Université de Montréal

Étude des déterminants génétiques et moléculaires de la scoliose idiopathique

Par

Dina Nada

Programme de Sciences biomédicales

Faculté de médecine

Thèse présentée à la Faculté des Études Supérieures

en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)

en Sciences biomédicales

Option Générale

Avril, 2017

Université de Montréal

Faculté des études supérieures

Cette thèse intitulée:

Étude des déterminants génétiques et moléculaires de la scoliose idiopathique

Présentée par :

Dina Nada

a été évaluée par un jury composé des personnes suivantes :

Dr. Stéphane Roy, président-rapporteur

Dr. Alain Moreau, directeur de recherche

Dr. Mark E Samuels, co-directeur de recherche

Dr. Gerardo Ferbeyre, membre du jury

Dr. Matthew B. Dobbs, examinateur externe

Dr. Hubert Labelle, représentant du doyen de la FES

Résumé

La scoliose idiopathique (SI) est une maladie complexe de la colonne vertébrale. Elle survient principalement à l'adolescence et affecte ~ 1-4% de la population mondiale pédiatrique avec une prévalence plus élevée chez les femmes. Dans la plupart des cas, la cause sous-jacente de la SI est inconnue, bien qu'une composante génétique soit bien reconnue. Un certain nombre de gènes et de loci candidats ont été proposés, mais peu ont été répliqués avec succès dans de multiples études. Les études d'association génomique (GWAS) ont identifié plusieurs gènes candidats prédisposant à la SI. Parmi ceux-ci, le locus *LBX1* est de loin celui qui a été répliqué avec le plus de succès dans différentes populations, bien que jusqu'à maintenant il n'ait pas été testé dans la population québécoise.

L'objectif principal de cette thèse, a été l'étude des déterminants génétiques de la SI par des approches complémentaires. Dans la première étude, de nouveaux gènes enrichis de variants rares, pouvant contribuer à la maladie ont été identifiés par séquençage d'exomes entiers (WES) dans une cohorte de patients québécois atteints de SI, suivi d'une deuxième phase de séquençage ciblé des 24 meilleurs gènes candidats dans une seconde cohorte indépendante. Parallèlement, nous avons effectué une approche par WES dans une famille multiplex unique constituée de trois sœurs affectées avec des parents sains. Nos résultats impliquent un nouveau gène, *FAT3*, non précédemment associé à la SI, comme un gène candidat d'importance pour cette condition. Dans la deuxième étude, nous avons effectué un GWAS pour tester les marqueurs génétiques associés à la dans la population québécoise. Les résultats complets de l'analyse GWAS dépassent le cadre de cette thèse. L'objectif principal de

l'analyse actuelle était de tester l'association du locus *LBX1*. Nos résultats appuient une association avec la région proche du gène LBX1 dans notre population. Dans la troisième étude, notre approche est une approche de gène candidat. Nous avons essayé d'élucider les corrélations génétiques et biochimiques des niveaux circulants en YKL-40 avec le risque de progression de la maladie. Une étude antérieure dans notre laboratoire a démontré que les patients atteints de SI ont un dysfonctionnement distinctif de la signalisation des récepteurs couplés aux protéines Gi inhibitrices (Gi), permettant leur classification en trois endophénotypes (FG1, FG2 et FG3). L'étude des profils d'expression de ces endophénotypes a montré une élévation significative de l'expression du gène *CHI3L1*, codant pour la glycoprotéine sécrétée YKL-40, chez les patients classés dans l'endophénotype FG1 par rapport à ceux classés dans les endophénotypes FG2 et FG3 qui sont plus enclins à développer une scoliose sévère. Nous avons démontré que les garçons du groupe FG1 présentent des niveaux plasmatiques significativement plus élevés en YKL-40 que les autres groupes tout comme les patients présentant une scoliose non-sévère. Les SNP dans le gène *CH13L1* ont montré une association significative avec les niveaux plasmatiques d'YKL-40 dans des cas non-sévère. L'analyse fonctionnelle *in vitro* a révélé que la glycoprotéine YKL-40 pourrait jouer un rôle protecteur dans le contexte de la SI en bloquant le défaut de signalisation Gi induit par l'élévation de l'ostéopontine.

En résumé, nous proposons une nouvelle association du gène *FAT3* avec la SI, répliquons l'association du gène *LBX1* dans une nouvelle cohorte et proposons un nouveau marqueur biochimique, YKL-40, associé à des formes non sévères de la SI.

Mots clés: Scoliose idiopathique, séquençage d'exome (WES), variants rares, GWAS, *FAT3*, *LBX1*, *CHI3L1,* YKL-40, endophenotypes, population québécoise

Abstract

Idiopathic Scoliosis (IS) is a complex disorder of the spine. It mostly occurs between 10 and 15 years old and affects ~1-4% of the global pediatric population with a much higher prevalence in females. In most cases the underlying cause of IS is unknown, although a genetic component is well recognized. A number of candidate genes and loci have been suggested, but few have been successfully replicated in multiple studies. Genome wide association studies (GWAS) have identified several candidate genes for IS susceptibility. Among these the *LBX1* locus is by far the most successfully replicated locus in different populations, although until now it has not been tested in the French-Canadian population. Few studies have attempted to detect rare causal variants in IS and this field of research is still in its infancy.

The primary goal of this thesis was to investigate the genetic component of IS through complementary approaches. In the first study, we aimed to find new genes enriched with rare variants, which might contribute to the disease. Hence, we performed whole exome sequencing (WES) in a French-Canadian IS cohort, followed by a second phase of targeted sequencing of the 24 best candidate genes in a replication cohort. In parallel, we performed WES in a unique multiplex family of three affected sisters with healthy parents. Our results implicate a novel gene, *FAT3*, not previously associated with IS, as a strong candidate for this condition. In the second study we performed a GWAS to test for genetic markers associated with IS in our French-Canadian cohort. The complete results of the GWAS analysis are beyond the scope of this thesis. The main objective of the current analysis was to test association of the *LBX1* locus. Our results support an association with the region near the

LBX1 gene in our French-Canadian population. In the third study, our approach is a candidate gene approach. We attempted to elucidate the genetic and biochemical correlates of circulating YKL-40 levels with the risk of spinal deformity progression in the context of IS. Our prior works have demonstrated that IS patients exhibit a distinctive G inhibitory (Gi) protein-coupled receptor signaling dysfunction, which enabled their classification into three distinct biological endophenotypes (FG1, FG2 or FG3). Previous microarray analysis revealed a significant elevation in the expression of *CHI3L1* gene, encoding for the secreted glycoprotein YKL-40, in IS patients classified in FG1 endophenotype when compared to FG2 and FG3 ones, which are more prone to develop a severe scoliosis. In this study, we demonstrated that IS males classified in FG1 endophenotype exhibit significant higher plasma YKL-40 levels than controls and other IS endophenotypes. Furthermore, the non-severe scoliosis group showed significant higher levels of YKL-40 than controls. SNPs in *CHI3L1* gene showed significant association with YKL-40 plasma levels in non-severe cases. Functional *in vitro* analysis showed that YKL-40 could play a protective role in the context of IS by altering the Gi-signaling dysfunction induced by the elevation of osteopontin in IS.

In sum, we propose a novel association of *FAT3* gene with IS, replicated the association of *LBX1* gene with IS in a new cohort, and propose a new biochemical marker, YKL-40, associated with non-severe forms of IS and has a plausible protective role in IS.

Key words: Idiopathic scoliosis, whole exome sequencing, rare variants, GWAS, *FAT3*, *LBX1*, *CHI3L1,* YKL-40, endophenotypes, French-Canadian population

Materials table

List of tables

List of figures

Abbreviations list

ABI	Applied Biosystems
ACAN	Aggrecan
AD	Autosomal dominant
AFLPs	Amplified fragment length polymorphisms
AIS	Adolescent idiopathic scoliosis
AKAP2	A-kinase anchoring protein 2
AR	Autosomal recessive
bps	Base pairs
BNC2	Basonuclin2
BRCA1	Breast cancer 1
BRCA2	Breast cancer 2
C17orf67	Chromosome 17 open reading frame 67
CDCV	Common disease common variant
CDH7	Cadherin 7
CDRV	Common disease rare variant
CFH	Complement factor H
CHI3L1	Chitinase 3-like 1
CNVs	Copy number variants
COL1A1	Collagen type I alpha1
COL1A2	Collagen type I alpha2
COL2A1	Collagen type II alpha1
COMP	Cartilage oligomeric matrix protein
CTLSO	Cervico-thoraco-lumbo-sacral orthosis
CTS	Computed tomography scan
3D	Three-dimensional
DOT1L	DOT1-like (DOT1 Like Histone Lysine Methyl transferase)
DNA	Deoxyribonucleic acid

DZ	Dizygotic twins
E	Environment
EpiG	Epigenetics
ESR1	Estrogen receptor1
ESR2	Estrogen receptor2
ELN	Elastin
FAM101A	Family with sequence similarity 101, member A
FAT3	FAT Atypical Cadherin 3
FBN1	Fibrillin-1
FBN2	Fibrillin-2
FG1	Functional group 1
FG2	Functional group 2
FG3	Functional group 3
G	Genes
Gi	G inhibitory
Giα	G inhibitory alpha
Gi PCR	G inhibitory protein-coupled receptor
GPER	G protein-coupled estrogen receptor
GPR126	G protein-coupled receptor 126
GWAS	Genome-wide association study
HOXB8	Homeobox B8
HOXB7	Homeobox B7
HOXA13	Homeobox A13
HOXA10	Homeobox A10
HSPG2	Heparan sulfate proteoglycan 2
IGF1	Insulin like growth factor 1
IL6	Interleukin-6
IL-17RC	Interleukin 17 receptor C
Indels	Insertion-deletions
IS	Idiopathic scoliosis

Kb	Kilobase pair
KCNJ2	Potassium voltage-gated channel subfamily J member2
LA	Linkage analysis
LBX1	Lady bird homeobox-1
LD	Linkage disequilibrium
LDL	Low-density lipoprotein
LOD	Logarithm of the odds
MAF	Minor allele frequency
MATN1	Matrilin 1, cartilage Matrix Protein
MB	Megabase pair
MiRNA	Micro ribonucleic acid
MMP3	Matrix metallopeptidase 3
MRI	Magnetic resonance imaging
MTNR1A	Melatonin receptor 1A
MTNR1B	Melatonin receptor 1B
MZ	Monozygotic
NGS	Next generation sequencing
P	Phenotype
PAX1	Paired box 1
PCP	Planar cell polarity
PCR	Polymerase chain reaction
PITX1	Paired-like homeodomain 1
POC5	POC5 centriolar protein
PTK7	Protein tyrosine kinase 7
RFLPs	Restriction fragment length polymorphisms
RNA	Ribonucleic acid
SOLiD	Sequencing by oligonucleotide ligation and detection
SOX9	SRY-Box 9
SNPs	Single nucleotide polymorphisms
SNVs	Single nucleotide variants

STRs	Short tandem repeats
TGFB1	Transforming growth factor beta 1
TLSO	Thoraco lumbo-sacral orthosis
TPH1	Tryptophan hydroxylase 1
VANGL1	VANGL Planar Cell Polarity Protein 1
VDR	Vitamin D receptor
WES	Whole exome sequencing
WGS	Whole genome sequencing
XLD	X-linked dominant
ZIC2	Zic family member 2

This work is dedicated to my family

.

"Pursue one great Decisive aim with force and determination"

Carl Von Clausewitz

"There is no substitute for hard work"

Thomas A. Edison

Acknowledgements

I would first like to express my deep appreciation to my supervisor Dr. Alain Moreau, who gave me the chance to join his lab and grow as a research scientist. My sincere thanks for your incredible support both at the academic and personal levels, for your continuous encouragement and for believing in me.

I would like to express my sincere gratitude and deep thanks to my co-supervisor Dr. Mark Samuels. Your passion for research has inspired me and taught me how to develop my research capabilities. My deep thanks for your wise guidance and valuable discussions which helped to deepen my scientific knowledge. Your advice and support have been invaluable.

I am so grateful to my thesis committee members Dr. Carolina Alfieri and Dr. Zoha Kibar for their valuable advices and support.

I sincerely thank all the jury members for kindly accepting to evaluate my thesis:

Dr. Stéphane Roy, Dr. Alain Moreau, Dr. Mark E Samuels, Dr. Gerardo Ferbeyre, Dr. Matthew B. Dobbs and Dr. Hubert Labelle.

I would also like to sincerely thank all my colleagues in the lab:

Dr. Cédric Julien for his incredible help and support. Anita Franco and Saadallah Bouhanik for their sincere help and support both on the personal and professional levels. My dear friends Lakshmi Suvarnan, Nancy Karam, Mohamed Elbakry and Niaz Oliazadeh for their nice company and agreeable time we spent together in the lab. Dr. Kristen Fay Gorman and Dr. Jean Francois Lavoie for their valuable scientific discussions. Dr. Dashen Wang, Mme Ginette

CHAPTER I

INTRODUCTION

1. Studying genetics of complex diseases

1.1 Human genome

The human genome is the hereditary material, which consists of the chemical deoxy ribonucleic acid (DNA). DNA contains the genetic information needed for a human being to be a functional organism. Most of the human genome resides within the cellular nuclei, and a small part is contained within the mitochondria in the cytoplasm. Each nucleated cell in the body has its own copy of the human genome, which includes from 20,000 to 50,000 genes according to the definition of the term. The DNA is composed of around 3.3 billion base pairs, however less than 1.5% of it encodes for proteins. Approximately 50% of the linear length of the genome is composed of single copy DNA sequences. The other 50% are repetitive DNA sequences [1].

1.2 Human genomic variation

Understanding genetic and genomic variation is the key for understanding genetics of human diseases. The DNA sequence is roughly 99.5% identical between any two unrelated individuals [1]. Some of the DNA sequence variations among individuals are responsible for their phenotypic differences and quantitative trait variations as well as disease susceptibility. Currently, many tens of millions of single variants and more than a million complex variants are discovered and reported [1]. A variant is sometimes defined as a polymorphism, when its frequency is more than 1% in a population. Common variants are often defined as those with a minor allele frequency > 5%, while uncommon variants have frequency 1-5% and rare variants are defined as those with frequency < 1%. There is not an absolute consensus in the

field of human genetics regarding the precise definitions of these terms; however, for this thesis these definitions will be employed.

The different types of variants or polymorphisms in the human genome include single nucleotide polymorphisms (SNPs), insertion- deletion (indels) polymorphisms, copy number variants (CNVs), inversion polymorphisms and large chromosomal rearrangements (see **Figure 1** and **Table I**).

SNPs are the most common type of variation in the human genome. This type of variation includes only a single base of DNA. On average a SNP occurs once every 1000 bps (base pairs) in the genome [1]. The SNP consortium was organized in 1999 in order to create a public resource of information on SNPs [2]. This was followed by the International HapMap project and the 1000 genomes project, with a goal of identifying and characterizing all common SNPs and SNP haplotypes in the world human population [3-6]. Most of the SNPs identified by these projects lie in the non-protein-coding regions. Nonetheless, more than 100,000 exonic SNPs have been identified and reported to date [1].

Insertion-Deletion (Indels) polymorphism or variation can be due to insertion or deletion either of a single bp or up to around 1000 bps. More than one million indels have been reported.

Copy number variants (CNVs) are similar to indels but there is variation in the number of copies of bigger segments of the genome which can reach up to hundreds of thousands bps. Minimum size of CNVs is not determined, but is typically defined as 1000 bp (1 kb). CNVs that include exons can influence gene dosage and might have significant role in diseases.

Inversions are rearrangements of the DNA sequence where a specific chromosomal region is reversed. Inversion can differ in size from few bps up to many mega bps.
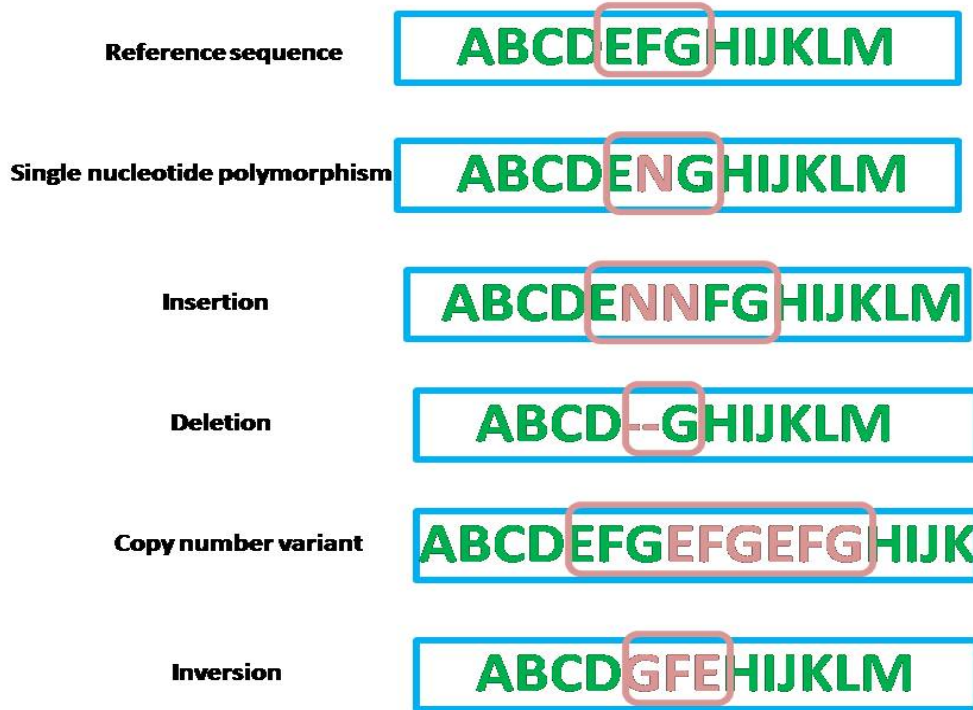
Figure 1: Demonstrating different types of common variation in the Genome.

Table I: Types of common variation in the Genome (modified from[1]).

Type of variation	Size (approximately)	Number of alleles
SNPs	1bp	2 (most of the times)
Indels	From 1bp to more than 100 bps	2, in case of microsatellites: 5 or more
CNVs	From 1kb to more than 1Mb	2 or more
Inversions	From few bps to more than 1 Mb	2

bp: Base pair, kb: kilobase pair, MB: megabase pair.

1.3   Simple versus complex diseases

Most diseases are presumed to involve interplay between genetic and environmental factors. Monogenic diseases or Mendelian diseases are those with a very high genetic component where the role of environmental factors is minor. This class of diseases is usually caused by variation in one single locus or gene. With the presence of an informative pedigree, the mode of inheritance of monogenic diseases can be easily identified. The mode of inheritance can be either autosomal dominant or co-dominant where the presence of one variant allele causes the disease to manifest while the presence of two variant alleles leads to a more severe form of the disease, or recessive where two variant alleles should be present in *trans* for the disease phenotype to be manifested. Other modes of inheritance include x-linked or y-linked and mitochondrial patterns. Most simple Mendelian diseases are rare with a low prevalence among populations, examples include Marfan syndrome [7], Huntington disease [8], cystic fibrosis [9], and fragile X syndrome [10].

On the other hand, in complex diseases, no single variant is responsible for the disease phenotype to manifest, but rather a complex interplay between multiple variants and environmental factors are responsible for diseases manifestation. Most major diseases affecting humans belong to this category including Alzheimer's disease, asthma, Parkinson's disease, and autoimmune diseases [11].

The difference between monogenic and complex diseases is not always obvious. The role of environmental factors could be very important in some monogenic diseases, for example phenylketonuria occurs in patients who carry the disease-causing gene only if they have phenylalanine in their diet. Moreover, it has been shown that the severity of two

monogenic disorders, cystic fibrosis and sickle cell anemia, may be modulated by other genes [12, 13]. Also single genes were reported to be responsible for the manifestation of complex diseases in some families e.g. *BRCA1* [14] and *BRCA2* [15] in breast cancer.

Most of the known clinically important genetic diseases and traits are catalogued on the Online Mendelian Inheritance in Man database (OMIM) (www.omim.org).

Penetrance represents the proportion of individuals who carry a specific genotype and manifest the associated phenotype. Penetrance varies in different diseases from approaching 100% in monogenic diseases to much lower penetrance in complex diseases.

Heritability is a measure of the genotypic variance that explains the phenotypic variance of a trait or a disease. High scores of heritability indicate strong genetic contribution and low heritability scores implies more effect of environmental factors. Heritability of different diseases is typically estimated by family and twin studies [16, 17]. A difference in the prevalence of the disease among relatives of an affected individual and its prevalence among relatives of a healthy individual indicates high heritability and more genetic influence on the disease. In twin studies, the prevalence of the disease in monozygotic twins (MZ) and dizygotic twins (DZ) is compared. MZ twins are genetically identical, while DZ twins share 50% of their DNA and both types of twins share many early and late environmental factors [18]. A higher concordance rate of the diseases in MZ twins compared to DZ twins implies high genetic impact and gives higher heritability scores. **Table II** summarizes the estimated heritability for some complex diseases together with the proportion of this heritability explained by known susceptibility variants for each trait.

Table II: Estimated heritability for some complex diseases, modified from [19].

Disease	Heritability (%)	Proportion of heritability explained
Idiopathic Scoliosis	38 [17]	<2 [20]
Alzheimer's disease	79	23.22
Bipolar disorder	77	2.77
Breast cancer	53	12.52
Coronary artery disease	49	25.15
Crohn's disease	55	13.43
Prostate cancer	50	31.16
Schizophrenia	81	0.39
Systemic lupus	66	13.2
Type 1 diabetes	80	13.63
Type 2 diabetes	42	27.93

1.4 Approaches to studying genetics of diseases
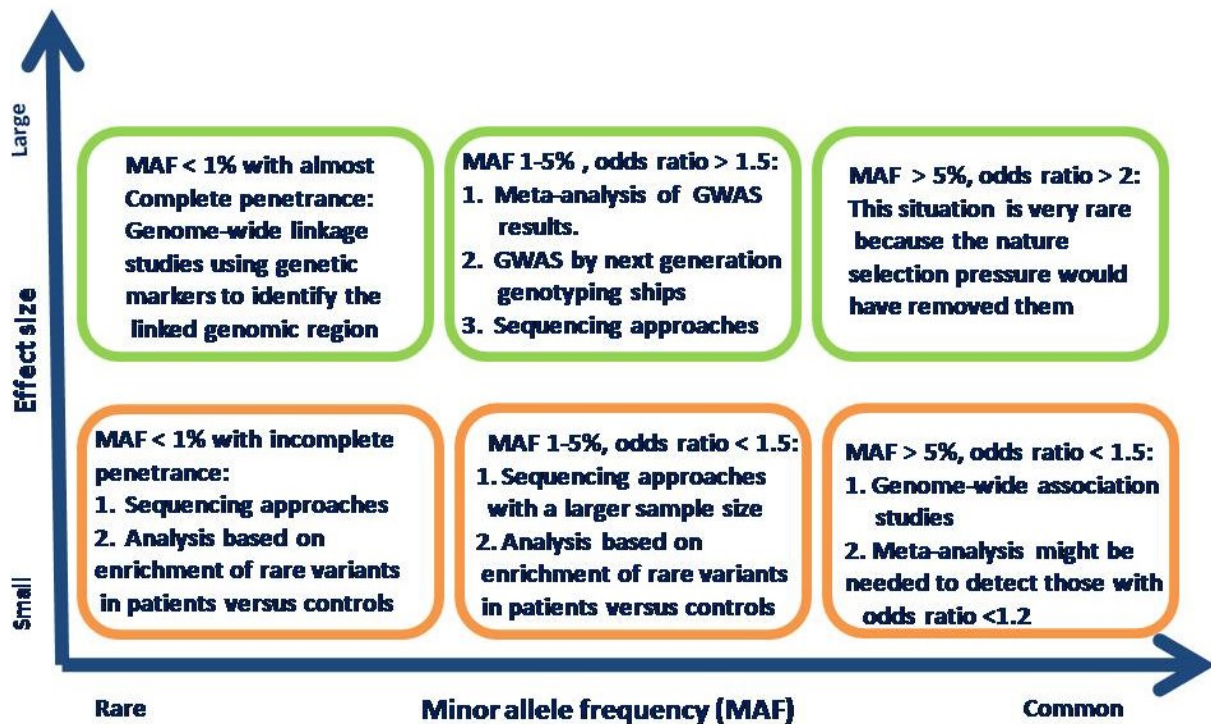
Figure 2: Different approaches to studying disease-associated variants (adapted from [21]).

1.4.1 Family-based approaches (Linkage studies)

Linkage analysis (LA) studies are based on the investigation of an informative family pedigree with the aim of identifying a chromosomal region, which is shared among affected family members and different from unaffected family members. Many types of genetic markers have been employed for linkage map construction. In early studies Restriction Fragment Length Polymorphism (RFLPs) and Amplified Fragment Length Polymorphisms (AFLPs) were more used [22, 23]. RFLPs are differences in length between different DNA segments arising from different positions of restriction enzyme sites caused by different types of genetic variations. AFLPs were markers in which Polymerase Chain Reaction (PCR)

added adaptors to the end of digested fragments followed by amplification. Microsatellites or Short Tandem Repeats (STRs) have also been used as markers for LA studies. These are present as multiallelic due to different numbers of repeated DNA sequence which is composed of 2-4 nucleotides [1]. Microsatellites are PCR-based, polymorphic and highly reproducible which made them good markers for LA studies. Currently, high density SNP panels for high resolution map for LA are being employed [24]. The ability to detect genomic loci harboring causative variants is highly affected by the density of SNP panels, knowledge of the amount of Linkage Disequilibrium (LD) in the genome and frequency of the haplotypes [25]. LD is a measure of co-inheritance of alleles of two different loci (usually close to each other on a given chromosome). The higher the LD the more chance that alleles of the two loci are co-inherited [1]. LA has proved to be powerful and successful in identifying genomic regions harboring causal mutations for monogenic diseases segregating in individual families [26, 27]. On the other hand, LA methods have limited success in mapping genomic risk regions for complex diseases. The reason behind this might be due to the presence of multiple variants of low or moderate effects within multiplex families, such that the power of LA algorithms is not enough to detect linkage [28]. Rischet al suggested that even studies which include 2000 sib pairs are unable to detect loci with a moderate effect[29].

1.4.2 Population-based approaches (Association studies)

The concept of association studies is to compare the frequency of genomic variants in cases versus controls to detect a statistical association between alleles of any of these variants and the disease of study. Association studies were initially performed with candidate genes, which were chosen due to a hypothesis based on their biological functions or on a previously

identified genomic region to be linked with the disease. With the advances of SNP genotyping technologies and the availability of high-density polymorphism databases, association studies were extended to study hundreds of thousands or millions of SNPs throughout the whole genome in an unbiased approach (GWAS: Genome wide association study). SNP associations together with large haplotype blocks involving the associated SNPs can be detected which implicates a disease susceptibility locus. Candidate genes association studies suffered limited success in replication studies. GWAS on the other hand proved their effectiveness in identifying many common variants to be associated with many complex diseases [30, 31]. GWAS became available for investigators after the completion of the Human Genome and the HapMap projects in 2003 and 2005 respectively [32, 33].

The first GWAS was published in 2005, which reported an association between a polymorphism in the complement factor H gene (*CFH*) and age-related macular degeneration [34]. This was followed by many successful GWAS published in 2006-2007 where many common variants were reported to be associated with several complex diseases such as Crohn's disease [35, 36], type I and II diabetes [36-38], obesity [39], coronary artery disease and rheumatoid arthritis [36]. As of now, hundreds of GWAS have been performed and contributed in the identification of thousands of variants associated with many complex diseases [31, 40-44]. However, care has to be taken in GWAS analysis and correction for population stratification should be done. It has been shown that population stratification can lead to false positive results where the differences in allele frequencies could be unrelated to the tested disease [45, 46]. In general, many reported GWAS associations have proved difficult to replicate, either due to differences in the study populations, definitions of the precise phenotype, or the presence of Type I (false positive) artifacts. Nonetheless, a

significant proportion of GWAS loci have been replicated in multiple studies; the most obvious being the major histocompatibility (HLA) locus on chromosome 6, which is universally observed to be strongly associated with many different autoimmune disorders.

1.4.3 Sequencing approaches

The main aim of the sequencing approach is to identify rare variants. In fact, sequencing techniques have the potential to identify both rare and common variants where the most comprehensive design to study variation in any disease would be to sequence the whole genome in a large population [47], however for studying common variants, dense SNP genotyping arrays remain for the moment less expensive than whole genome sequencing. That said, sequencing approaches designed to detect rare variants would detect common ones as well obviously.

Previously, DNA was sequenced by the Sanger method [48], a technology which was although a breakthrough in the field of molecular genetics, yet was very expensive for sequencing large number of individuals for research laboratories. More recently, large-scale sequencing has been universally performed using the so-called Next Generation Sequencing (NGS) technologies. NGS refers to sequencing approaches where massive sequencing reactions are carried out in parallel, typically without the need for electrophoresis. The 454 pyrosequencing method was introduced in 2004 [49] and is now commercialized via Roche. The Solexa reversible termination method was introduced in 2006 [50] and is now commercialized by Illumina. Sequencing by oligonucleotide ligation and detection (SOLiD) method was introduced in 2007 by ABI (Applied Biosystems) [51], now commercialized by Life Technologies. Illumina introduced the HiSeq and MiSeq variants of the Solexa

chemistry method, and Life Technologies introduced Ion Torrent techniques [52] to replace SOLiD. NGS technologies lead to drastic reduction in the cost and time of the sequencing approaches, which consequently dramatically affected the field of study of genetics of complex diseases. However, whole genomic sequencing (WGS) is still expensive for laboratory research. In 2007, a sequence capture technology was introduced by Nimblegen whereby microarrays were used to capture specific sites of the genome to be sequenced (typically protein-coding exons) [53]. More recently sequence capture has been adapted to solution hybridization using biotin-labelled RNA oligonucleotide probes to target desired genomic content to be studied (**Figure 3**). Targeted sequencing gave researchers a less expensive alternative to WGS in terms of total cost, although the cost per base pair is higher due to the targeting and capture steps in the protocol. Targeted sequencing may focus on subsets of genes belonging to biological pathways of interest, or loci previously linked or associated with a disease. More generally, targeting is used to capture and analyze all known genomic protein coding regions, which comprise about1.5 % of the genome and have the potential of harboring many or most of high penetrance disease causing variants (whole exome sequencing, WES) [54, 55]. Sequencing approaches can be done for families or case/ control populations; where either sequencing multiply affected families (multiplex) searching for rare variants segregating with the disease or sequencing a population of cases and healthy controls and looking for genes enriched with rare variants in cases versus controls. It is important to note that targeted sequencing approaches have more potential for artifacts than either Sanger or whole genome sequencing, due to the hybridization capture step itself as well as the requirement for more cycles of PCR. Certain regions of the genome (high GC-content exons for example) cannot be uniquely targeted, while some protein-coding exons are

captured much more efficiently than others for unclear reasons. A high level of statistical variance among all targeted exons in WES requires a higher level of mean coverage of targeted regions, to ensure that most targeted regions are covered at sufficient read depth to detect heterozygous rare variants.
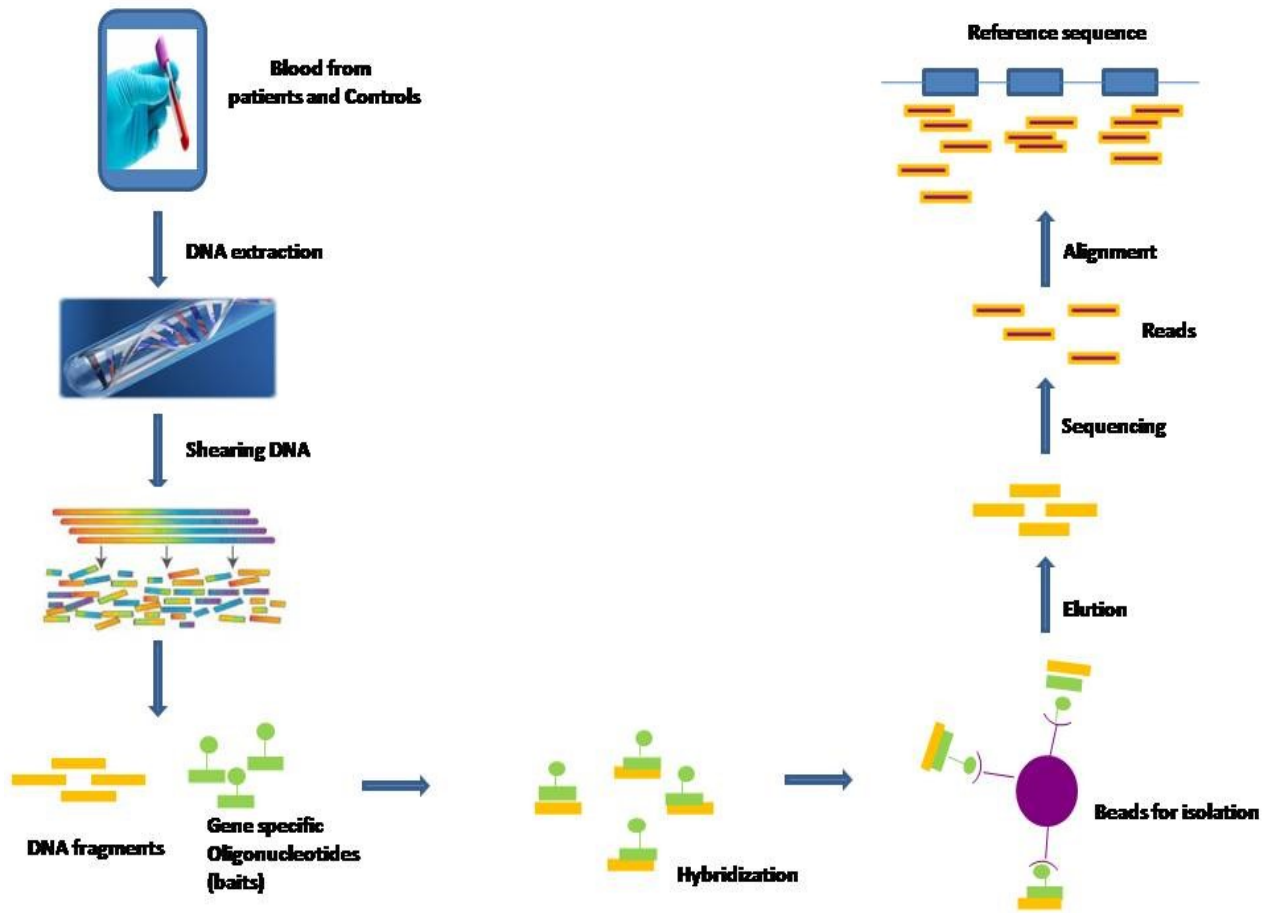
Figure 3: Schematic for targeted next generation sequencing. DNA is extracted from the study subjects, then sheared and allowed to hybridize to the baits (oligonucleotides specific for the target regions). These baits are tagged to be easily isolated by capture on immobilized beads (for example). The captured DNA is eluted, amplified and sequenced on one of several next generation platforms (Illumina HiSeq, LifeTech Proton for example). For mutation detection the sequenced DNA fragments (reads) are aligned and compared to the human genome reference sequence (adapted from [57]).

The most important challenge of the sequencing approaches is the analysis of the results. If the analysis of rare variants is based on single variant associations, large sample sizes and/ or variant effect sizes will be needed to attain sufficient statistical power in order to detect the associations. To overcome this problem, new statistical methods have been developed to boost the statistical power of analytical methods. The main idea of these methods is to look for multiple rare variants in either a gene or biologically relevant pathway rather than testing only single variants, which is routinely done in GWAS. Among those methods is the collapsing gene burden test [56] which is a simple test where the number of rare variants (typically of MAF ≤1%) in one gene is compared between patients and controls. This test is used under the assumption that all rare potentially protein altering variants act in the same phenotypic direction with the same magnitude.

2. Idiopathic Scoliosis (IS)

2.1 Definition

The term scoliosis (skŏ'lē-ō'sîs), "Σκολίωση", first used by the Hippocrates (A.D. 460–370) and Galen (A.D.131-201), is a Greek word which means "crookedness". It is a three-dimensional (3D) spinal deformity with a Cobb angle of 10° or more in the coronal plane measured on a standing radiograph. The Cobb angle is the anglulation used to measure the degree of curvature of the spine (**Figure 4**), it is the angle formed between a line drawn parallel to the superior endplate of the upper vertebra of the scoliotic curve and that drawn parallel to the inferior endplate of the lower vertebra of the scoliotic curve. The deformity of the spine is most often combined with asymmetric trunk and extremities [58].
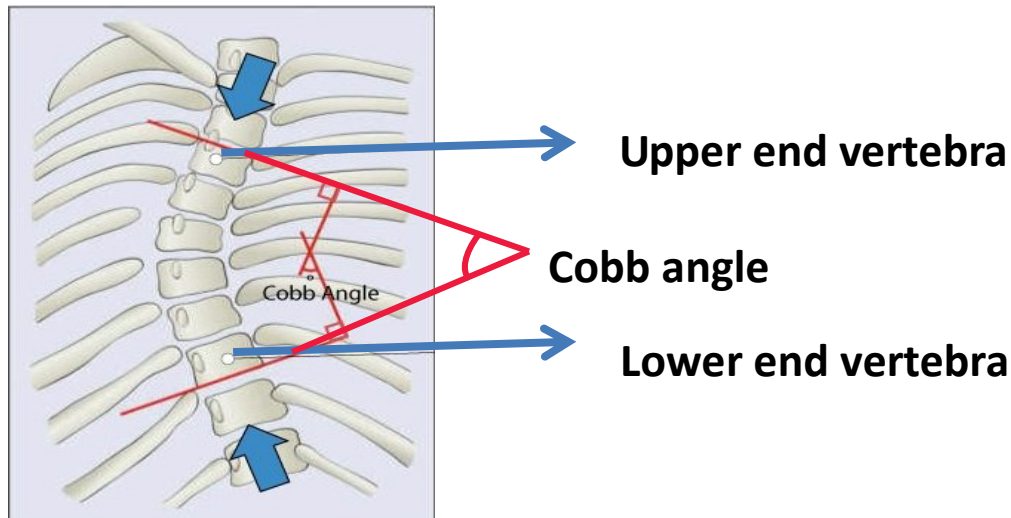
Figure 4: The Cobb angle used for measurement of the degree of curvature of the spine (adapted from https://www.coreconcepts.com.sg/article/cobb-angle-and-scoliosis).

2.2 Classification

There are many types of scoliosis where the cause is well-known, such as congenital, neuromuscular, degenerative, and syndromic while Idiopathic scoliosis is the type whose etiology remains most elusive. Idiopathic scoliosis comprises approximately 80% among all types of scoliosis [59]. Idiopathic scoliosis can be further classified according to the pathological type into structural or functional, according to the magnitude of the curve where the severity increases with increasing the Cobb angle, according to the anatomical site of the curve into cervical, thoracic, thoracolumbar or combinations if there is more than one angle of curvature. Idiopathic scoliosis is often classified according to the age of onset into infantile (0-3) years, juvenile (4-9) years, adolescent (10-15) years which is the most common type, and an adult type which occurs above the age of 18 [58].

2.3 Epidemiology

The overall prevalence of IS ranges somewhat widely, with estimates varying from 0.47-5.2% [60]. There are substantial differences among the studies with regard to prevalence of IS, potentially due to different definitions of scoliosis, different age groups, different study designs and methods of diagnosis [61-67]. The most important factors which affect the prevalence of IS are ethnicity, age and gender. The prevalence of IS from 17 countries has been summarized in a meta-analysis study which included 36 studies [68]. The reported IS prevalence while different in detail, overlaps substantially in different regions: in Australia and the Middle East it was 1.9%; in Asia it ranged from 0.4 to 2.5%; in North America it was from 0.4 to 3.9%; in Europe (Spain) it was reported to be from 0.7% to 7.5%. This study reported a general prevalence of curves $\geq 10^o$ to be 1.34%, curves $\geq 20^o$ to be 0.22% and severe curves to be 0.07%. Generally prevalence decreases in areas near the equator and increases towards the north, a latitude cline that has also been observed for other common disorders including some autoimmune diseases [58, 69]. Age is another factor, which affects the prevalence. It was reported that the prevalence of IS is 3.12% in the age 16-17 years, 1% in the age 11-12 years and 0.12% in the age 6-7 years [64]. Another German study reported a prevalence of 6.5% in the group of age 11-13 years and 11.1% in the group of age 14-17 years [61]. It is well known that the prevalence of IS is greater in females than males, although the female to male ratio also varies widely among different studies from 1.5 to 11 due to the previously mentioned differences among those studies [70, 71]. A study which included a very large cohort of IS patients, recruited children from 10 years old and followed them until 19 years old, the prevalence of curves $\geq 10^o$ were reported to be 2.5%, curves $\geq 20^o$

were 1.4% and severe curves were 0.33% [72]. The female: male ratios were 2.7, 4.5, 8.1 and 8.4 for curves $\geq 10^{o}, \geq 20^{o}, \geq 40^{o}$ and very severe respectively [72]. Patients who are more at risk of developing severe cases are the youngest, shortest, premenarchal and the higher the initial Cobb angle [73].
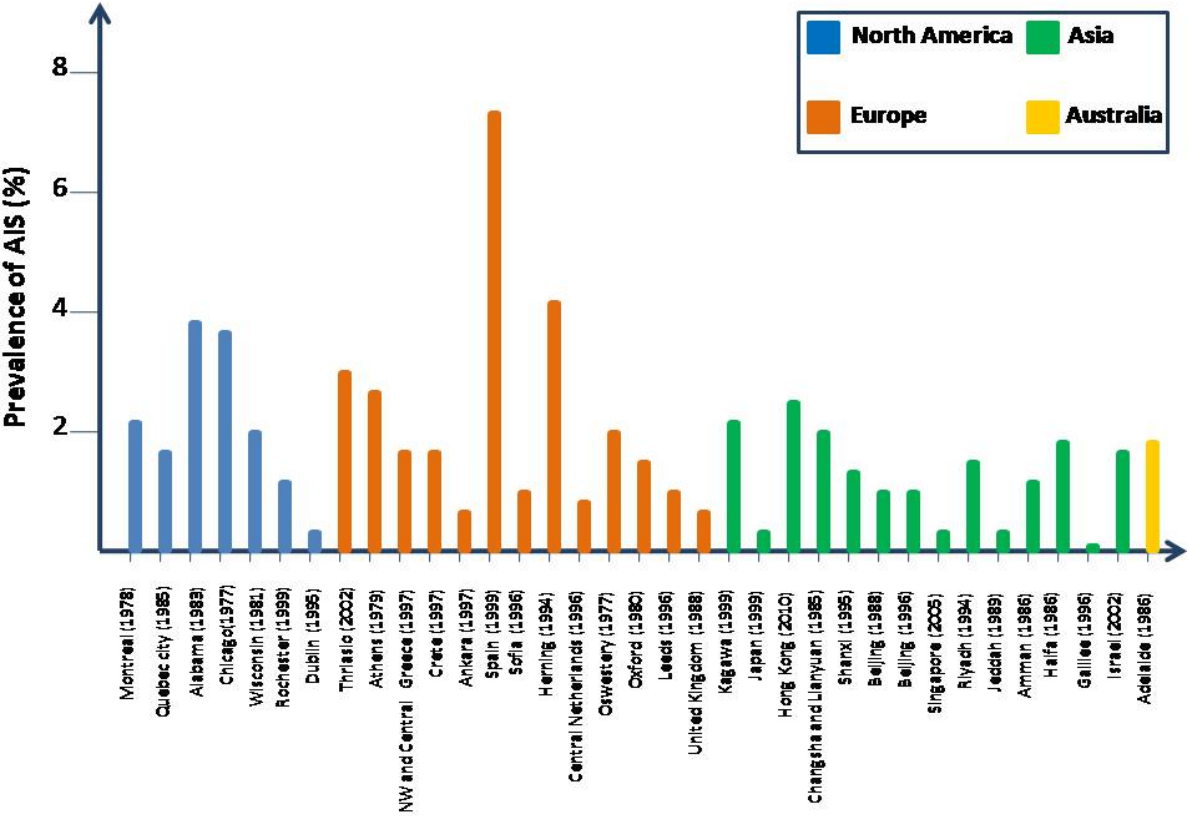
Figure 5: Prevalence of AIS worldwide. The prevalence of AIS (Adolescent Idiopathic scoliosis) obtained from a meta-analysis of 36 school scoliosis screening studies performed in 17 countries (adapted from [58]).

2.4 Diagnosis

2.4.1 Screening

Screening of scoliosis in schools is done with the purpose to detect the condition at an early stage, which enables early intervention. Clinical signs of scoliosis include asymmetry of the trunk, shoulder and scapular prominence and uncentered head over the pelvis (**Figure 6**). The most common screening test for IS is the Adam's forward bending test where the patient bends forward and examined from behind, if one side is higher than the other it is a sign of scoliosis while if the back is completely straight so the patient is normal (**Figure 6**). An apparatus called scoliometer is used to measure the angle of the trunk deformity which when is 7° or more is referred to a physician for further examination (**Figure 6**). In some countries of Europe and East Africa another test is used called Moiré topography in addition to the forward bending test with or without a scoliometer [58]. There is an ongoing debate for the effectiveness of screening for IS. Not all health professionals agree on applying a screening program because there is a high number of false positives and negatives [74]. Adbor et al suggested that screening can reduce the rates of treatment and is therefore cost effective [75]. But many other researchers disagree because of the low sensitivity of the test and the non-predictable natural history [75]. However, the American Academy of Orthopedic Surgeons, the Scoliosis Research Society, the American Academy of Pediatrics and the Pediatric Orthopedic Society of North America decided that females should be screened twice at the age of 10 and 12 years, while males should be screened only once at the age of 13 or 14 years[76].
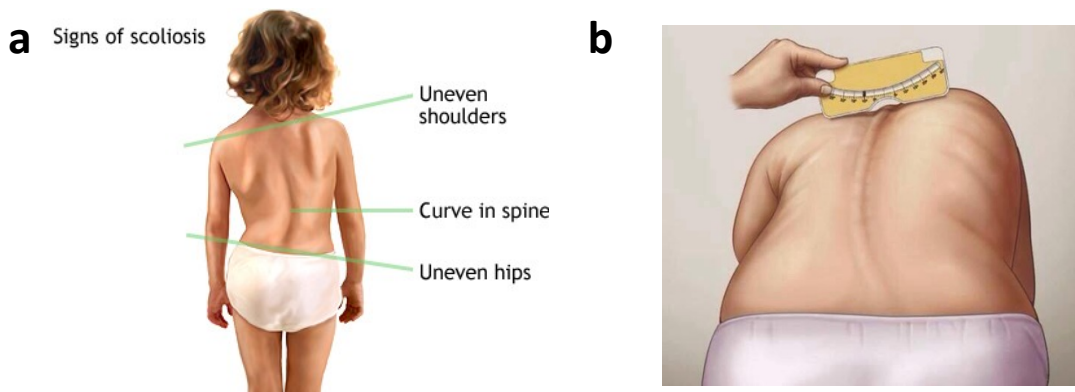
Figure 6: a. Signs of Scoliosis **b.** Adam's forward bending test using a scoliometer for measuring the trunk deformity (http://umm.edu/health/medical/reports/articles/scoliosis), (https://www.pinterest.com/ottawascoliosis/school-screening-for-scoliosis-what-we-dont-have-i/).

2.4.2 History

The evaluating physician starts by investigating the patient's history specially any family history, the age of onset and the history of progression of the disorder. Associated neurologic and respiratory symptoms as pain are investigated. Further pulmonary examination is done in case of history of failure to thrive in childhood and difficulty in breathing. Previous surgical and medical history is important to rule out other syndromes.

2.4.3 Physical examination and imaging

At the beginning and the end of the pubertal spurt, the skeletal maturity should be assessed. This help to predict the growth potential remaining and subsequently the potential for curve progression. The most reliable method to measure skeletal maturity is the Risser sign [77]. The process of ossification of the iliac epiphysis is divided into six stages up to its total fusion with the iliac crest and the formation of the iliac apophysis. Skeletal maturity is

indicated by Risser (4-5). Height also should be measured and recorded. The physician performs a standing posterior-anterior and lateral radiograph of the spine including the hip joints. The radiography is done posterior-anterior to reduce the amount of radiation to important organs in the upper torso [78]. These radiographs are viewed such that the heart is on the left side, this view is exactly like during the clinical assessment where the patient is examined from behind during Adam's forward bending test. The curve magnitude is measured using the Cobb angle method. Other malformations in the ribs or vertebrae could also be detected. The curves are denoted by the direction of its convex side and the site of its apex. Other imaging methods like Computed Tomography Scan (CTS) and MRI are not necessarily done for IS patients except in some cases with pain, severe curvature, rapid progression of the deformity or suspected tumors, and in cases who present abnormal neurological examination as well [78]. In addition, patients who require surgical treatment undergo MRI scan to rule out any intraspinal abnormality.

2.5 Management

Management of IS patients generally ranges from observation, physiotherapy, bracing and treatment. Each patient is assessed individually and a treatment plan is tailored for each patient according to his case. Prognostic factors are very important to help the physician decide the best treatment plan for his patient. The patient's prognosis depends mainly on the age of onset, Cobb angle and the Risser grade [79]. Factors which contribute positively to disease progression include family history, defects in connective tissue (skin and joints), flattening of kyphosis in the thorax [80], angle of deformity of trunk exceeding 10° and growth spurt [81].

2.5.1 Physiotherapy

IS patients undergo physiotherapy as part of a conservative therapy [82-84]. Physiotherapy can either be in the form of exercises or physical treatment like electro stimulation of the paravertebral muscles [82-84]. Whether physiotherapy is effective for IS is under debate [84-86]. Several publications reported a positive effect of exercises on reducing the risk of disease progression [87-89] and reducing the Cobb angle [90-93]. On the other hand, many reviews report that those studies do not offer sufficient scientific evidence of efficacy in the reduction of scoliosis progression [84-86]. Of note, certain physiotherapy can be useful to support brace treatment and to prepare IS patients for surgical operations[82, 83].

2.5.2 Bracing

Bracing is recommended for IS patients with the aim to reduce further curve progression, alleviate back pain, prevent respiratory dysfunction and ameliorate esthetics [82]. Bracing is indicated for skeletally immature patients (Risser<3) with Cobb angle 30°-$40^{o.}$ Bracing is also indicated for curves around 25°-30° that showed a progression of 5° or more. There are many types of braces which can be classified according to the material into rigid and soft, according to the time it is being worn into full time, part time and night time, or according to its design based on the site of curvature into cervical, thoracic, lumbar and sacral [94]. Examples of rigid TLSO (thoraco-lumbo-sacral orthosis) braces, which are designed for thoracic, low thoracic, thoracolumbar and lumbar scoliosis are Cheneau brace, Boston, Gensingen, Lapadula, lyonese, Sibilla, Sforzesco and Progressive action shortBrace [95]. Examples of soft braces are Spinecor, Olympe, TriaC, Spinealite and St Etienne [95]. An example of the CTLSO (cervico-thoraco-lumbo-sacral orthosis) brace is the Milwaukee

brace. The Charleston and the Providence braces are examples of night time braces [95]. In most studies, the evidence for the effectiveness of bracing for IS patients is of low quality [95]. It is hard to compare studies which addressed the effectiveness of brace treatment because of the inconsistency of the methods of evaluation and the presence of many different additional factors that could have affected their results [95].

2.5.3   Surgical treatment

Surgical treatment is indicated for severe cases when the curve magnitude is $\geq 45^{\circ}$ who are equal or less than Risser 2 and with curves ≥ 50 who are equal or greater than Risser 3. The aim of surgeries is to stop the curve progression and improve the balance and alignment of the spine. During surgeries spine fusion is performed by instrumentation and bone grafting. Fusion techniques have started early in the 1960s when Harington introduced the hook and rod constructs [96]. Later Luque introduced wires for segmental fixation [97]. Nowadays, there is third generation segmental fixation using pedicle screws [78]. In all these techniques bony anchors which include the hooks, wires and pedicle screws are placed to the vertebrae and connecting them to rod constructs [78]. Fusion surgeries can be done anteriorly, posteriorly or both. Preoperative planning should consider many factors including the stage of skeletal maturity, curve flexibility, type and magnitude of curvature and spinal balance [78].

2.6 Etiopathogenesis and different hypotheses

2.6.1 Neuromuscular

A postural equilibrium dysfunction has been suggested to play a causative role in IS [98]. Consequently, defective visual and vestibular functions have been investigated by many studies to understand their role in IS [99-103].Visual, vestibular and proprioceptive neural pathways interconnect together in certain areas of the brainstem. Lesions in these areas were found to be associated with scoliosis [104] and induction of scoliosis has been done by damaging these areas in animal studies [105]. Altered function of cerebral cortical and sub cortical has been suggested to contribute to pathogenesis of IS [101, 106, 107]. A significant reduction in the length of the spinal cord relative to the vertebral column in severe IS cases has been documented by MRI [108]. This lead to and supported the theory of disproportional neuro-osseous growth in the etiology of IS [109-112]. Differences in the volume of different regions of the brain between IS and matched controls has been reported [113]. Ten regions were shown to be significantly larger in IS patients, and twelve regions were found to be significantly smaller in IS patients. These anatomical differences of different regions of the brain have been suggested to contribute to neurological defects detected in IS patients [114]. Defects in the paraspinal muscles have also been encountered among scoliotic patients [115-120]. Several studies suggested the involvement of the paraspinal muscles in the etiology and progress of IS [121-125]. Many authors showed that there is an increase in the myoelectric response near the apex of the curve on the convex side [125-129]. The interpretation of these observations is still unclear. However, it has been proposed that IS is caused by a neuromuscular dysfunction and an asymmetry of the paraspinal muscles due to changes of the

23

motor drive of the spinal cord produced by a change in the sensory input or by central mechanism which affect the balance of forces in this area and lead to IS [98].

2.6.2   Biomechanics

Scoliotic curve progression is defined by spinal biomechanical changes [130-132]. The association between vertebral growth and IS has been extensively studied and there is extensive debate in the literature about it [133-136]. Many studies reported that IS patients are significantly taller than matched healthy controls [137-140]. It has been suggested that this deviation of growth pattern is responsible for IS because of the higher tendency of taller spines under loading to buckle out of the sagittal plane [132, 141, 142]. However, this theory has not been proved [98]. Biomechanical studies have used finite element models which are patient-specific to study the effect of biomechanical factors on the progression of IS curve. Gravity, combined with anterior spinal overgrowth were shown as aggravating factors for thoraco-lumbar scoliosis [113]. Curve progression was more enhanced by reduction of the mechanical stiffness of the disc in addition to the aforementioned factors. These disc changes could be secondarily to a primarily existing deformity and hence affecting the curve severity rather than being a primarily cause [143]. The soft tissues responsible for sustaining the stability of the spine could also play a role in the etiopathogenesis of IS. Of note, 82% of IS patients were reported to have high level of distortion of the ligamentum flavum elastic fibers when compared with controls. Moreover, 23% of IS patients showed dysfunctional metabolism of fibrillin which suggests that the elastic fiber system might play a role in the pathogenesis of IS[144].

2.6.3 Biochemistry

All the biochemical factors that have been reported in the literature and suggested to contribute to the pathogenesis of IS, are reviewed in the review in the Annex "Biochemistry of Idiopathic Scoliosis from discovery to diagnostic biomarkers": Current scientific literature reveals potential biomarkers of the onset and progression of IS, which include hormones, systemic factors, hematological factors, and bone metabolism factors. Hormones that include insulin like growth factor-1, melatonin, estrogen, ghrelin, and leptin, have been implicated as potential diagnostic biomarkers in IS [145-149]. Systemic factors, such as osteopontin and Gi proteins; proteins involved in bone metabolism, such as matrilin-1, cartilage oligomeric matrix protein (COMP), osteocalcin; and the hematological protein calmodulin, are the other potential candidates for diagnosing IS [150-154]. These diagnostic biomarkers do not all act in isolation but are interconnected through signaling pathways. There is crosstalk between melatonin and estrogen; estrogen and growth hormones; or between melatonin and calmodulin [155, 156]. Similarly, increased levels of osteopontin regulate Gi protein-coupled receptor (GiPCR) dysfunction [157]. Other studies implicate single nucleotide polymorphisms (SNPs) in calmodulin, melatonin receptors, or matrilin-1, as potential risk factors of increasing the spinal curvature [158-160]. While theories, such as the double neuro-osseous theory implicate increased sensitivity of the hypothalamus to leptin in the development of IS, the downregulation of leptin has also been indicated in the etiopathogenesis of IS [149, 161]. Although data on the potential candidates for diagnosis appears promising, it is acknowledged that replication studies with more cases should generate a sufficient statistical power required to confirm the diagnostic potential of blood biomarkers.

2.6.4 Genetics

A genetic component of IS is well documented. Clinical studies have noted the familial incidence of scoliosis since the 1930s, where the inheritance of scoliosis in five generations was reported by Garland et al [162]. 97% of IS patients have positive familial history [113]. Segregation studies report an increased risk of IS among first degree relatives of affected individuals, and that the risk decreases as the degree of relation to the affected individual becomes more distant [163, 164]. Moreover, twin studies have shown that the rate of concordance of IS in monozygotic twins is almost double that of the dizygotic twins [165]. It has also been reported that the rate of curve progression is very similar in twins [165]. The mode of inheritance of IS is still unclear [166] with an apparent high level of heterogeneity between different families [167-169]. The genetic nature of the disease is presumed therefore to be complex. The incomplete penetrance and variable expressivity of IS suggest complex genetics with multiple genes and environmental factors involved in the disease [170, 171]. It has been reported that there is high level of inconsistency between the results of different genetic studies in IS [172]. A recent study conducted on a huge number of twins in the Swedish twin registry estimated that the genetic component of IS susceptibility is 38%, while the environmental component is 62% [17].

2.7 Biological endophenotypes associated with IS

The inconsistency between the results of different studies in IS may be at least partly due to the heterogeneity among IS patients both at the clinical level and the genetic level. The patients differ in their age of onset, type of curvature, degree of severity, rate of progression, sex and ethnic background. By studying homogeneous cohorts, chances are higher to find

common genetic background. Studies from our lab have classified IS patients into three biological endophenotypes named FG1, FG2, and FG3 (functional groups 1, 2 and 3) [151, 173]. This classification was based on a common pathological mechanism in which differential impairment of Gi-coupled receptor (G inhibitory protein-coupled receptor) signalling pathway resulting from selective inhibition of one or more of the isoforms of Giα subunit (Gi_1- Gi_2- Gi_3). These endophenotypes represent a unique heritable trait conserved among affected family members contrasting with the usual phenotypic heterogeneity, within the same family, in terms of curve type and severity. Moreover, some unaffected relatives of the patients can carry the same endophenotype conferring a high risk to those individuals[174]. The use of these endophenotypes allows patient stratification that has been correlated to specific molecular expression profiles and different prognosis [175]. In a complex genetic disease such as IS, those endophenotypes have potential utility to identify risk genes specific to each of these endophenotypes.

3.   Genetic studies in Idiopathic Scoliosis

During the last three decades, there has been huge number of studies trying to uncover genetic factors in IS. This task is difficult due to the complexity of the disease, as evidenced by the high level of inconsistency among different genomic studies. These studies are either family based and conducted with linkage analysis, or population (case/control) based and conducted with association studies. The general goal of these studies was either to identify specific loci in the genome or certain genes, which could be related to the disease or trying to explain the mode of inheritance of this disorder.

3.1 Family based studies and suggested modes of inheritance

3.1.1 Genome wide linkage studies

Extensive effort was made to identify specific loci in the genome, which might be linked to IS by genome-wide studies as well as to understand its mode of inheritance. Several regions in the genome have been suggested to be linked to IS through family-based studies. Those with significant statistics are listed in **Table III**. However, there is a general lack of replication of specific loci by different groups. As an exception to this, two loci were replicated by two different groups; chromosome 9q31-q34 which was introduced by Miller at al [176] and replicated by Ocaka et al [169], and chromosome 17q25 which was first identified by Ocaka et al [169] and confirmed by Clough et al [177]. Gao et al [178], reported linkage of the chromosomal region 8q12 and suggested *CDH7* as a candidate IS gene. Mutations in this gene lead to CHARGE syndrome in which a large percentage of patients have scoliosis. However, these results could not be replicated by another group [179]. The mode of inheritance for IS is inconsistent among the different studies. Of note, 65% of the studies reported AD to be the mode of inheritance for IS [180]. X-linked dominant mode of inheritance had also been suggested [166, 181]. Raggio et al suggested both AD and AR modes of inheritance in their study [182]. Recently a study suggested AD, AR and multifactorial modes of inheritance for IS [183].

Table III: Genomic regions that have been linked to Idiopathic Scoliosis (modified from[172]).

Locus	Suggested mode of inheritance	Statistics	Reference
17p11	AD	Zmax=3.2	Salehi et al.[184]
19p13.3	AD	LOD=4.48	Chan et al.[185]
Xq22.3-q27.2	XLD	LOD=2.23	Justice et al.[181]
9q31.2-q34.2	AD	Zmax=3.64	Ocaka et al.[169]
3q12.1	AD	Zmax=3.00	Edery et al.[168]
5q13.3		Zmax=3.01	
12p	AR	HLOD=3.2	Raggio et al.[182]
	AD	HLOD=3.7	
6q15-q21	—	P<2x10-10	Marosy et al.[186]
10q23-q25.3		P<2x10-7	

AD: autosomal dominant, XLD: X-linked dominant, AR: autosomal recessive

3.1.2 Candidate gene linkage studies

Based on the biological roles of some genes, several have been tested by linkage analysis and/or by transmission disequilibrium. A study demonstrated that an intragenic microsatellite polymorphism in the 3' region of the matrilin 1 gene (*MATN1*) was linked to IS[187]. Other genes tested by studies which fall in this category did not show linkage with

IS; fibrillin (*FBN1*), elastin (*ELN*), collagen I A1, collagen IA2, and collagen II A1 (*COL1A1*, *COL1A2, COL2A1*), Aggrecan (*ACAN*) and melatonin receptor 1A (*MTNR1A*) [188-192].

3.2   Population based studies

3.2.1   Genome wide association studies

Since 2011, a number of GWAS of IS have been published. The most important findings are listed in **Table IV**. A GWAS study in a Japanese cohort reported an association between a locus at 10q 24.31 and IS [42]. They reported common SNPs in the vicinity of the *LBX1* gene, of which the most significant was rs11190870. Several replication studies subsequently confirmed this association in both Asian and Caucasian populations [193-197]. A recent study showed that a risk allele of this SNP is associated with higher level of transcription of the promoter. Moreover in zebra fish high expression of *LBX1* leads to the deformation of normal body curvature [198]. Another candidate gene, *GPR126* (G-protein coupled receptor 126), has been reported by two studies to be associated with IS. The first study involved three populations; Japanese, Han Chinese and European [199]. The second study involved a Chinese population [200]. It has also been shown that the knockdown of this gene in zebra fish lead to delayed ossification of the spine [199]. The intergenic variant rs12946942 has been reported by Miyake et al [201]. This variant lies on chromosome 17 near the two genes *SOX9* and *KCNJ2*, mutations in which are associated with campomelic dysplasia and Andersen Tawil syndrome, for which scoliosis is an aspect of the phenotype. *BNC2* (basonuclin-2) and *PAX1* (paired box1) are another two interesting genes suggested by two GWAS [202, 203]. However, they both need to be confirmed by additional studies. *BNC2* codes for a zinc finger transcription factor and its overexpression was shown to induce

body curvature in zebra fish [202]. *PAX1* encodes a transcription factor which is involved in the development of the spine [203].

Table IV: The most important GWAS results for Idiopathic Scoliosis (modified from [204]).

SNP	Gene name	Statistics	References
rs11190870	*LBX1*(downstream)	p=1.2x10-19	Takahashi et al.[42], Fan et al. [193] Gao et al.[194], Jiang et al.[195], Liang et al.[196], Londono et al.[197], Grauers et al. [205]
rs657507	*GPR126* (intron)	p=1.6x10-4	Kou et al. [199], Xu at al.[200], Grauers et al.[205]
rs12946942	Intergenic	p= 4x10-8	Miyake et al.[201], Grauers et al.[205]
rs3904778	*BNC2*	p=2.5x10-13	Ogura et al.[202]
rs6137473	*PAX1*	p=2.2x10-10	Sharma et al.[203]

SNP: single nucleotide polymorphism

3.2.2   Candidate genes association studies

During the last two decades, a large number of associations of candidate genes have been explored. Gorman et al [172] have comprehensively reviewed those studies. They reported that the earlier candidate gene studies (from 1992 to 2006) were mostly family based linkage studies, while most recent candidate gene studies (after 2006) were case-control association studies, due to the low power of family based methods in detecting common variants in complex diseases. Genes, which have been reported for their positive associations with either IS susceptibility and/ or severity are summarized in **Table V**.

Table V: Genes which have been reported to be associated with Idiopathic Scoliosis (modified from [172]).

Gene name	SNPs	Statistics	References
MATN1	rs1149048	p=0.0034	Chen et al. [160]
MMP3	rs3025058	p=0.01	Aulisa et al. [206]
IL6	rs1800895	p=0.014 (G/C) p<0.001 (G/G)	Aulisa et al. [206]
VDR	rs1544410	p=0.0054	Suh et al.[207]
MTNR1B	rs4753426	p=0.045	Qiu et al. [159]
TPH1	rs10488682	p=0.002 (allele) p=0.001 (genotype)	Wang et al. [208]
ESR1	rs9340799	p=0.002 p=0.01	Inoue et al [209] Wu et al. [210]
ESR2	rs1256120	p=0.037 (susceptibility) p= 0.005 (severity)	Zhang et al. [211]
GPER	rs3808351 rs10269151 rs4266553	p=0.004 p=0.048 p=0.028	Peng et al. [212]
IGF1	rs5742612	p=0.042	Yeung et al.[213]
IL-17RC	rs708567	p=0.023 (genotype) p=0.028 (allele)	Zhou et al.[214]
DOT1L	rs12459350	p=0.001	Mao et al.[215]
C17orf67	rs4794665	p=0.006	Mao et al. [215]
TGFB1	rs1800469	p<0.01	Ryzhkov et al.[216]

3.3 Sequencing studies

Because sequencing technologies were very expensive in the past and only very recently have become more affordable, few studies have attempted to detect rare causal variants in IS and this field of research is still in its infancy. Sequencing using either whole exome or targeted gene panels has identified a few genes, which might contribute to the occurrence and/or severity of scoliosis. Buchan et al documented an enrichment of rare potentially protein-altering variants in the *FBN1* and *FBN2* genes in severe IS cases versus controls, by performing WES (whole exome sequencing) in 91 severe cases and 337 controls, followed by targeted sequencing of these two genes in a larger cohort [217]. *HSPG2*, a gene coding for an extracellular matrix protein, was identified by Baschal et al [218]. In their study they reported several rare potentially damaging variants by performing WES for three patients in a multigenerational family followed by sequencing this gene in an independent cohort of 100 IS patients. *POC5*, which encodes for a centriolar protein, was identified by Patten at al. In this study they performed exome sequencing for three patients in a multigenerational family as a follow up study to another LA study to the same family and suggested linkage to two loci 3q12.1 and 5q13.3. Three rare variants in *POC5* were reported and were shown to cause spinal deformity in zebrafish [219]. Haller et al analyzed exome sequence data of 391 severe IS cases and 843 controls. In a genome-wide gene burden test no individual gene achieved statistical significance, therefore they further collapsed genes according to gene ontology pathways and observed excess variation among genes implicated in the extracellular matrix, particularly collagen genes [220]. More recently in a study where

WES was performed for a multiplex family, a variant in the gene *AKAP2* was found to be co-segregating with the phenotype [221].

3.4 Genome wide expression profiling studies

Genome wide expression profiling methods are those, which detect all the transcriptionally active genes and the relative amounts of their RNAs in the tested cells or tissues. While sequencing a genome shows the alterations that might lead to an effect, expression profiling shows the actual transcriptional state of the cell or tissue in a given state. DNA microarrays were usually used in these analyses, wherein the extracted RNAs from biological samples after processing, are allowed to hybridize to their probes fixed on a solid surface, and hence genes which are expressed can be determined [222]. The probes can include subsets of biological candidate genes, or ultimately all known protein-coding or otherwise transcribed genes. This method allows for the detection of genes whose expression changes with a disease state, and that might therefore play a role in the pathophysiology of the disease. Few studies have performed this approach in IS. Using genome wide expression profiling, Fendri et al found 145 genes differentially expressed in primary osteoblasts from vertebrae of IS patients compared to healthy controls. The most significant genes were Homeobox genes (*HOXB8, HOXB7, HOXA13, HOXA10), ZIC2, FAM101A, COMP* and *PITX1.*Clustering analysis showed that these genes belong to biological pathways important for the development of bones specifically of skeletal elements, differentiation and the structural integrity of vertebrae [223]. Another gene expression study reported the differential expression profiles of genes of the bone marrow mesenchymal stem cells and related

pathways in IS patients [224]. For expression studies in general, DNA microarrays have mostly been replaced by high throughput sequencing technologies (RNA-Seq).

3.5 Summary

In sum, the estimated heritability of IS is 38%, whereas common known variants explain less than 2% of the genetic component of the disease. On the other hand, few sequencing studies have been performed which have identified a few potential candidate genes, so this field of research is still in its infancy. Hence, according to the current state of knowledge, the genetic architecture of IS is very incompletely understood, and there is a room for much additional discovery work. New genes may identify new pathways or clarify already identified pathways in terms of modulating them to improve treatment of patients at various stages of disease progression.

4. Objectives

The main objective of our thesis is to gain more knowledge and deeper insight in the genetic and biochemical basis of IS. To identify the genetic determinants of IS we employed multiple genetic approaches.

The first approach is exome sequencing and targeted sequencing aiming to detect genes enriched with rare variants in IS.

The second approach is a GWAS approach where the goal of our analysis is to test whether the *LBX1* gene locus is associated with French-Canadian cohort. *LBX1* locus is so far the most solidly confirmed locus for its association with IS in different populations, but it has never been tested in French-Canadian population.

The third approach is a candidate gene approach. We chose to study the gene *CHI3L1,* encoding for the secreted factor YKL-40, which has been identified in a previous study in our lab (Gorman et al, manuscript submitted), in which a genome wide expression profiling method has shown that *CHI3L1* is significantly over expressed in a subgroup of IS patients (biological endophenotype FG1), compared to the two other groups (FG2 and FG3) which are more prone to develop a severe scoliosis. We attempted to elucidate the biochemical correlates of circulating YKL-40 levels and by combining it with the analysis of *CHI3L1* gene variants, with the risk of spinal deformity progression in the context of IS.

CHAPTER II

RESULTS

1. Article 1

Rare variants in the *FAT3* gene are associated with familial and isolated Idiopathic Scoliosis

Dina Nada[1,2], Cédric Julien[1], Simon Papillon[3], Jacek Majewski[3], Mark E. Samuels[4,5], Alain Moreau[1,2,6,7, *]

[1]Viscogliosi Laboratory in Molecular Genetics of Musculoskeletal Diseases, Centre de recherche du CHU Sainte-Justine, Montreal, QC, Canada;

[2]Program of Biomedical Sciences, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada;

[3]Department of Human Genetics, McGill University, Montreal, QC, Canada;

[4]Centre de Recherche du CHU Ste-Justine, Montreal, QC, Canada;

[5]Department of Medicine, Université de Montréal, Montreal, QC, Canada;

[6]Department of Biochemistry and Molecular Medicine, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada;

[7]Department of Stomatology, Faculty of Dentistry, Université de Montréal, Montreal, QC, Canada.

* Correspondence to: Alain Moreau PhD, Viscogliosi Laboratory in Molecular Genetics of Musculoskeletal Diseases, Centre de recherche du CHU Sainte-Justine (room 2.17.027), 3175 Chemin de la Côte-Ste-Catherine, Montreal, H3T 1C5, QC, Canada;

E-mail: alain.moreau@recherche-ste-justine.qc.ca

This study was approved by the institutional review boards of Sainte-Justine University Hospital, Montreal Children's Hospital, Shriners Hospital for Children in Montreal and McGill University, as well as the Affluent and Montreal English School Boards

Contribution of authors:

Dina Nada: Participated in design of the study, performed experimental work, participated in bioinformatics and statistical analyses of data, and as lead author of the manuscript.

Cédric Julien: Participated in design of the study, bioinformatics and statistical analyses of data.

Simon Papillon: Performed bioinformatics analyses to confirm gene structure.

Jacek Majewski: Supervised bioinformatics analyses to confirm gene structure.

Mark E Samuels: Participated in and supervised genomics experiments and participated in writing the manuscript.

Alain Moreau: Supervised and supported all aspects of the study, and participated in writing the manuscript.

Rare variants in the *FAT3* gene are associated with familial and isolated Idiopathic Scoliosis

Dina Nada[1,2], Cédric Julien[1], Simon Papillon[3], Jacek Majewski[3], Mark E. Samuels[4,5], Alain Moreau[1,2,6,7, *]

[1]Viscogliosi Laboratory in Molecular Genetics of Musculoskeletal Diseases, Sainte-Justine University Hospital Research Center, Montreal, QC, Canada;

[2]Program of Biomedical Sciences, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada;

[3]Department of Human Genetics, McGill University, Montreal, QC, Canada;

[4]Sainte-Justine University Hospital Research Center, Montreal, QC, Canada;

[5]Department of Medicine, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada;

[6]Department of Biochemistry and Molecular Medicine, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada;

[7]Department of Stomatology, Faculty of Dentistry, Université de Montréal, Montreal, QC, Canada.

* Correspondence to: Alain Moreau PhD, Viscogliosi Laboratory in Molecular Genetics of Musculoskeletal Diseases, Sainte-Justine University Hospital Research Center (room 2.17.027), 3175 Chemin de la Côte-Ste-Catherine, Montreal, H3T 1C5, QC, Canada;

E-mail: alain.moreau@recherche-ste-justine.qc.ca

Word count: 3996

Abstract

Whole exome sequencing was performed on a cohort of French-Canadian patients with idiopathic scoliosis, and analyzed in a collapsing gene burden test for rare protein-altering variants using a case control statistics. No single gene achieved statistical significance, therefore targeted exon sequencing was performed for 24 genes with the smallest p-values in an independent replication cohort of severely affected females. One gene, *FAT3*, achieved statistical significance in the replication cohort using a similar collapsing gene burden test, with multiple different rare variants in the gene present in the replication samples. Independently, we sequenced the exomes of all members of a rare multiplex family having three affected sisters and unaffected parents. All three sisters were compound heterozygous for two rare protein-altering variants in *FAT3*. The parents were singly heterozygous for each variant. The two variants in the family were also present in the case control cohort. *FAT3* is expressed in primary osteoblasts, but no significant difference was found in the level of RNA expression in cells from scoliotic patients versus controls. Our results strongly implicate *FAT3* as a new genetic factor in the etiology of idiopathic scoliosis.

Introduction

Idiopathic Scoliosis (IS) is a common complex disorder of the spine. It is a three-dimensional deformity of the skeleton characterized by a lateral curvature of ≥10° on a standing radiograph (Cobb method), combined with vertebral rotation. It is the most common form of spinal disorder. It mostly occurs at the age of adolescence and affects1-4% [1] of the global pediatric population with higher prevalence in females who are generally more severely affected than males [2].In most cases the underlying cause of idiopathic scoliosis is unknown, although a genetic component is well recognized [3 4]. Twin and family studies have documented high rates of concordance among twins and increased risk to relatives of IS patients [5, 6]. The mode of inheritance is still unclear [7]. The genetic nature of the disease is complex, with an apparent high level of heterogeneity between different families [8-10]. A number of candidate genes and loci have been suggested by different studies, but few have been successfully replicated [11]. Human genetic studies have used both linkage and association methods. The results of linkage studies have been poorly reproducible [11]. Genome wide association studies (GWAS) have identified several candidate genes for IS susceptibility including *CHL1*, *LBX1*, *GPR126*, *BNC2*, and *PAX1* [12-16]. The associated common single nucleotide polymorphisms (SNPs) identified to date only explain a small portion of the genetic component of the disease. Genetic interactions [17] and rare variants [18] might well explain this "missing heritability" [19].

Few studies have attempted to detect rare causal variants in IS and this field of research is still in its infancy. Sequencing using either whole exome or targeted gene panels, has identified several genes which might contribute to the occurrence and or severity of scoliosis; such as *FBN1*, *FBN2* [20], *HSPG2* [21], *POC5* [22] and *AKAP2* [23]. Another study suggested that accumulation of rare variants in a group of genes of the extracellular matrix might contribute to disease risk [24]. In sum, the genetic component of idiopathic scoliosis is incompletely understood, leaving significant room for further research.

In this study we performed whole exome sequencing (WES) in a French-Canadian IS cohort, followed by a second phase of targeted sequencing of the 24 best candidate genes in a replication cohort. In parallel, we performed WES in a unique multiplex family of three affected sisters with healthy parents. Our goal was to find new genes enriched with rare variants, which might contribute to the disease. Our results implicate a novel gene, *FAT3*, not previously associated with IS, as a strong candidate for this condition.

Materials and methods

Ethics approval

This study was approved by the institutional review boards of Sainte-Justine University Hospital, The Montreal Children's Hospital, The Shriners Hospital for Children, and McGill University, as well as The Affluent and Montreal English School Boards. Written informed consents were given by parents or legal guardians and assents were given by minors.

Subjects

All patients with IS were examined by orthopedic surgeons from the three pediatric centers participating to this study. A diagnosis of IS required both history and physical examination with a minimum curvature in the coronal plane of 10°, showed by a standing posteroanterior spinal radiograph, by the Cobb method with vertebral rotation and without any congenital or genetic disorder. Healthy children were recruited from Montreal's schools and examined by one orthopedic surgeon.

First French-Canadian IS cohort (discovery cohort)

We selected seventy-three unrelated IS cases and seventy age- and gender-matched healthy controls; all were of French-Canadian ancestry. Fifty of the cases were severe (Cobb angle ≥ 40) and twenty three were moderate (Cobb angle < 40).

Second French-Canadian IS cohort (replication cohort)

Ninety-six patients of French-Canadian origin were selected for the replication study, unrelated to each other or to the cases in the discovery cohort. Since 93% of the initial cohort were females and 68%were severe cases, the second cohort were chosen to be all females and severely affected. Thirty-six healthy French-Canadian females were recruited from Montreal's schools, plus an additional sixty French-Canadian females from the CARTAGENE project [25 26].

French-Canadian multiplex family

A rare multiplex French-Canadian family with three affected sisters and healthy parents was ascertained and analyzed by WES analysis. The proband was diagnosed at the age of 13 years old with IS with a right lumbar curve and Cobb angle measuring 15°. Her first sister was diagnosed with IS with a left lumbar curve measuring 23° and the second sister was also diagnosed with IS with right thoracic curve measuring 13°.

DNA extraction

Blood was obtained by standard venipuncture. Genomic DNA was extracted from peripheral leukocytes using PureLink®genomic DNA kit (Thermo Fisher Scientific, Waltham, Massachusetts, USA).

Exome and targeted sequencing

Whole-exome sequencing (WES) for a discovery French-Canadian IS cohort

Exome capture was performed using Agilent SureSelect[XT] Human All Exon 50 Mb v3 according the manufacturer's recommendations. Sequencing was done using SOLiD 5500xl from Applied Biosystems

by Life technologies at the Sainte-Justine University Hospital genomic platform.The average coverage of targeted sites was approximately 100X (Supplementary Materials and Methods S1).

Targeted sequencing for selected genes in a replication French-Canadian IS cohort

Twenty-four genes were chosen for resequencing in a second French-Canadian cohort. Enrichment of coding exons of these genes was done using Roche Nimblegen EZ Choice custom baits, with bar code multiplexing of 96 samples per lane of sequencing. Sequencing was done on an Illumina HiSeq 2000 at the McGill University and Genome Quebec Innovation Centre (MUGQIC). The average coverage of targeted sites was approximately 400X (Supplementary Materials and Methods S1).

WES for a French-Canadian family

Exome capture for the multiplex family was performed using Agilent SureSelect^{XT} Human All Exon 50 Mb v3 according the manufacturer's recommendations. Sequencing was done on an Illumina HiSeq2500 at the Sainte-Justine University Hospital genomic platform (Supplementary Materials and Methods S1).

Bioinformatics

Specifics of our bioinformatics analysis, the pipeline and subsequent variant filters are included in the Supplementary Materials and Methods S1. Only protein coding, and near intronic regions were analyzed. Our analysis included SNPs, and small insertions or deletions (indels). The SIFT (Sorting Intolerant from Tolerant) [27], PolyPhen-2 (Polymorphism Phenotyping v2) [28] and MutationTaster2 [29] algorithms were used to predict possible impact of amino acid substitutions on the structure and function of a human FAT3 protein in IS patients harboring different *FAT3* gene variants.

Sanger sequencing

Sanger sequencing was performed at the Genome Quebec Innovation Centre at McGill University. Primers were designed using the program Primer3. Sanger sequence chromatograms were analyzed using Mutation Surveyor (Soft Genetics, Inc.). Exons 25 and 26 of *FAT3* were not initially sequenced in the replication cohort because the custom baits that were used to capture the selected genes for sequencing were designed according to the RefSeq gene model, which did not include those 2 alternative exons. Hence, we performed Sanger sequencing for the 2 additional exons in 72 patients of the replication cohort. DNA of the other patients was not available. Numbering of variants in *FAT3* is based on NCBI reference sequence entries NM_001008781.2 and NP_001008781.2.

Statistical Analysis

In both phases of the case/control study, we employed a collapsing gene burden test for significance, under the assumption that all rare, potentially protein-altering variants act in the same phenotypic direction with the same magnitude, independent of specific allele frequencies. In the instances where an individual carried two rare variants in the same gene, this was counted as a single event. In the first, whole exome discovery phase, chi-square p-values were calculated to compare the accumulation of rare variants (MAF < 0.01) in genes throughout the exome in patients versus controls, assuming a threshold of $p=6\times10^{-6}$ (0.05/8150), based on the number of genes harboring at least one rare variant among either the cases or controls in the exome data set. In the second, targeted gene phase (24 selected genes), Fisher's exact test was used to calculate the p-values for comparisons between patients and controls where $p=2\times10^{-3}$ (0.05/24) was considered significant.

Validation of *FAT3* gene structure

The gene model for *FAT3* used by RefSeq does not appear to be supported by long individual human cDNA clones, and seems to be based on homology to several long rodent cDNAs. Therefore, to confirm the gene structure, we analyzed in-house brain RNA-Seq data and WGBS data (Whole Genome Bisulphite) from an unrelated individual not part of our cohorts, as well as from GENCODE public annotations. We also profiled *FAT3* expression using GTExTranscriptome Portal (http://www.gtexportal.org/home/gene/FAT3).

Cell culture and RNA extraction

Primary osteoblasts were derived from bone specimens that were obtained from IS patients intraoperatively and from traumatic individuals unaffected by IS as controls. Briefly, cells were grown in culture dishes 10cm^2 in Alpha Modification of Eagle's Medium (αMEM) containing 10% fetal bovine serum (FBS, Hyclone) and 1% penicillin/ streptomycin (antibiotic/antimycotic, Invitrogen) at 37°C and 5% CO_2. Cells were grown until they reached confluency. Then the cells were washed by phosphate-buffered saline (PBS 1x) twice and were treated with 1ml TRIzol, lysed and transferred to 1.5 ml tube and stored at -80°C. RNA was extracted using TRIzol (Life Technologies), following the manufacturer's instructions.

Quantitative RT-Polymerase Chain Reaction (qRT-PCR)

Expression analyses by qRT-PCR were done in triplicate using *GAPDH* and *PPIA* (Peptidylprolyl isomerase A) as normalizing housekeeping genes (Supplementary Materials S1).

Results

Identification of twenty-four candidate genes from WES of the discovery cohort

We studied an initial discovery cohort of 73 unrelated IS patients (68 females and 5 males), with 70 age- and gender-matched controls, all of French-Canadian ancestry. Fifty of the patients were considered severely affected as their Cobb angles were at least 40°, and the other twenty-three were considered as moderate cases. We performed WES for this cohort, followed by variant annotation and filtering to identify rare variants contributing to IS. Because the frequency of individual rare variants in the cohort was too low to yield sufficient statistical power, we looked for genes harboring an overall excess of rare variants in the discovery patient cohort. We performed a collapsing gene burden test, in which we compared the enrichment of rare variants per gene in patients versus controls. To define rare variants we used a minor allele frequency (MAF<1%) as an initial cutoff, and a MAF <0.5% as a more stringent cutoff according to 1000 Genomes European ancestry (EUR) and the Exome Sequencing Project European ancestry (ESP-EA). Due to the total cohort size and technical limits of SOLiD sequencing, only 8150 genes harbored at least one such rare variant among all case and control samples. Therefore, we set a threshold for statistical significance at $0.05/8150 = 6 \times 10^{-6}$. None of the 8150 genes met the required p-value threshold. We therefore selected the 24 best candidate genes for follow up validation in a separate replication cohort (Table 1). For our selection of candidate genes, we took into consideration both the p-value and the absolute number of patients and controls who carried rare variants.

Table 1.Genes selected from the discovery cohort with SNPs of MAF < 1%

Gene	Cases with non-synonymous SNPs	Controls with non-synonymous SNPs	Uncorrected Fisher-exact p-value
GLP1R	10	0	0.001
DMRT3	10	1	0.009
ITGA8	7	0	0.014
A1CF	7	0	0.014
GPR179	14	4	0.022
FAT3	12	3	0.027
CEACAM18	6	0	0.028
TTC21A	6	0	0.028
NFRKB	6	0	0.028
GMPR2	6	0	0.028
SLC3A1	6	0	0.028
IMMT	6	0	0.028
ZNF189	6	0	0.028
CD1B	10	2	0.031
SEC16A	10	2	0.031
CCDC50	8	1	0.034
SLC22A16	12	4	0.062
R3HCC1L	7	1	0.063
IL16	7	1	0.063
HPS4	6	1	0.116
DPEP3	6	1	0.116
LY75	6	1	0.116
TENM3	6	1	0.116
ITGA4	4	0	0.120

Targeted sequencing of the selected twenty-four genes in a replication cohort

The replication cohort was chosen to be more homogeneous; all cases were severely affected females. For comparison, in our initial discovery cohort, 93% of IS patients were females and only 68% were severe cases. Our replication cohort comprised 96 patients and 96 gender matched controls. Coding exons of the 24 candidate genes were sequenced in the 192 replication samples, using a custom capture library. After calling and annotating variants, we filtered to remove poor quality calls and variants with a MAF greater than 1% (according to 1000 Genomes (EUR) and ESP-EA). For the replication study, we again employed a collapsing gene burden test. Of the twenty-four candidate genes, only one gene; *FAT3*, showed significant enrichment for rare variants in the IS patients (Table 2). Specifically, in *FAT3*, 21 of the replication IS cases harboured rare protein altering variants versus only 2 of the controls (Odds ratio (OR) 13.16, p-value=2.38×10^{-5}, Fisher-exact test) (Table 2). Interestingly when we included the synonymous variants, the p-value became more significant (OR 8.27, p-value=3.23×10^{-6}) (Table 3). We repeated the same analysis including variants with a MAF up to 2%, in which case the significance although still high was reduced (see Supplementary Tables S1 and S2). None of the other twenty three genes had p-values approaching the required threshold. The non-synonymous variants in *FAT3* in both the discovery and replication cohorts were distributed across much of the protein (Figure 1A).

Table 2. Statistical analysis for all selected genes and SNPs in the replication cohort with SNPs with MAF ≤ 1% (after removing the synonymous SNPs)

GENE	Non-synonymous rare SNPs	Cases with non-synonymous rare SNPs	Controls with non-synonymous rare SNPs	Un corrected Fisher-exact p-value
FAT3	17	21	2	2.38E-05
HPS4	13	8	2	0.100
DMRT3	3	4	0	0.121
IL16	7	7	2	0.169
ITGA8	6	5	1	0.211
A1CF	4	4	1	0.368
NFRKB	4	4	1	0.368
TENM3	6	4	1	0.368
SEC16A	5	8	4	0.372
LY75	8	9	5	0.406
R3HCC1L	7	10	6	0.434
CD1B	2	2	5	0.444
TTC21A	3	8	12	0.479
CCDC50	2	2	0	0.497
SLC22A16	2	2	0	0.497
GPR179	11	19	15	0.571
CEACAM18	1	3	1	0.621
ITGA4	2	3	1	0.621
DPEP3	3	6	4	0.747
GLP1R	3	6	4	0.747
GMPR2	0	0	0	1.000
IMMT	0	0	0	1.000
SLC3A1	1	1	1	1.000
ZNF189	0	0	0	1.000

Table 3. Statistical analysis for all selected genes and SNPs in the replication cohort with SNPs with MAF ≤ 1% (including the synonymous SNPs)

GENE	Rare SNPs	Cases with rare SNPs	Controls with rare SNPs	Un corrected Fisher-exact p-value
FAT3	26	30	5	3.23E-06
TENM3	16	19	8	0.036
A1CF	9	11	3	0.049
ITGA8	9	9	2	0.058
SEC16A	12	13	5	0.081
HPS4	15	8	2	0.100
IL16	11	11	4	0.104
NFRKB	7	7	2	0.169
DMRT3	4	5	1	0.211
LY75	10	10	5	0.282
CEACAM18	2	4	1	0.368
ITGA4	3	4	1	0.368
GPR179	12	20	15	0.455
R3HCC1L	8	12	8	0.479
CCDC50	2	2	0	0.497
DPEP3	4	7	4	0.537
CD1B	3	3	5	0.721
GLP1R	4	7	5	0.767
GMPR2	0	0	0	1.000
IMMT	3	6	5	1.000
SLC3A1	5	6	5	1.000
SLC22A16	3	3	2	1.000
TTC21A	5	14	13	1.000
ZNF189	1	1	0	1.000

Exome sequencing of an independent IS family

In parallel, we studied a very rare multiplex family in which three sisters were affected with IS, while the parents were unaffected (Fig. 1). This family was not included in our previous cohorts. We performed WES for all 5 members of this family. As in the case control exome and candidate gene sequencing, we restricted analysis to rare (MAF ≤ 1%), potentially protein altering SNPs or small indels. We analyzed the exome data with different inheritance models, given the unaffected status of the parents. First, we considered a dominant *de novo* mutation model in which the three sisters would share a heterozygous variant absent in the parents. Second, we considered a recessive model; either homozygous variant in the three sisters which is heterozygous in the parents or compound heterozygous in which the three sisters have two heterozygous variants in the same gene, each coming from one parent. No genes were found consistent with the *de novo* dominant or recessive homozygous models. However, the presence of compound heterozygous variants in *FAT3* was found consistent with that recessive model. Of note, the selection of candidate genes from the WES cohort was done before we performed the family study and the selection of the targeted resequencing genes was unbiased. The two variants of *FAT3* gene in the multiplex family are non-synonymous: p.L517S and p.L4544F (Fig 1). The first variant was present in four cases and one control in the replication cohort, and the second variant was present in one case in the discovery cohort. Both variants in *FAT3* were confirmed by Sanger sequencing of DNA from all members of the family (Fig 1D).

Validation of *FAT3* gene structure and identification of a novel unannotated exon

The gene model for *FAT3* used by RefSeq appears to be supported only by long individual rodent cDNA clones in the NCBI database, whereas there are only fragmentary human cDNA clones documented in the public genome browsers. Therefore to confirm the human *FAT3* gene structure, we analyzed our in-house brain RNA-Seq data and WGBS data for one individual. Our results were consistent with the RefSeq gene model (NM_001008781.2) with two exceptions. Just upstream of the 3' terminal exon we found evidence for two alternative exons which were either included or excluded together in various RNA-Seq reads. The two exons are also annotated by the GENCODE project website (version 24). In addition, we identified a previously uncharacterized exon located 125kb upstream of the first annotated exon, supported by multiple individual reads splicing this sequence to the second (but first protein-coding) exon (Supplementary Fig S1). This novel exon lies in a hypomethylated CpG island, a feature that is characteristic of active promoters (Supplementary Fig S2). Because the 5'-most exon annotated by GENCODE (exon 2 of our gene model) begins precisely at the splice acceptor junction, we suspect that the GENCODE raw data probably included exon 1 in some junction reads, which were not aligned to the genome across exon 1 due to the very long first intron. We also profiled *FAT3* expression using GTEx Transcriptome Portal (http://www.gtexportal.org/home/gene/FAT3) and observed a strong enrichment in brain and artery tissues (Supplementary Fig S3).

Sanger sequencing of exons 25 and 26 of *FAT3*

The alternative exons 25 and 26 were not captured in the replication capture sequencing because they are not annotated by RefSeq. Hence, we performed direct Sanger sequencing for these two exons in 72 cases of the replication cohort (DNA was not available for the rest of the cases). No rare potentially

protein-altering variants were observed among the sequenced cases for either of these two (very small) exons (data not shown).

Consequences of the *FAT3* rare variants

We identified in IS patients 26 non-synonymous SNVs (18 previously reported in public databases and 8 novel) in *FAT3* gene (Table 4). Prediction of the functional consequences of the non-synonymous SNVs was performed using three different algorithms SFIT, PolyPhen-2 and MutationTaster2. Of note, two variants were predicted as likely pathogenic by all three software tools, 13variants as likely pathogenic by two of the three algorithms, and one variant is a frame shift mutation (Table 4). To test if these rare variants affect the expression of *FAT3*, we performed a qPCR expression analysis using RNA extracted from primary osteoblasts obtained from seven scoliotic patients who had rare variants in *FAT3* from the discovery and replication cohorts, and seven controls (traumatic patients who did not have scoliosis from whom we could extract osteoblasts). No statistically significant difference in averaged *FAT3* expression was observed between both groups (Supplementary Fig S4).

Table 4. Prediction of *FAT3* variant effects on the function of the protein.

Reference position *	Mutation DNA level (hg19)	Mutation protein level	SNP ID	SIFT[a]	POLYPHEN-2[b]	MUTATION TASTER[c]	MAF 1000 Genome (EUR) %	MAF ESP (EA) %
1	chr11:92086828T>C	p.L517S	rs139595720	T	D	N	1	0.83
2	chr11:92087676A>T	p.N800Y	rs188857169	D	D	D	0	0.012
3	chr11:92087959G>A	p.R894Q	rs80293525	T	D	D	0	0.472
4	chr11:92088151T>C	p.L958P	rs76869520	D	D	D	0	0.036
5	chr11:92532013A>G	p.N1945S	novel	T	B	D		
6	chr11:92532651A>G	p.I2158V	novel	T	B	N		
7	chr11:92533254C>T	p.H2359Y	rs80046666	T	B	N	0	0.012
8	chr11:92533405G>A	p.R2409Q	novel	T	D	D		
9	chr11:92533555A>G	p.Q2459R	rs118056487	T	D	D	0	0.574
10	chr11:92533558G>A	p.R2460Q	rs200944979	T	D	D	0	0.1
11	chr11:92534695T>C	p.I2839T	rs200241295	T	B	N	0	0.048
12	chr11:92565003C>A	p.P3233T	novel	T	D	D		
13	chr11:92569867C>T	p.R3408W	rs200404766	D	D	N	0	0.255
14	chr11:92570856G>T	p.A3418S	rs201449521	D	B	D	1	0.252
15	chr11:92573811CT>C	p.S3485fs	novel					
16	chr11:92577352G>A	p.A3607T	rs200032318	T	D	D	0	0.496
17	chr11:92577469C>A	p.Q3646K	novel	T	B	D		
18	chr11:92577590G>A	p.R3686H	rs138237129	T	B	N	0	0.117
19	chr11:92577659G>T	p.S3709I	rs75081660	T	B	D	1	1.36
20	chr11:92613978G>A	p.R4070Q	rs201379307	T	B	D	0	0.036
21	chr11:92616191C>T	p.T4190M	rs186899262	T	D	D	0	
22	chr11:92616217G>A	p.A4199T	rs201053443	T	D	D	0	0
23	chr11:92620226A>G	p.N4333S	novel	T	D	D		
24	chr11:92623798G>A	p.G4398D	rs142403035	D	P	D	0	0.25
25	chr11:92624166T>C	p.C4521R	novel	D	B	D		
26	chr11:92624235C>T	p.L4544F	rs187159256	D	B	N	0	0.272

[a]SIFT : D, damaging, T, tolerated.
[b]POLYPHEN-2: D, probably damaging, P, possibly damaging, B, benign.
[c]MUTATIONTASTER: D, disease causing, N, polymorphism.
 * Position of each variant of FAT3 protein is illustrated in Figure 1 panel A.

Discussion

Using whole-exome sequencing with a combined two-stage case/control and multiplex family model, we discovered a novel association between the gene *FAT3* and idiopathic scoliosis (IS). Most other studies have looked for individual rare variants in families [23], rather than collapsing these variants by gene. *POC5* and *HSPG2* were initially identified from such familial studies, and only then were further investigated in independent cohorts [21 22]. Only two studies employed a similar approach to ours, looking at rare variant burden at the gene level. Buchan et al [20] with a two-stage approach beginning with whole exome sequencing of 91severe IS cases and a collapsing gene burden test [30], followed by targeted gene resequencing in a second, much larger cohort. As in our study, no single gene achieved genome-wide significance, but the gene with the best p-value, *FBN1*, was followed up in a replication cohort similar to our approach and replicated together with the related gene *FBN2*. In the second study, Haller et al analyzed exome sequence data of 391 severe IS cases and 843 controls. Again, in a genome-wide gene burden test no individual gene achieved statistical significance, therefore they further collapsed genes according to gene ontology pathways and observed excess variation among genes implicated in the extracellular matrix, particularly collagen genes [24]. It is unclear whether *FAT3* would have been included among this group. No collagen or fibrillin genes were among the 24 candidates in our replication cohort.

Idiopathic scoliosis is a highly heterogeneous disease in terms of both phenotype and etiology, therefore finding a common genetic background in isolated cases is challenging. Several of the individual rare variants we observed in our case cohort were recurrent, suggestive of at least a modest founder effect. This is consistent with the elevated incidence of scoliosis in Quebec, given that our cohort was almost completely of French-Canadian ancestry. Nonetheless, there were a relatively large number of different rare variants in our cases versus matched controls. Our success in identifying an

association to *FAT3* with this strategy may also have relied on a very homogeneous phenotype definition in terms of sex and severity. It is also worth noting that our controls are not random population controls, but are effectively discordant since they are of individuals whose physical exam and the lack of family antecedents excludes a diagnosis of idiopathic scoliosis or related spinal disorders. It would be interesting to revisit the total variant data sets from the previous population studies [20 24] with respect to *FAT3*; however those data are not available to us. More generally, our results indicate that two-stage approaches for rare variant detection in common complex diseases can yield good gene candidates for further study, even without additional criteria relying on previously known biology of the disease.

Interestingly, *FAT3* is near another gene *MTNR1B* melatonin receptor 1B, in which a polymorphism has been associated with IS [31]. The SNP in question, rs4753426, lies slightly proximal to the 5' end of *MTNR1B*, and about 72kbdistal to the 3' end of *FAT3*. We speculate that the observed association may be functionally related to *FAT3* rather than *MTNR1B* function, especially as the association is strongest in Asian populations where there is typically more extended linkage disequilibrium.

FAT3 is a member of the FAT gene family comprised of *FAT1*, *FAT2*, *FAT3* and *FAT4*, all of which are members of the cadherin superfamily homologous to the *Drosophila* gene *Fat* [32] regulating planar cell polarity (PCP) in the Drosophila wing [33]. Members of the FAT cadherin subfamily have conserved structures from flies to vertebrates [34]. *FAT3* contains multiple repeats of a cadherin repeat domain (involved in Ca^{+2} binding), a single laminin G domain and three EGF-like Ca^{+2} binding domains. The rare non-synonymous variants we observed in our discovery and replication cohorts are distributed across much of the protein, including some in these conserved domain regions (Fig 1A). Mutations in each of the FAT genes has been reported in many types of cancers; early T-cell precursor

acute lymphoblastic leukemia [35], ovarian [36], and pancreatic [37]. It is presumed that they all represent somatic not inherited mutations, although it is difficult to confirm this among the various sequencing studies. There is no particular known co-morbidity between IS and such cancers. More interestingly, multiple rare variants in *FAT3* were reported in families affected by the developmental disorder Hirschsprung disease [38]; two of the reported variants are present in our first stage discovery case cohort. As far as we know, there is no phenotypic component related to Hirschsprungin our cohorts. Although Hirschsprung disease is not obviously developmentally related to scoliosis, there are scattered reports in the literature of co-morbidity of these conditions [39-41]. Given the wide variety of developmental functions ascribed to the *FAT* genes, genetic associations of either common or rare variants to multiple complex disorders do not seem unlikely.

Somatic mutations in *FAT3* affect cell adhesion and interaction mechanisms, beside affecting the Wnt pathway [35]. Members of the FAT family proteins work synergistically and antagonistically to affect many aspects of tissue morphogenesis [42]. It has been shown that FAT3 and FAT4 act synergistically during fusion of the vertebral arches [42] through conserved interactions with components of planar polarity pathways. *Fat3* knockout mice have planar polarity defects [42]. A recent study demonstrated that a targeted mutation in the zebrafish *D. rerio* ptk7 gene, whose encoded protein functions in cell communication, leads to both congenital and idiopathic scoliosis according to the timing of gene loss of function. Furthermore, mutation of the gene led to the disruption of both planar cell polarity (PCP) and Wnt/ß-catenin signaling, consistent with the contribution of these pathways to the disease [43]. PCP pathways play an important role in regulating the polarity and behavior of different cells in different tissues [44]. Le Pabic *et al* suggested that PCP might be involved in skeletal morphogenesis as well [44]. They proposed a model whereby FAT3 coordinates the polarity and differentiation of chondrocytes affecting skeletal morphology. *FAT3* is highly expressed in the nervous

system and affects the neuronal morphology [45], beside its expression in the intervertebral discs [46], vertebral bone and other bone cells.

We directly compared *FAT3* gene expression levels in bone cells in a subset of our patients harboring rare variants in the gene to a group of controls lacking such variants. However, these rare variants in *FAT3* appeared to have no statistically significant effect on expression of the gene at least in this cell type. Somewhat unexpectedly, the statistical support for association of rare variants in *FAT3* with IS was stronger when synonymous variants were included. It has been shown that synonymous variants can affect mRNA splicing [47], mRNA stability and protein expression [48], and even in one case protein conformation and function [49]. We were not able to explore this directly due to lack of available biological materials from the particular cases in our cohort harboring such rare synonymous variants. However, the rare non-synonymous variants in our cases were not obviously clustered near exonic splice junctions.

In summary, our results implicate *FAT3* as a strong gene candidate contributing to either the occurrence or progression (or both) of idiopathic scoliosis.

Acknowledgements

Competing Interests

This work led to a patent application (pending) own by Sainte-Justine University Hospital (CHU Sainte-Justine). The authors declare no other potential conflicts of interest relevant to this manuscript.

References

1. Cheng JC, Castelein RM, Chu WC, et al. PRIMER. NATURE 2015

2. Asher MA, Burton DC. Adolescent idiopathic scoliosis: natural history and long term treatment effects. *Scoliosis and Spinal Disorders* 2006;**1**(1):2

3. Wise CA, Barnes R, Gillum J, et al. Localization of susceptibility to familial idiopathic scoliosis. *Spine* 2000;**25**(18):2372-80

4. Wynne-Davies R. Familial (idiopathic) scoliosis. *J Bone Joint Surg* [Br] 1968;**50**:24-30

5. Kesling KL, Reinker KA. Scoliosis in Twins: A Meta-analysis of the Literature and Report of Six Cases. *Spine* 1997;**22**(17):2009-14

6. Ward K, Ogilvie J, Argyle V, et al. Polygenic inheritance of adolescent idiopathic scoliosis: a study of extended families in Utah. *American Journal of Medical Genetics* Part A 2010;**152**(5):1178-88

7. Gorman KF, Julien C, Oliazadeh N, et al. *Genetics of Idiopathic Scoliosis*. eLS 2014

8. Edery P, Margaritte-Jeannin P, Biot B, et al. New disease gene location and high genetic heterogeneity in idiopathic scoliosis. *European Journal of Human Genetics* 2011;**19**(8):865-69

9.   Ocaka L, Zhao C, Reed JA, et al. Assignment of two loci for autosomal dominant adolescent idiopathic scoliosis (AIS) to chromosomes 9q31. 2-q34. 2 and 17q25. 3-qtel. *Journal of Medical Genetics* 2007

10.   Alden KJ, Marosy B, Nzegwu N, et al. Idiopathic scoliosis: identification of candidate regions on chromosome 19p13. *Spine* 2006;**31**(16):1815-19

11.   Gorman KF, Julien C, Moreau A. The genetic epidemiology of idiopathic scoliosis. *European Spine Journal* 2012;**21**(10):1905-19

12.   Sharma S, Londono D, Eckalbar WL, et al. A PAX1 enhancer locus is associated with susceptibility to idiopathic scoliosis in females. *Nature Communications* 2015;**6**

13.   Sharma S, Gao X, Londono D, et al. Genome-wide association studies of adolescent idiopathic scoliosis suggest candidate susceptibility genes. *Human Molecular Genetics* 2011;**20**(7):1456-66

14.   Takahashi Y, Kou I, Takahashi A, et al. A genome-wide association study identifies common variants near LBX1 associated with adolescent idiopathic scoliosis. *Nature Genetics* 2011;**43**(12):1237-40

15.   Kou I, Takahashi Y, Johnson TA, et al. Genetic variants in GPR126 are associated with adolescent idiopathic scoliosis. *Nature Genetics* 2013;**45**(6):676-79

16.   Ogura Y, Kou I, Miura S, et al. A functional SNP in BNC2 is associated with adolescent idiopathic scoliosis. *The American Journal of Human Genetics* 2015;**97**(2):337-42

17. Zuk O, Hechter E, Sunyaev SR, et al. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences* 2012;**109**(4):1193-98

18.   Asimit J, Zeggini E. Testing for rare variant associations in complex diseases. *Genome Medicine* 2011;**3**(4):1

19.   Marian AJ. Elements of 'missing heritability'. *Current Opinion in Cardiology* 2012;**27**(3):197-201

20.   Buchan JG, Alvarado DM, Haller GE, et al. Rare variants in FBN1 and FBN2 are associated with severe adolescent idiopathic scoliosis. *Human Molecular Genetics* 2014;**23**(19):5271-82

21.   Baschal EE, Wethey CI, Swindle K, et al. Exome sequencing identifies a rare HSPG2 variant associated with familial idiopathic scoliosis. G3: *Genes| Genomes| Genetics* 2015;**5**(2):167-74

22. Patten SA, Margaritte-Jeannin P, Bernard J-C, et al. Functional variants of POC5 identified in patients with idiopathic scoliosis. *The Journal of Clinical Investigation* 2015;**125**(3):1124-28 doi: 10.1172/JCI77262[published Online First: Epub Date]|.

23. Li W, Li Y, Zhang L, et al. AKAP2 identified as a novel gene mutated in a Chinese family with adolescent idiopathic scoliosis. *Journal of Medical Genetics* 2016:jmedgenet-2015-103684

24. Haller G, Alvarado D, Mccall K, et al. A polygenic burden of rare variants across extracellular matrix genes among individuals with adolescent idiopathic scoliosis. *Human Molecular Genetics* 2016;**25**(1):202-09

25. Godard B, Marshall J, Laberge C. Community engagement in genetic research: results of the first public consultation for the Quebec CARTaGENE project. *Public Health Genomics* 2007;**10**(3):147-58

26. Awadalla P, Boileau C, Payette Y, et al. Cohort profile of the CARTaGENE study: Quebec's population-based biobank for public health and personalized genomics. *International Journal of Epidemiology* 2012:dys160

27. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research* 2003;**31**(13):3812-14

28. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nature Methods* 2010;**7**(4):248-49

29. Schwarz JM, Cooper DN, Schuelke M, et al. MutationTaster2: mutation prediction for the deep-sequencing age. *Nature Methods* 2014;**11**(4):361-62

30. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *The American Journal of Human Genetics* 2008;**83**(3):311-21

31. Yang P, Liu H, Lin J, et al. The Association of rs4753426 Polymorphism in the Melatonin Receptor 1B (MTNR1B) Gene and Susceptibility to Adolescent Idiopathic Scoliosis: A Systematic Review and Meta-analysis. *Pain Physician* 2015;**18**(5):419-31

32. Lin D-C, Hao J-J, Nagata Y, et al. Genomic and molecular characterization of esophageal squamous cell carcinoma. *Nature Genetics* 2014;**46**(5):467-73

33. Katoh Y, Katoh M. Comparative integromics on FAT1, FAT2, FAT3 and FAT4. *International Journal of Molecular Medicine* 2006;**18**(3):523

34. Sadeqzadeh E, Bock CE, Thorne RF. Sleeping giants: emerging roles for the fat cadherins in health and disease. *Medicinal Research Reviews* 2014;**34**(1):190-221

35. Neumann M, Heesch S, Schlee C, et al. Whole-exome sequencing in adult ETP-ALL reveals a high rate of DNMT3A mutations. *Blood* 2013;**121**(23):4749-52

36. Network CGAR. Integrated genomic analyses of ovarian carcinoma. *Nature* 2011;**474**(7353):609-15

37. Furukawa T, Sakamoto H, Takeuchi S, et al. Whole exome sequencing reveals recurrent mutations in BRCA2 and FAT genes in acinar cell carcinomas of the pancreas. *Scientific Reports* 2015;**5**

38. Luzón-Toro B, Gui H, Ruiz-Ferrer M, et al. Exome sequencing reveals a high genetic heterogeneity on familial Hirschsprung disease. *Scientific Reports* 2015;**5**

39. McDonald-McGinn DM, Emanuel BS, Zackai EH. 22q11. 2 Deletion Syndrome. 2013

40. Reish O, Gorlin RJ, Hordinsky M, et al. Brain anomalies, retardation of mentality and growth, ectodermal dysplasia, skeletal malformations, Hirschsprung disease, ear deformity and deafness, eye hypoplasia, cleft palate, cryptorchidism, and kidney dysplasia/hypoplasia BRESEK/BRESHECK: New X-linked syndrome? *American Journal of Medical Genetics* 1997;**68**(4):386-90

41. Brooks AS, Breuning MH, Osinga J, et al. A consanguineous family with Hirschsprung disease, microcephaly, and mental retardation (Goldberg-Shprintzen syndrome). *Journal of Medical Genetics* 1999;**36**(6):485-89

42. Saburi S, Hester I, Goodrich L, et al. Functional interactions between Fat family cadherins in tissue morphogenesis and planar polarity. *Development* 2012;**139**(10):1806-20

43. Hayes M, Gao X, Yu LX, et al. ptk7 mutant zebrafish models of congenital and idiopathic scoliosis implicate dysregulated Wnt signalling in disease. *Nat Commun* 2014;**5**:4777 doi: 10.1038/ncomms5777[published Online First: Epub Date]|.

44. Le Pabic P, Ng C, Schilling TF. Fat-Dachsous signaling coordinates cartilage differentiation and polarity during craniofacial development. *PLoS Genet* 2014;**10**(10):e1004726

45. Deans MR, Krol A, Abraira VE, et al. Control of neuronal morphology by the atypical cadherin Fat3. *Neuron* 2011;**71**(5):820-32

46. Rock R, Schrauth S, Gessler M. Expression of mouse dchs1, fjx1, and fat-j suggests conservation of the planar cell polarity pathway identified in drosophila. *Developmental Dynamics* 2005;**234**(3):747-55

47. Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nature Reviews Genetics* 2002;**3**(4):285-98

48. Nackley AG, Shabalina S, Tchivileva IE, et al. Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. *Science* 2006;**314**(5807):1930-33

49. Kimchi-Sarfaty C, Oh JM, Kim I-W, et al. A" silent" polymorphism in the MDR1 gene changes substrate specificity. *Science* 2007;**315**(5811):525-28

Figure legends

Figure 1. Multiplex IS family. One family in our cohort (ID1581) consisted of three affected sisters and two unaffected parents. Panel A indicates the FAT3 protein organization as annotated by NCBI is 4557 amino acids long and includes multiple functional homology domains. The positions of the 26 rare variants identified in our study among the IS cases are labelled from 1 to 26 and are indicated by vertical arrows above the protein schema. The location of two heterozygous mutations present in this multiplex IS family are indicated by the red boxes. Panel B shows a simplified pedigree and segregation of the *FAT3* mutations. Panel C is a sequence alignment with different species showing that both mutations affect an invariantly conserved amino acid sequences in FAT3 orthologues. Panel D, represents sequence chromatograms showing those heterozygous mutations.
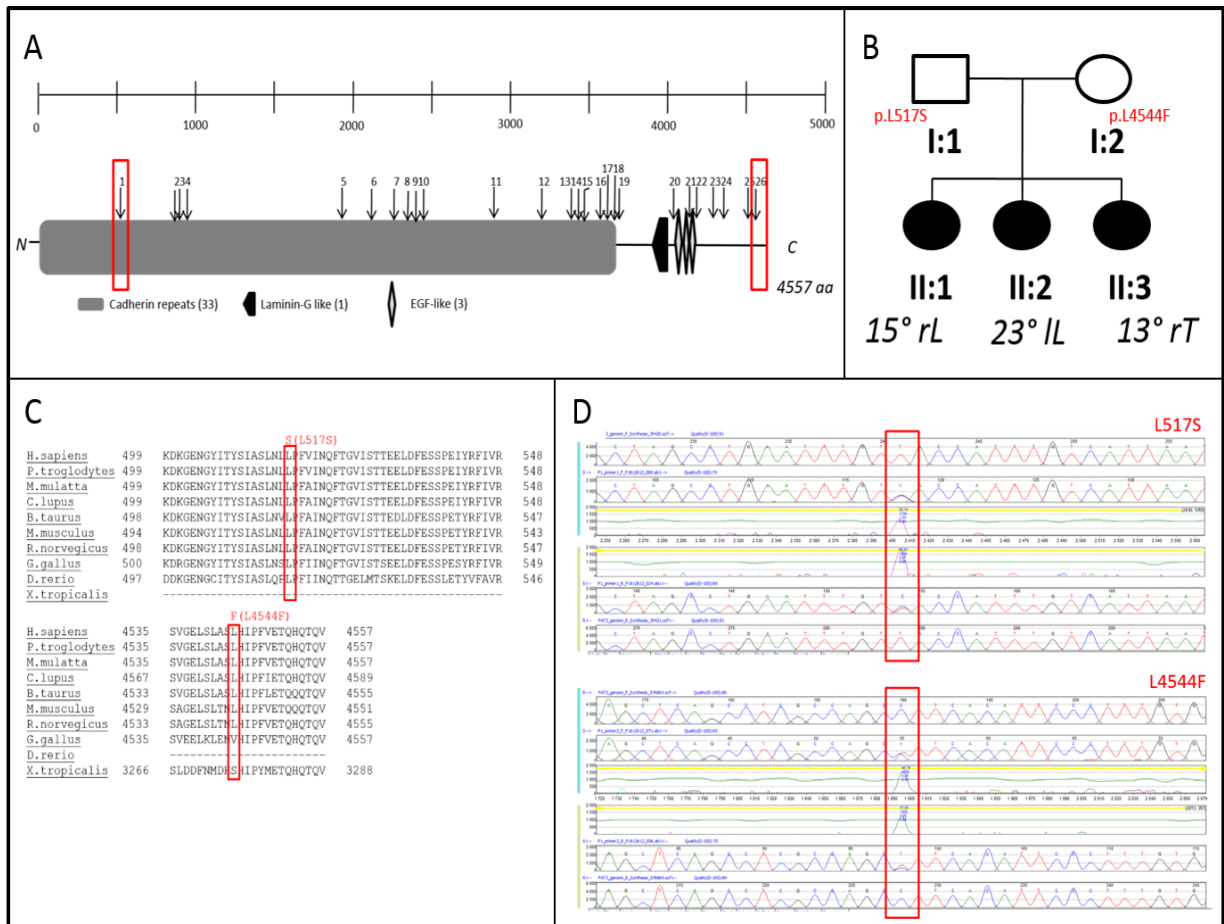
Figure 1

S1: Supplementary materials and methods

SOLiD 5500xl WES sequencing of the discovery cohort

Libraries were constructed using a modified version of the Fragment Library Preparation 5500 Series SOLiD™ Systems User Guide. Genomic DNA was fragmented with a Covaris®S2 System, 3 μg amounts as measured with a Bioanalyzer (Agilent) were used for library construction. Truncated adaptors were utilized to minimize nonspecific capture during SureSelectin-solution hybridization. Therefore, the P1-T and barcode-T-0XX adaptors were replaced by:

Tr5500P1

5′-   CCTCTCTATGGGCAGTCGGTGA*T -3′
3'- C*C*GGAGAGATACCCGTCAGCCACT   -5'
Tr5500IA
5′-   CGCCTTGGCCGTACAGC   -3′
3'   T*GCGGAACCGGCATGTCG*T*C   -5'

* Phosphorothioate bond

In addition, Library PCR Primer 1 and Library PCR Primer 2 were replaced by SureSelect Pre-Capture Primers provided in the SureSelect AB Barcoding Library Kit (Agilent). Exome capture was performed using Agilent SureSelectXT Human All Exon 50 Mb v3 according to the manufacturer's recommendations. Final libraries were quantified using the SOLiD™ Library TaqMan® Quantitation Kit. Standard steps were taken thereafter to create enriched, templated beads for the SOLiD 5500xl system. Pools of 8 libraries were loaded on each flowchip (6 lanes). Sequencing was performed in paired-end 50 bases in forward and 25 bases in reverse.

HiSeq 2500 WES sequencing for the French Canadian family

Genomic DNA was quantified using the QuantiFluor® dsDNA System (Promega) and 3ug was used as input. Libraries were constructed using the SureSelect[XT] Target Enrichment System for Illumina Paired-End Multiplexed Sequencing Library protocol and the SureSelect[XT] Human All Exon 50 Mb v5 capture kit (Agilent). Final libraries were qualified using a Bioanalyzer (Agilent) and quantified using the KAPA Library Quantification kit for Illumina. The clustering was done on Illumina cBot using 16pM of pooled libraries. Pools of 4 librairies were loaded on each lane of a High Output flow cell (8 lanes). Sequencing was performed on a HiSeq2500 for 125 cycles in paired-end using HCS 2.2.38 and RTA 1.18.61.

HiSeq2000 targeted genes sequencing in the replication cohort:

gDNA was quantified using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies). Libraries were generated robotically using the KAPA HTP Library Preparation Kit Illumina® platforms (Kapa Biosystems) as per the manufacturer's recommendations. TruSeq adapters and PCR primers were purchased from BioO. Libraries were quantified using the Kapa Illumina GA with Revised Primers-SYBR Fast Universal kit (D-Mark). Average size fragment was determined using a LaChip GX (PerkinElmer) instrument. Twenty ngs of 48 libraries were pooled together (total of 1000 ngs per capture) prior to proceeding with the enrichment of the targeted regions using the Roche Nimblegen EZ Choice custom baits. Captures were performed robotically according the manufacturer's recommendations. Final libraries were quantified using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies) and the Kapa Illumina GA with Revised Primers-SYBR Fast Universal kit (D-Mark). Average size fragment was determined using a LaChip GX (PerkinElmer) instrument. The clustering was done on a Illumina cBot using 11pM of each capture pool (2 captures per lane) and the

flowcell was ran on a HiSeq 2000 for 100 cycles in paired-end mode using HCS 2.2.58 and RTA 1.18.63 and using the manufacturer's instructions.

Bioinformatics:

The discovery cohort was analysed with the following pipeline:*.xsq files were converted to csfasta and qual file using XSQTools. 5' and 3' reads were filtered independently using SOLiD aware software (http://bioinformatics.oxfordjournals.org/content/26/6/849.full) and re-balanced to retain only read pairs i.e. singletons removed. The process was tested to ensure the settings used did not remove too much information, just poor quality reads. Reads were mapped in color space using bfast+bwa-0.7.0a to hg19 at the library level and then merged by sample. SNPs and INDELs were called with samtools 0.1.19 in batch mode i.e. all sample used during calling. SNPEff 3.3h was used to add genetic variant information and effect prediction (http://snpeff.sourceforge.net/). GATK indel Realiger (2.5-2) was used to help resolve indels, Picard Mark Duplicates (1.96) to label PCR duplicates, and GATK base recalibration (2.5-2, SOLiD specific settings) to re-calibrate base qualities due to various sources of systematic technical error. Variants were annotated using Gemini 0.11.1a (http://gemini.readthedocs.io/en/latest/index.html).

For quality control validation, Sanger sequencing was performed for more than 100 different variants throughout the exome and results showed consistency in 85% of the genotypes obtained using both sequencing techniques. Based on this we set our quality criteria of coverage> 10x, call rate> 90%, map quality > 20. We retained only variants having Minor allele frequency < 0.05 in 1000 genomes, Exome Sequencing Project, ExAC and dbSNP.

Both the replication French-Canadian IS cohort and the French-Canadian family were analysed using the same pipeline which was slightly different from the discovery cohort. The reads were trimmed

and aligned to the reference human genome (hg19) using Picard, BWA (0.5.9) (Li and Durbin, 2009) and Samtools (v.0.1.12a) (Li, et al., 2009). Variants were called using Pileup and varFilter commands, followed by filtering to keep Single nucleotide polymorphisms (SNPs) and insertion-deletions of Phred-like quality scores of more than 20 and 50 respectively. Our coverage was approximately 400x for the targeted sequencing and 100x for the five members of the family. Variants were annotated using ANNOVAR (Wang, et al., 2010), according to the type of mutation and frequency in the different data bases. We considered both SNPs and small insertion/ deletions (indels). We filtered those which are not in either exonic region or in a neighbouring potential splice site.

Quantitative Polymerase Chain Reaction (qPCR):

Total RNA was treated with DNase and reverse transcribed using the Maxima First Strand cDNA synthesis kit with ds DNase (Thermo Scientific). Before use, RT samples were diluted 1:5. Gene expression was determined using assays designed with the Universal Probe Library from Roche (www.universalprobelibrary.com). For each qPCR assay, a standard curve was performed to ensure that the efficacy of the assay is between 90% and 110%. QPCR reactions were performed using PERFECTA QPCR FASTMIX II (Quanta), 2 μM of each primer and 1 μM of the corresponding UPL probe. The Viia7 qPCR instrument (Life Technologies) was used to detect the amplification level and was programmed with an initial step of 20 sec at 95˚C, followed by 40 cycles of 1 sec at 95˚C and 20 sec at 60˚C.

References

Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 2009;**25**(14):1754-1760.

Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. Bioinformatics 2009;**25**(16):2078-2079.

Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic acids research 2010; **38**(16):e164-e164.

Supplementary tables

Table S1. Statistical analysis for all selected genes and SNPs in the replication cohort with SNPs with MAF < 2% (including the synonymous SNPs)

GENE	Rare SNPs	Cases with rare SNPs	Controls with rare SNPs	Un corrected Fisher-exact p-value
FAT3	31	32	8	2.86E-05
TENM3	17	25	11	0.015
A1CF	9	11	3	0.049
SEC16A	16	20	11	0.116
DMRT3	5	6	1	0.118
HPS4	18	12	5	0.125
IL16	13	12	5	0.125
ITGA8	10	10	4	0.163
NFRKB	7	7	2	0.169
LY75	16	11	6	0.310
ITGA4	3	4	1	0.368
DPEP3	5	8	4	0.372
R3HCC1L	8	12	8	0.479
GMPR2	1	2	0	0.497
CEACAM18	3	10	7	0.613
SLC22A16	5	11	8	0.630
CD1B	5	15	12	0.679
TTC21A	7	21	24	0.734
CCDC50	3	4	6	0.747
GLP1R	4	7	5	0.767
IMMT	4	7	5	0.767
GPR179	16	24	22	0.866
SLC3A1	5	6	5	1.000
ZNF189	1	1	0	1.000

Table S2. Statistical analysis for all selected genes and SNPs in the replication cohort with SNPs with MAF < 2% (after removing the synonymous SNPs)

GENE	Non-synonymous rare SNPs	Cases with non-synonymous rare SNPs	Controls with non-synonymous rare SNPs	Un corrected Fisher-exact p-value
FAT3	19	23	5	0.0006
SEC16A	8	15	10	0.391
HPS4	15	9	2	0.058
DMRT3	3	4	0	0.121
IL16	9	9	3	0.133
TENM3	7	10	4	0.163
ITGA8	6	5	1	0.211
A1CF	4	4	1	0.368
NFRKB	4	4	1	0.368
SLC22A16	3	4	1	0.368
LY75	12	10	6	0.434
R3HCC1L	7	10	6	0.434
DPEP3	4	7	4	0.537
TTC21A	4	13	17	0.552
ITGA4	2	3	1	0.621
CCDC50	3	4	6	0.747
GLP1R	3	6	4	0.747
CEACAM18	2	9	7	0.795
GPR179	15	24	22	0.866
CD1B	3	9	10	1.000
GMPR2	0	0	0	1.000
IMMT	0	0	0	1.000
SLC3A1	1	1	1	1.000
ZNF189	0	0	0	1.000

Supplementary figures
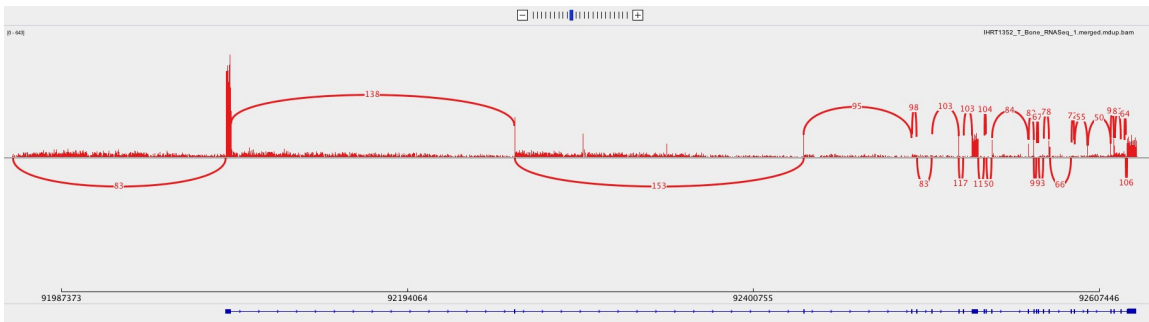
Supplementary Figure S1. Showing reads (83 reads) supporting the newly identified exon upstream the first annotated exon.
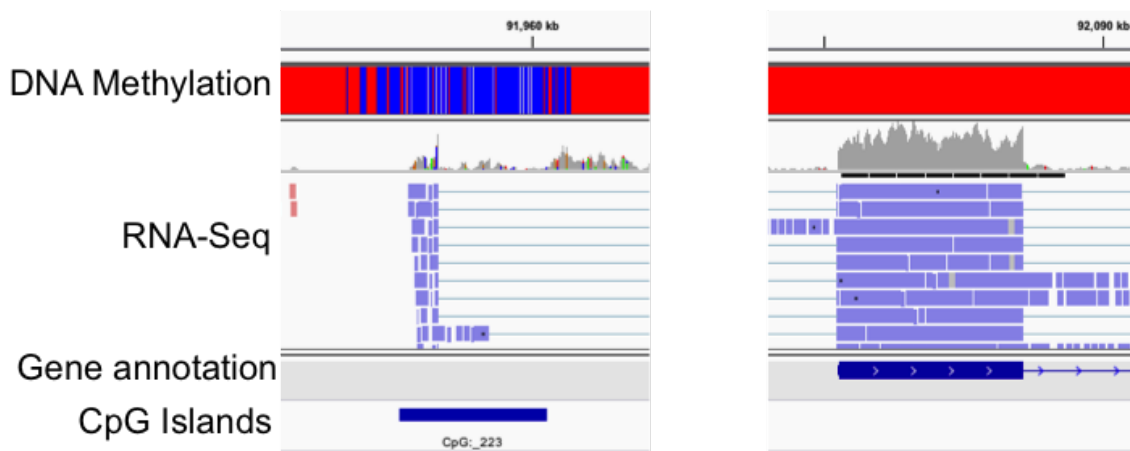
Supplementary Figure S2. A novel promoter of *FAT3*. Spliced reads map to a region 125kb upstream of the annotated *FAT3* first exon. This novel promoter (left) is a hypomethylated CpG island, a feature characteristic of active promoters whereas the first annotated exon (right), lacking those features, is putatively an internal exon.
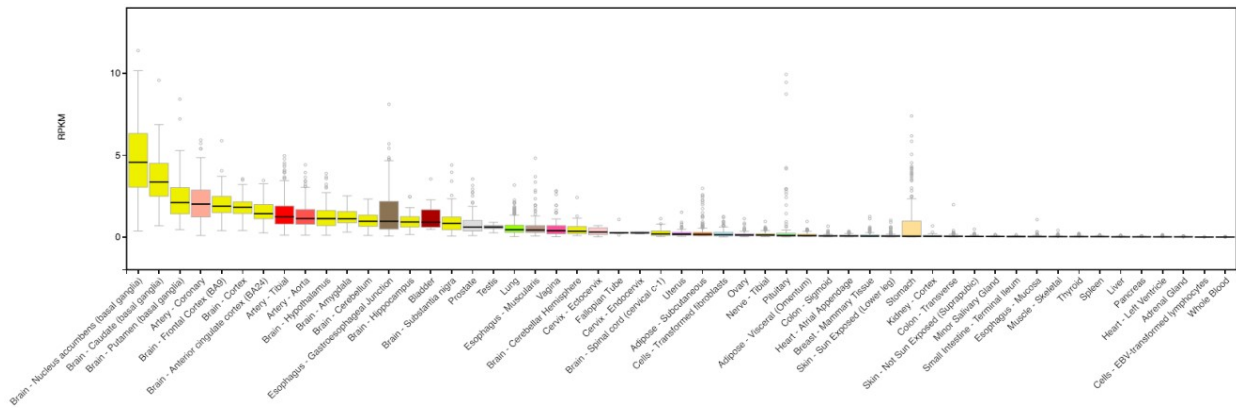
Supplementary Figure S3. Tissue-specific expression of *FAT3* in GTEx Analysis Release V6 (dbGaP Accession phs000424.v6.p1). Boxplots show enrichment of *FAT3* expression in brain and artery.
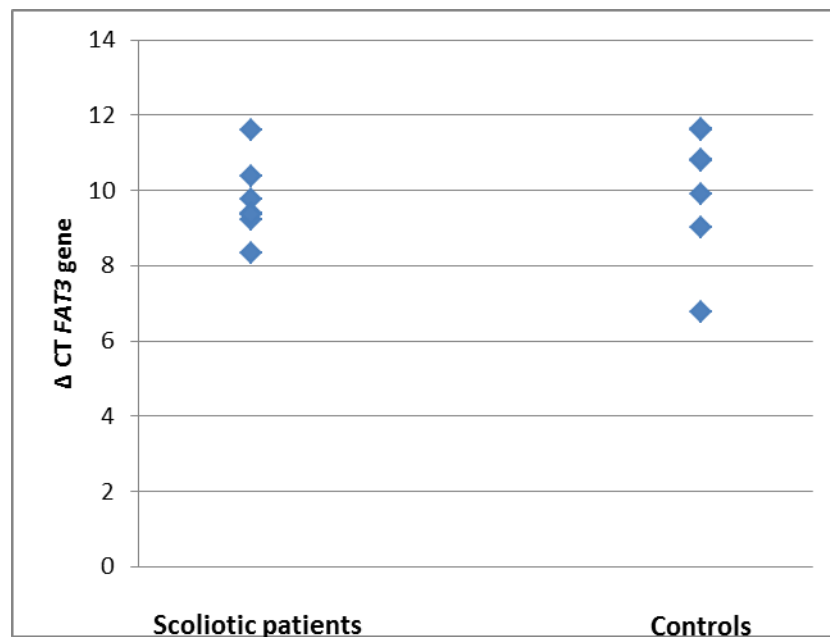
Supplementary Figure S4.*FAT3* expression levels in primary osteoblasts extracted from 7 scoliotic patients bearing rare variants (from both discovery and replication cohorts) versus 7 controls (trauma patients with no scoliosis).

2. Article 2

A Replication Study for Association of *LBX1* locus with

Adolescent Idiopathic Scoliosis in French-Canadian Population

Dina Nada, MSc[1,2], Cédric Julien, PhD[1], Mark E. Samuels, PhD[3,4] and Alain Moreau PhD[1,2,5,6]

[1] Viscogliosi Laboratory in Molecular Genetics of Musculoskeletal Diseases, Sainte-Justine University Hospital, Research Center, Montreal, QC, Canada

[2] Program of Biomedical Sciences, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada

[3] Sainte-Justine University Hospital Research Center, Montreal, QC, Canada

[4] Department of Medicine, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada

[5] Department of Biochemistry and Molecular Medicine, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada

[6] Department of Stomatology, Faculty of Dentistry, Université de Montréal, Montreal, QC, Canada

*Corresponding author: Prof. Alain Moreau, Sainte-Justine University Hospital Research Center, Scientific Director of Viscogliosi Laboratory in Molecular Genetics of Musculoskeletal Diseases (room 2.17.027), 3175 Cote-Sainte-Catherine Road, Montreal, Quebec, H3T 1C5, Canada.

Phone: 514-345-4931 ext. 5722; Fax: 514-345-4801

E-mail: alain.moreau@recherche-ste-justine.qc.ca

Contribution of authors:

Dina Nada: Participated in design of the study, experimental work, bioinformatics and statistical analyses of data, and as lead author of the manuscript.

Cédric Julien: Participated in design of the study, experimental work, bioinformatics and statistical analyses of data and participated in writing the manuscript.

Mark E Samuels: Supervised genomic analyses and participated in writing the manuscript.

Alain Moreau: Supervised all aspects of the study and participated in writing the manuscript.

A Replication Study for Association of *LBX1* locus with Adolescent Idiopathic Scoliosis in French-Canadian Population

Dina Nada, MSc[1,2],Cédric Julien, PhD[1], Mark E. Samuels, PhD[3,4]and Alain Moreau PhD[1,2,5,6]

[1]Viscogliosi Laboratory in Molecular Genetics of Musculoskeletal Diseases, Sainte-Justine University Hospital, Research Center, Montreal, QC, Canada

[2]Program of Biomedical Sciences, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada

[3]Sainte-Justine University Hospital Research Center, Montreal, QC, Canada

[4]Department of Medicine, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada

[5]Department of Biochemistry and Molecular Medicine, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada

[6]Department of Stomatology, Faculty of Dentistry, Universitéde Montréal, Montreal, QC, Canada

*Corresponding author: Prof. Alain Moreau, Sainte-Justine University Hospital Research Center, Scientific Director of Viscogliosi Laboratory in Molecular Genetics of Musculoskeletal Diseases (room 2.17.027), 3175 Cote-Sainte-Catherine Road, Montreal, Quebec, H3T 1C5, Canada.

Phone: 514-345-4931 ext. 5722; Fax: 514-345-4801

E-mail: alain.moreau@recherche-ste-justine.qc.ca

Structured Abstract

Study Design. A case-control association study.

Objectives. To investigate the relationship between *LBX1* polymorphisms and Adolescent Idiopathic Scoliosis (AIS) in French-Canadian population.

Summary of Background Data. It is widely accepted that genetic factors contribute to AIS. Although the *LBX1* locus is so far the most successfully replicated locus in different AIS cohorts, these associations were replicated mainly in Asian populations, with few studies in Caucasian populations of European descent.

Methods. We recruited 1568 participants (667 AIS patients and 901 healthy controls) in the French-Canadian population. Genomic data was generated using the Illumina HumanOmni 2.5M BeadChip. An additional 121 AIS cases and 51 controls were genotyped for specific SNPs by multiplex PCR using standard procedures. BEAGLE 3 was used to impute the following markers: rs7893223, rs11190878 and rs678741 against the 1000-genomes European cohort phased genotypes given that they were absent in our GWAS panel. Resulting genotypes were combined then used for single marker and haplotyped-based association.

Results. Four markers showed association with AIS in our cohort at this locus; rs11190870 the most studied marker, rs7893223, rs594791, and rs11190878. When we restricted the analysis to severe cases only, four additional SNPs showed associations: rs11598177, rs1322331, rs670206 and rs678741. In addition, we analyzed the associations of the observed haplotypes and dihaplotypes formed by these SNPs. The haplotype TTAAGAAA and its homozygous dihaplotype showed the highest association with our severe group and was the highest risk haplotype. The haplotype CCGCAGGG was significantly more associated with the control group, and its homozygous or heterozygous dihaplotype

was less frequent in the severe group compared to the control group, suggesting that CCGCAGGG may represent a protective haplotype.

Conclusions. We have replicated the association of the *LBX1* locus with AIS in French-Canadian population, a novel European descent cohort, which is known for its unique genetic architecture.

Keywords: adolescent idiopathic scoliosis, genome-wide association study, single nucleotide polymorphism, haplotypes, *LBX1*, French-Canadian, disease severity

Level of evidence: 3

Mini Abstract

We examined the association of the scoliosis phenotype to markers near the known *LBX1* locus in the French-Canadian population. We confirmed association of this locus with scoliosis at both SNP haplotype and dihaplotype levels. We identified haplotypes either associated with disease severity or acting in a protective manner.

Introduction

Idiopathic Scoliosis (IS) is a common complex disease and one of the most prevalent childhood deformities. It presents both an immediate medical challenge and a chronic condition that affects individuals throughout their lives. It is mostly first identified/ diagnosed between the age of 10 and 15 years old. On average Adolescent Idiopathic Scoliosis (AIS) affects 1-4%[1] of the global pediatric population. Although the syndrome is phenotypically complex, it is clinically characterized by a 3D spinal deformity with unknown cause.[2] Most patients requiring corrective surgery are females who have five times greater risk of curve progression than males; the cause of this gender bias is not well understood.[3]

A genetic component of AIS is well documented. Clinical studies have noted the familial incidence of scoliosis since the 1930s.[4] Segregation studies reported an increased risk of AIS among first degree relatives of affected individuals, and that this risk decreases as the degree of relation to the affected individual becomes more distant.[5] Moreover, twin studies have shown that the rate of concordance of AIS in monozygotic twins is almost double that of the dizygotic twins.[6] A recent study conducted on a huge number of twins in the Swedish twin registry estimated that the genetic component of AIS susceptibility is 38%, while the environmental component is 62%.[7]

A number of genome wide association studies (GWAS) have been conducted on AIS since 2011, however few loci were successfully replicated. A GWAS study in a Japanese cohort reported an association between a locus at the 10q24.31 and AIS.[8] They reported common SNPs in the vicinity of the *LBX1* (lady bird homeobox1) gene, of which the most significant SNP was rs11190870. Several replication studies subsequently confirmed this association in different populations.[9-13] One of those studies identified two haplotypes involving rs11190870 that were highly associated with AIS.[14] Other genes identified by GWAS for AIS susceptibility include: G-protein coupled receptor *GPR126*, *SOX9*, cell adhesion molecule *CHL1,* basonuclin 2 *BNC2* and transcription factor *PAX1*.[15-19] Although the *LBX1* locus is so far the most successfully replicated locus in different Asian populations, these associations were replicated in few studies in Caucasian populations of European descent: two from the United States [12,14] and one in a Scandinavian population.[20]

We performed a GWAS to test for genetic markers associated with AIS in our primarily French-Canadian cohort. The objective of the current analysis tested for the association of the *LBX1* locus in a French-Canadian population which was not done before. Our results support an association with the region near the *LBX1* gene in our population.

Materials and methods

Study population, clinical descriptions and samples

This study was approved by the institutional review boards of Sainte-Justine University Hospital, Montreal Children's Hospital, Shriners Hospital for Children in Montreal and McGill University, as well as the Affluent and Montreal English School Boards. We recruited prospectively 1056 individuals from primary schools in Quebec; 864 diagnosed with AIS and 192 considered unaffected with AIS. Diagnosis of AIS was based on consistency between history and physical examination with AIS diagnosis where there's a minimum Cobb angle of 10° in the coronal plane associated with a vertebral rotation, on a standing radiograph. Each patient was examined by an orthopedic surgeon at Sainte-Justine University Hospital, Montreal, Quebec, Canada. To verify the absence of the phenotype in the school controls, each healthy individual was also examined by the same orthopedic surgeon using the Adam forward bending test with a scoliometer. Additional unaffected controls comprising 750 adult individuals were obtained from the CARTaGENE project, with screening of clinical exclusion criteria using the historical data provided by CARTaGENE.[21,22] Ethnicity based on self-reporting, verified by principal component analysis of genotype data (see below). For schools-derived affected and control individuals, DNA was obtained from peripheral blood leukocytes following standard venipuncture. Informed written consent was obtained from parents or legal guardians of all participants in addition to assent given by the minors. For CARTaGENE-derived controls, consent was previously provided through inclusion in the biobank.

Genome wide association (GWAS)

Genomic DNA samples from our cohort of cases and controls were genotyped by the Illumina Human Omni 2.5M BeadChip, at Genome Quebec Innovation Centre (McGill University). Genotype data from CARTaGENE participants were obtained previously,[23] and merged with our cohort data. Quality control (QC) measures were applied to genotype data as previously described.[24,25] We used PLINK[26] and R[27] software packages to: 1) filter gender mismatches, 2) filter for poor calling at both the

sample level (< 2%) and SNP level (< 2%), 3) assess sample heterozygosity, 4) filter SNPs with a minor allele frequencies (MAFs) less than 1%, and 5) filter SNPs in Hardy-Weinberg disequilibrium.

Linkage disequilibrium (LD) pruning was performed on the filtered genotype data prior to ancestral and relatedness testing by applying respectively EIGENSTRAT (Principal Component Analysis or PCA, analysis of self-reported ethnicity)[28] andPLINK[26] identity-by-descent (IBD). To account for population admixture, we performed LD pruning using a window size of 1500 SNPs, a window increment of 150 SNPs and an LD cut-off defined by r2>0.2. PCA was conducted on the LD-pruned SNPs and the 10 top PC vectors were analyzed. Self-reported French-Canadian individuals falling outside the main core cluster were removed from further analyses. Another analysis was performed on the main core cluster to look for any remaining population substructures. Using the IBD approach, ancestral outliers and related samples (pi_hat>0.1875) were removed. These QC procedures retained over 1.4 million SNPs among 667 AIS patients (545 females and 122 males) and 901 controls (476 females and 425 males, controls consisting of 170 individuals from schools and 731 individuals from CARTaGENE. The genomic inflation factor given by PLINK was estimated to be $\lambda=1.13$.

PCR-based genotyping

An additional 121 AIS cases and 51 controls were genotyped for specific SNPs by multiplex PCR using standard procedures with 20ng of template genomic DNA and HotStarTaq DNA polymerase enzyme (QIAGEN).PCR reactions were run on the QIAxcel (QIAGEN) to assess the amplification, followed by the single base extension using iPlex ThermoSequenase. Genotypes were determined by MALDI-TOF mass-spectrometry and data were analyzed using MassARRAY Typer Analyser software.

Imputation procedure

BEAGLE 3 was used to impute the following markers: rs7893223, rs11190878 and rs678741 against the 1000-genomes European cohort phased genotypes (N=503). Resulting imputed genotypes were then used for single marker and haplotyped-based association.

Statistical analyses

The association of SNPs near the *LBX1* locus on chromosome 10 was evaluated using PLINK v1.09. In our analyses, we performed separate analysis with all AIS patients (N=788) and severe cases only defined as spinal curves $\geq$40° (N=202, average spinal curve=54.9°, SD=11.9°, 87% females). Differences in allelic frequencies between cases and control subjects were evaluated by calculating Odd Ratios (ORs), together with their 95% confidence intervals (CI) and p-values. Two steps for haplotype association tests were performed using, first, the

'--chap' option in PLINK to calculate ORs, CI and, second, the '--hap-logistic' and '--mperm' options to calculate p-values and perform permutation tests (n=50000) of each estimated haplotypes having a frequency higher than 1%. Haplotype visualization was done using Haploview. The r-squared values and color display were used to visualize linkage disequilibrium between the candidate SNPs.

Results

Genotyping and single-marker association near the *LBX1* locus

Our cohort comprised 788 AIS cases and 952 controls (after merging the GWAS data with the subsequently genotyped individuals). Of the affected cases, 202 were classified as severe (Cobb angle $\geq$ 40$^{\text{o}}$), 295 were classified as non-severe (Cobb angle 10-39$^{\text{o}}$); the remaining cases had not reached skeletal maturity hence severity could not be assessed. Our mapping panel of 2.5 million SNPs represents the highest resolution GWAS for AIS to date.

We first tested the association of rs11190870, the marker reported several times in the literature for its association with AIS. Indeed, this marker was significantly associated with our AIS cohort (OR=0.82, p=0.0047). Trait severity (measured by Cobb angle of curvature) had an impact on this association to *LBX1*, which gave higher relative risk and more significant p-value for the SNP when severely affected individuals only were compared to the complete set of controls (OR=0.66, p=0.0003).

In the literature there are other markers reported to be associated with AIS near the *LBX1* locus, these markers lie in a region 40kb around the *LBX1* gene. We analyzed SNPs that lie in this region and were present in our GWAS panel. After considering those with a MAF≤ 2%, 12 SNPs remained. One of these (rs594791: OR=0.82, p=0.0054) showed association with AIS similar to rs11190870.When only severely affected individuals were considered, three additional SNPs in this region (rs11598177, rs1322331 and rs679206) showed similar associations (**Table 1**).

Imputation and haplotype analysis of the *LBX1* association

The five SNPs that showed highest association in our cohort are in strong linkage disequilibrium with each other (r^2 ranges from 0.76 to 0.99, **Figure1**). It has been previously reported[14] that two specific haplotypes with opposite effects are associated with AIS in a North American cohort. Those haplotypes are each composed of 3 SNPs: rs7893223, rs11190870 and rs11190878. Of these rs7893223 and rs11190878 were absent in our data set. rs678741 is another associated SNP at the *LBX1* locus[29] also not in our genotype panel. We imputed genotypes of these three additional SNPs (rs7893223, rs11190878 and rs678741) using BEAGLE 3 and the 1000-genomes European cohort phased genotypes (**Table 1, Figure 1**). Two of these SNPs (rs7893223 and rs11190878) showed the most significant association in our cohort when we analyzed both whole cohort versus controls and when severe cases

only were analyzed versus controls. The third SNP rs678741 showed significant association only when severe cases were implicated.

We next analyzed the haplotypes formed by these three imputed SNPs combined with the five associated SNPs of our dataset. We observed four major haplotypes in our cohort (**Table 2**). When all the cases are considered, two haplotypes were significantly associated (CCGCAGGG, OR=0.77, p=0.0002, p perm (p permutation)=0.0008) and (TCGCAGGA, OR=2.54, p=0.0005,p perm=0.0037). Two haplotypes were significantly associated when severe cases only were included: TTAAGAAA (OR=1.42, p=0.0016,p perm=0.0107) and CCGCAGGG (OR=0.64, p=$2x10^{-5}$, p perm=0.0003).

We also analyzed the various dihaplotype genotype combinations of our observed haplotypes (**Table 3**). The haplotype TTAAGAAA was homozygous in 41% of the severe cases; this was the most highly associated dihaplotype among our severe cases suggesting it to be the most at risk dihaplotype. TTAAGAAA/CCGCAGGG heterozygotes were significantly more represented in the control group (43%) than in the severe AIS group (33%; OR=0.65, p=0.008), and all AIS group (37%, OR=0.76, p=0.006), and homozygotes for CCGCAGGG were slightly significantly under-represented in the severe group (10%) compared to controls (16%, OR=0.59, p=0.0331), suggesting that CCGCAGGG may be a resistance-to-risk haplotype. Finally, the dihaplotype composed of TTAAGAAA and TCGCAGGA was over-represented in a small proportion of total AIS cases (3.2%) compared to controls (1.2%, OR=2.8, p=0.0032).

Discussion

In this study we have replicated the reported association of the *LBX1* locus with AIS in a French-Canadian cohort, the first genetic study of this ethnic group. This locus was first observed by a Japanese study, which presented three SNPs with rs11190870 as the most significantly associated SNP.[8] Since

then, this locus has been extensively studied and confirmed by many Asian studies.[9,10,13] The association of this locus with cohorts of European descent has been also confirmed by few studies; one included two non-Hispanic white cohorts,[12] one with a Caucasian cohort taken from the general U.S. population and not otherwise specified[14] and one with a Scandinavian population.[20] Although these latter cohorts are more closely related to ours than the Asian cohorts, our French-Canadian cohort is nonetheless a different one with a unique genetic architecture. Thus verifying the *LBX1* association is an important first step in understanding the genetics of AIS in our population.

Four markers showed association with AIS in our cohort at this locus; rs11190870 the most studied marker, rs7893223, rs594791, and rs11190878. When we restricted the analysis to severe cases only, four additional SNPs showed associations: rs11598177, rs1322331, rs670206 and rs678741. Two studies from the Asian populations reported contradictory results concerning association of *LBX1* with the severity of AIS. Gao et al. reported no association with this locus and the severity of the cases,[9] whereas Jiang et al. reported an association between rs11190870 and both the susceptibility and the severity of the cases.[10] In our study, we support the latter result where the trait severity showed an impact on the association to *LBX1* and gave higher relative risk and more significant p-values.

In addition to individual SNPs, we analyzed the associations of the observed haplotypes and dihaplotypes formed by these SNPs in our cohort. The haplotype TTAAGAAA and its homozygous dihaplotype showed the highest association with our severe group and was the highest risk haplotype. The haplotype CCGCAGGG was significantly more frequent in the control group, and its homozygous dihaplotype showed lower frequency in the severe group compared to the control group and may represent a protective haplotype. A third haplotype TCGCAGGA showed association with all cases only but not with just severe cases, although this was potentially due to a lack of statistical power as this haplotype has a frequency of 0.017 in our severe cohort. Our results support the interpretation that there

are both at-risk and protective genetic factors in the population with respect to idiopathic scoliosis, and further demonstrate the utility of haplotype and even dihaplotype analyses for this complex trait.[14]

LBX1 encodes a homeobox-like transcription factor, related to the Drosophila ladybird late (lbl) gene.[30] LBX1 is highly expressed in skeletal muscle and both adult and fetal spinal cords,[8] which is consistent with a genetic role in spinal deformities such as scoliosis. Of note, a recent publication by Zhu Z et al 2017 reported a significant asymmetric expression of Wnt/beta-catenin pathway in the bilateral paraspinal muscle of AIS patients, including beta-catenin, TNIK, and LBX1.[31] The interplay between Wnt/beta-catenin pathway and *LBX1* remains to be clarified but previous works published by Guo L et al (2016) showed that overexpression of *LBX1* or either of its zebrafish homologs in zebrafish induces spinal deformities and downregulates the expression of wnt5b, a ligand of the non-canonical Wnt/planar cell polarity (PCP) pathway.[32] Interestingly, injection of mRNA for wnt5b or RhoA, a key downstream effector of Wnt/PCP signalling attenuated the *LBX1* induced curvature of the body axis suggesting additional deregulated mechanisms. Among others, *LBX1* overexpression could deregulate somatosensory fate on relay neurons in the hindbrain[33] and dorsal spinal cord.[34] Given the higher prevalence of somatosensory disorders in AIS patients than in the general population[35-37] and the fact that previous reports showed that children with abnormal somatosensory functions are more susceptible to the development of scoliosis than normal ones[2] it could be hypothesized that the contribution of LBX1 to the pathophysiology of AIS is due to its role in somatosensory function.[8] Indeed, animal models have been shown to develop scoliosis when subjected to damage to the sensory area of the spinal cord or posterior rhizotomy.[38-41] Alternatively these could be unlinked phenotypes resulting from a pleiotropic effect of LBX1 dysfunction.

The SNPs that were shown to be associated with AIS in previous studies and confirmed in our study lie in a region encompassing the two head to head genes *LBX1* and *LBX1-AS1* (RNA gene)

(**Figure 1**). *LBX1-AS1* is a long non-coding RNA which encodes for the antisense RNA of *LBX1* transcript. Zhu et al previously suggested that *LBX1-AS1* might contribute to AIS risk.[29] Very recently, Liu et al demonstrated an association between rs1322331 TT genotype and the decreased expression of *LBX1-AS1* and suggested that this in turns increases the expression of *LBX1* and consequently increases AIS susceptibility.[42]

In summary, we confirm the association of the *LBX1* locus with AIS with regard to both risk and disease severity in a French-Canadian cohort. This locus is the most solidly confirmed region to be associated with the disease, but nonetheless it only explains a very small percentage of the estimated genetic variance of AIS.[17] Thus, additional genetic factors must be present in populations, which may be either common (or identifiable by GWAS) or rare (and identifiable by genome or exome sequencing). Grauers et al further confirmed the association of rs11190870 with AIS in 1739 patients versus 1812 controls but they failed to find rare causing variants in *LBX1* by sequencing the exonic, promoter and 5'UTR regions of the gene in 100 severe cases.[20]

The interaction between the *LBX1* locus with other common or rare variants will help to build complex models that explain more clearly the etiology of the disease. A very interesting study recently documented genetic interaction between common and rare variants of two functionally related genes involved in the same phenotype.[43]

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

Shriners Hospital for Children in Montreal. We also thank Ms. Anita Franco, project coordinator, for her technical assistance.

References

1. Cheng JC, Castelein RM, Chu WC, et al. *Nature Reviews Disease Primer*, 2015, (1): 1-20.

2. Kouwenhoven J-WM, Castelein RM. The pathogenesis of adolescent idiopathic scoliosis: review of the literature. *Spine* 2008;33:2898-908.

3. Wise CA, Gao X, Shoemaker S, et al. Understanding genetic factors in idiopathic scoliosis, a complex disease of childhood. *Current genomics* 2008;9:51-9.

4. Garland HG. Hereditary scoliosis. *British medical journal* 1934;1:328.

5. Ward K, Ogilvie J, Argyle V, et al. Polygenic inheritance of adolescent idiopathic scoliosis: a study of extended families in Utah. *American journal of medical genetics Part A* 2010;152:1178-88.

6. Kesling KL, Reinker KA. Scoliosis in Twins: A Meta-analysis of the Literature and Report of Six Cases. *Spine* 1997;22:2009-14.

7. Grauers A, Rahman I, Gerdhem P. Heritability of scoliosis. *European Spine Journal* 2012;21:1069-74.

8. Takahashi Y, Kou I, Takahashi A, et al. A genome-wide association study identifies common variants near LBX1 associated with adolescent idiopathic scoliosis. *Nature genetics* 2011;43:1237-40.

9. Gao W, Peng Y, Liang G, et al. Association between common variants near LBX1 and adolescent idiopathic scoliosis replicated in the Chinese Han population. *PloS one* 2013;8:e53234.

10. Jiang H, Qiu X, Dai J, et al. Association of rs11190870 near LBX1 with adolescent idiopathic scoliosis susceptibility in a Han Chinese population. *European Spine Journal* 2013;22:282-6.

11. Liang J, Xing D, Li Z, et al. Association between rs11190870 polymorphism near lbx1 and susceptibility to adolescent idiopathic scoliosis in East Asian population: a genetic meta-analysis. *Spine* 2014;39:862-9.

12. Londono D, Kou I, Johnson TA, et al. A meta-analysis identifies adolescent idiopathic scoliosis association with LBX1 locus in multiple ethnic groups. *Journal of medical genetics* 2014;51:401-6.

13. Fan Y-H, Song Y-Q, Chan D, et al. SNP rs11190870 near LBX1 is associated with adolescent idiopathic scoliosis in southern Chinese. *Journal of human genetics* 2012;57:244-6.

14. Chettier R, Nelson L, Ogilvie JW, et al. Haplotypes at LBX1 have distinct inheritance patterns with opposite effects in adolescent idiopathic scoliosis. *PloS one* 2015;10:e0117708.

15. Sharma S, Londono D, Eckalbar WL, et al. A PAX1 enhancer locus is associated with susceptibility to idiopathic scoliosis in females. *Nature communications* 2015;6: 6452.

16. Ogura Y, Kou I, Miura S, et al. A functional SNP in BNC2 is associated with adolescent idiopathic scoliosis. *The American Journal of Human Genetics* 2015;97:337-42.

17. Kou I, Takahashi Y, Johnson TA, et al. Genetic variants in GPR126 are associated with adolescent idiopathic scoliosis. *Nature genetics* 2013;45:676-9.

18. Sharma S, Gao X, Londono D, et al. Genome-wide association studies of adolescent idiopathic scoliosis suggest candidate susceptibility genes. *Human molecular genetics*2011;20:1456-66.

19. Miyake A, Kou I, Takahashi Y, et al. Identification of a susceptibility locus for severe adolescent idiopathic scoliosis on chromosome 17q24. 3. *PloS one* 2013;8:e72802.

20. Grauers A, Wang J, Einarsdottir E, et al. Candidate gene analysis and exome sequencing confirm LBX1 as a susceptibility gene for idiopathic scoliosis. *The Spine Journal* 2015; 15: 2239-46.

21. Awadalla P, Boileau C, Payette Y, et al. Cohort profile of the CARTaGENE study: Quebec's population-based biobank for public health and personalized genomics. *International journal of epidemiology* 2012:dys160.

22. Godard B, Marshall J, Laberge C. Community engagement in genetic research: results of the first public consultation for the Quebec CARTaGENE project. *Public Health Genomics* 2007;10:147-58.

23. Hodgkinson A, Idaghdour Y, Gbeha E, et al. High-resolution genomic analysis of human mitochondrial RNA sequence variation. *Science* 2014;344:413-5.

24. Turner S, Armstrong LL, Bradford Y, et al. Quality control procedures for genome-wide association studies. *Current protocols in human genetics* 2011:1.19. 1-1.. 8.

25. Weale ME. Quality control for genome-wide association studies. *Genetic Variation: Methods and Protocols* 2010:341-72.

26. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* 2007;81:559-75.

27. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013, 2014.

28. Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics* 2006;38:904-9.

29. Zhu Z, Tang NL-S, Xu L, et al. Genome-wide association study identifies new susceptibility loci for adolescent idiopathic scoliosis in Chinese girls. *Nature communications* 2015;6.

30. Jagla K, Dollé P, Mattei M-G, et al. Mouse Lbx1 and human LBX1 define a novel mammalian homeo gene family related to the Drosophila lady bird genes. *Mechanisms of development* 1995;53:345-56.

31. Zhu Z, Xu L, Leung-Sang TN, et al. Genome-wide association study identifies novel susceptible loci and highlights Wnt/beta-catenin pathway in the development of adolescent idiopathic scoliosis. *Human molecular genetics* 2017; Feb 28. doi: 10.1093/hmg/ddx045. [Epub ahead of print]

32. Guo L, Yamashita H, Kou I, et al. Functional Investigation of a Non-coding Variant Associated with Adolescent Idiopathic Scoliosis in Zebrafish: Elevated Expression of the Ladybird Homeobox Gene Causes Body Axis Deformation. *PLoS Genet* 2016; 12: e1005802.

33. Sieber MA, Storm R, Martinez-de-La-Torre M, et al. Lbx1 acts as a selector gene in the fate determination of somatosensory and viscerosensory relay neurons in the hindbrain. *Journal of Neuroscience* 2007; 27: 4902-9.

34. Gross MK, Dottori M, Goulding M. Lbx1 specifies somatosensory association interneurons in the dorsal spinal cord. *Neuron* 2002; 34: 535-49.

35. Cheng JC, Guo X, Sher AH. Posterior tibial nerve somatosensory cortical evoked potentials in adolescent idiopathic scoliosis. *Spine* 1998;23:332-7.

36. Cheng JC, Guo X, Sher AH, et al. Correlation between curve severity, somatosensory evoked potentials, and magnetic resonance imaging in adolescent idiopathic scoliosis. *Spine* 1999; 24: 1679.

37. Guo X, Chau WW, Hui-Chan CW, et al. Balance control in adolescents with idiopathic scoliosis and disturbed somatosensory function. *Spine* 2006;31:E437-E40.

38. Taffs L. Experimental scoliosis in primates: a neurological cause. *Bone & Joint Journal* 1982;64:503-7.

39. Davies J, Taffs L. Scoliosis caused by section of dorsal spinal nerve roots. *Bone & Joint Journal* 1984;66:27-9.

40. Barrios C, Tunon MT, De Salis JA, et al. Scoliosis induced by medullary damage: an experimental study in rabbits. *Spine* 1987;12:433-9.

41. Suk S, Song H, Lee C. Scoliosis induced by anterior and posterior rhizotomy. *Spine* 1989;14:692-7.

42. Liu S, Wu N, Zuo Y, et al. Genetic Polymorphism of LBX1 is Associated with Adolescent Idiopathic Scoliosis in Northern Chinese Han Population. *Spine* 2017.

43. Timberlake AT, Choi J, Zaidi S, et al. Two locus inheritance of non-syndromic midline craniosynostosis via rare SMAD6 and common BMP2 alleles. *eLife* 2016;5:e20125.

Figure legends:

Figure 1. Associated SNPs in the vicinity of the *LBX1* and the *LBX1-AS1* genes. The relative positions of the 8 studied SNPs are shown with respect to *LBX1* and the *LBX1-AS1* genes on chromosome 10. The arrows denote the direction of transcription of the genes. The pattern of linkage disequilibrium between the SNPs is also shown as determined by the HapMap project. The position of the *LBX1* gene: 101 226 976-101 228 960, *LBX1-AS1* gene: 101229594-101238859. The positions of the SNPs on chromosome 10 are as follows; rs7893223:101210403, rs11190870:101219449, rs11598177:101220398, rs1322331:101226831, rs594791:101236038, rs679206:101237692, rs678741: 101237823, rs11190878: 101250150.
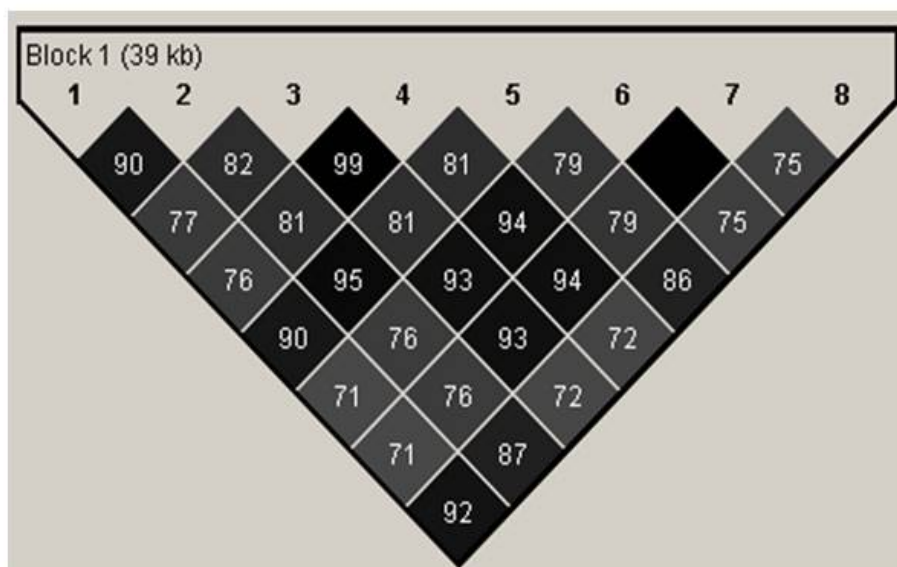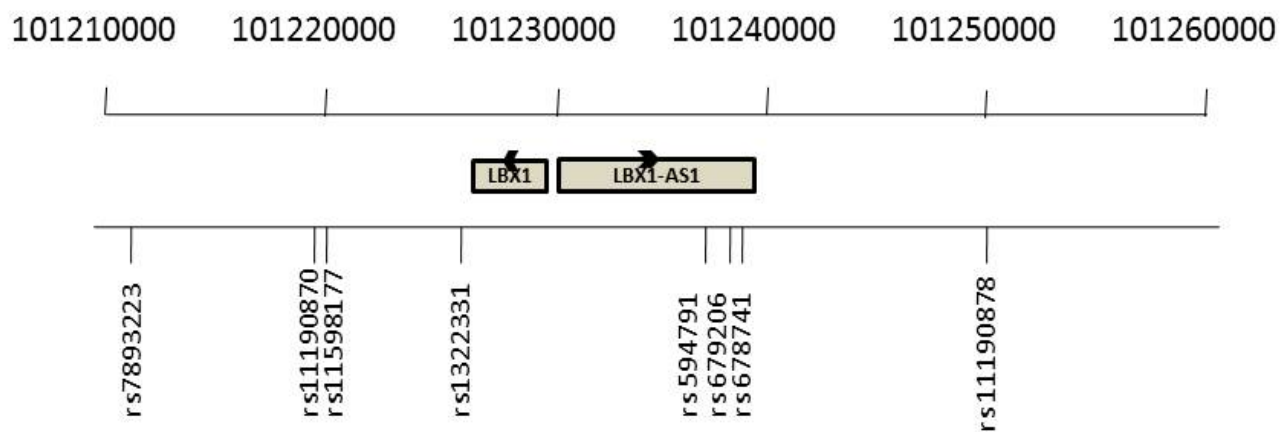
Figure 1

Tables:

Table 1. Association tests of individual SNPs

SNP	Allele	MAF[1] Ctrls (N=952)	MAF AIS (N=788)	OR [ci 0.95] AIS vs Ctrls	p-value	MAF severe (N=202)	OR [ci 0.95] Severe vs Ctrls	p-value
rs7893223*	C	0.41	0.35	0.77 [0.67-0.89]	0.00030	0.30	0.63 [0.50-0.79]	8×10^{-5}
rs11190870	C	0.42	0.37	0.82 [0.72-0.94]	0.00468	0.32	0.66 [0.52-0.83]	0.00030
rs11598177	G	0.46	0.42	0.88 [0.77-1.01]	0.06343	0.37	0.69 [0.55-0.86]	0.00106
rs1322331	C	0.46	0.43	0.88 [0.77-1.01]	0.06537	0.37	0.69 [0.55-0.86]	0.00092
rs594791	A	0.42	0.37	0.82 [0.72-0.94]	0.00536	0.32	0.65 [0.52-0.82]	0.00019
rs679206	G	0.45	0.42	0.90 [0.78-1.03]	0.11160	0.36	0.70 [0.56-0.87]	0.00159
rs678741*	G	0.45	0.42	0.90 [0.78-1.03]	0.11160	0.36	0.70 [0.56-0.87]	0.00159
rs11190878*	G	0.40	0.34	0.78 [0.68-0.89]	0.00037	0.29	0.62 [0.49-0.78]	5×10^{-5}

The * denotes SNPs imputed using BEAGLE 3 using the 1000-genomes European cohort phased genotypes

SNPs are put in order according to their position on chromosome 10 from left to right. [1]MAF: Minor allele frequency

Table 2. Association tests of multi-SNP haplotypes

Haplotypes	Ctrls freq	AIS freq	OR [CI 0.95] AIS vs Ctrls	p-value	Severe freq	OR [CI 0.95] Severe vs Ctrls	p-value
TTAAGAAA	0.538	0.565	1.11 [0.97-1.28]	0.116	0.626	1.42 [1.14-1.78]	0.0016
CCGCAGGG	0.395	0.334	0.77 [0.67-0.88]	0.00018	0.285	0.64 [0.48-0.77]	$2x10^{-5}$
TTGCGGGA	0.037	0.050	1.40 [1.01-1.94]	0.043	0.042	1.20 [0.70-2.05]	0.509
TCGCAGGA	0.009	0.024	2.54 [1.47-4.42]	0.00046	0.017	1.86 [0.84-4.11]	0.145

Table 3. Dihaplotype association tests

Allele 1	Allele 2	Freq (N Controls)	Freq (N AIS)	Freq (N Severe AIS)	OR [CI 0.95] AIS vs Ctrls	p-value	OR [CI 0.95] Severe vs Ctrls	p-value
TTAAGAAA	TTAAGAAA	0.281 (268)	0.321 (253)	0.411 (83)	1.21 [0.99-1.49]	0.0727	1.78 [1.30-2.44]	0.0003
CCGCAGGG	CCGCAGGG	0.156 (149)	0.127 (100)	0.099 (20)	0.78 [0.59-1.02]	0.0706	0.59 [0.36-0.96]	0.0331
TTAAGAAA	CCGCAGGG	0.433 (412)	0.368 (290)	0.332 (67)	0.76 [0.63-0.92]	0.0061	0.65 [0.47-0.89]	0.0081
TTAAGAAA	TTGCGGGA	0.043 (41)	0.065 (51)	0.059 (12)	1.54 [1.01-2.35]	0.0444	1.40 [0.72-2.71]	0.3125
CCGCAGGG	TTGCGGGA	0.029 (28)	0.027 (21)	0.020 (4)	0.90 [0.51-1.60]	0.7290	0.67 [0.23-1.93]	0.4503
TTAAGAAA	TCGCAGGA	0.012 (11)	0.032 (25)	0.015 (3)	2.80 [1.37-5.73]	0.0032	1.29 [0.36-4.67]	0.6985

3. Article 3

Association of Circulating YKL-40 Levels and *CHI3L1* Gene Variants with Idiopathic Scoliosis and the Risk of Spinal Deformity Progression

Dina Nada[1,2], Cédric Julien[1], Pierre H. Rompré[3], Marie-Yvonne Akoume[1], Kristen F. Gorman[1,4] Mark E. Samuels[5,6], Emile Levy[5,7] Alain Moreau[1,2,8,9,*]

[1] Viscogliosi Laboratory in Molecular Genetics of Musculoskeletal Diseases, Sainte-Justine University Hospital, Research Center, Montreal, QC, Canada

[2] Program of Biomedical Sciences, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada

[3] Faculty of Dentistry, Université de Montréal, Montreal, QC, Canada

[4] Department of Biological Sciences, California State University, Chico, CA, USA

[5] Sainte-Justine University Hospital Research Center, Montreal, QC, Canada

[6] Department of Medicine, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada

[7] Department of Nutrition, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada

[8] Department of Biochemistry and Molecular Medicine, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada

[9] Department of Stomatology, Faculty of Dentistry, Université de Montréal, Montreal, QC, Canada

*Corresponding author: Alain Moreau PhD,
E-mail: alain.moreau@recherche-ste-justine.qc.ca

Author contributions

Dina Nada: participated in design of the study, performed experimental work, participated in statistical analyses of data, and as lead author of the manuscript.

Cédric Julien and Pierre Rompré: performed statistical analyses of data.

Marie-Yvonne Akoume: participated in functional assesment

Kristen Fay Gorman: participated in active discussions and analyses of data

Mark E Samuels: participated in and supervised genomics experiments and participated in writing of the manuscript.

Emile Levy: supervised acylated ghrelin dosage and participated in the writing of the manuscript

Alain Moreau: supervised and supported all aspects of the study, and participated in writing of the manuscript.

Association of Circulating YKL-40 Levels and *CHI3L1* Gene Variants with Idiopathic Scoliosis and the Risk of Spinal Deformity Progression

Dina Nada[1,2], Cédric Julien[1], Pierre H. Rompré[3], Marie-Yvonne Akoume[1], Kristen F. Gorman[1,4] Mark E. Samuels[5,6], Emile Levy[5,7] Alain Moreau[1,2,8,9,*]

[1] Viscogliosi Laboratory in Molecular Genetics of Musculoskeletal Diseases, Sainte-Justine University Hospital, Research Center, Montreal, QC, Canada

[2] Program of Biomedical Sciences, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada

[3] Faculty of Dentistry, Université de Montréal, Montreal, QC, Canada

[4] Department of Biological Sciences, California State University, Chico, CA, USA

[5] Sainte-Justine University Hospital Research Center, Montreal, QC, Canada

[6] Department of Medicine, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada

[7] Department of Nutrition, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada

[8] Department of Biochemistry and Molecular Medicine, Faculty of Medicine, Université de Montréal, Montreal, QC, Canada

[9] Department of Stomatology, Faculty of Dentistry, Université de Montréal, Montreal, QC, Canada

*Corresponding author: Alain Moreau PhD,
E-mail: alain.moreau@recherche-ste-justine.qc.ca (AM)

Short title: Role of YKL-40 in Idiopathic Scoliosis Pathogenesis

Key Words: Idiopathic Scoliosis, *CHI3L1* gene, YKL-40, polymorphisms, French-Canadians, Spinal deformity progression

Abstract

Background

This study aimed to investigate the relationship of circulating YKL-40 levels, a secreted glycoprotein encoded by *chitinase 3-like 1* (*CHI3L1*) gene and single nucleotide polymorphisms (SNPs) regulating its expression with the risk of spinal deformity progression in idiopathic scoliosis (IS) patients in French-Canadian population.

Methods

Plasma samples and DNA obtained from 787 IS patients and 239 age- and gender-matched controls were used in the study. Enrolled participants were further subclassified according to their gender, biological endophenotypes and scoliosis severity phenotype. Circulating concentrations of YKL-40 was determined by ELISA and 12 SNPS were analyzed by multiplex polymerase chain reaction and genotyping. Functional effects of YKL-40 was tested on primary osteoblasts obtained from IS patients by cellular dielectric spectroscopy (CDS) assay.

Results

Significant differences in circulating YKL-40 levels between the tested groups were observed, being significantly increased in IS males classified in biological endophenotype FG1 and IS patients exhibiting less severe spinal deformities ($<40°$). Among the 12 studied SNPs, we find a significant association between 8 SNPs (rs55700740, rs946259, rs880633, rs1538372, rs4950881, rs946261, rs946262, rs10920576) and YKL-40 plasma levels in IS patients. The same SNPs are significantly associated with YKL-40 plasma levels in non-severe cases. Treatment of primary osteoblast from IS patients with YKL-40 rescued Gi-coupled receptor signalling dysfunction induced by osteopontin, a cytokine being proposed to play an active role in IS pathomechanism.

Conclusion

Collectively our findings suggest that YKL-40 could act as a protecting disease-modifying factor in the context of IS in a subgroup of patients by interfering with OPN signaling activity.

Introduction

Idiopathic scoliosis is a prevalent spinal deformity that affects an average of 1-4% of the global pediatric population [1]. It is characterized by an abnormal three-dimensional curvature of the spine with an onset that can occur between birth and sexual maturity. Thus, it has been classified as infantile, juvenile, or adolescent based on when a curve first presents itself [2]. Adolescent Idiopathic Scoliosis (AIS) is the most common form of scoliosis and occurs between the ages of 10 and 15 years, with girls affected more severely than boys [3]. Although the etiology of IS remains unclear, the fact that the syndrome is influenced by genetic factors has been widely accepted [4, 5]. The great phenotypic heterogeneity of IS, and that multiple loci have been identified in genetic studies, [6] suggests that the IS is more likely to be multifactorial.

Phenotypic complexity and possible genetic heterogeneity have complicated efforts to articulate our understanding of IS etiology using traditional genetic approaches. Previous work demonstrated that IS patients have a distinctive systemic signaling dysfunction for G inhibitory (Gi)-coupled receptors [7, 8]. The differential Gi signaling dysfunction among IS patients allow their classification into three distinct biological endophenotypes (FG1, FG2 and FG3), based on the maximum Gi signaling response in cells (osteoblasts and other cell types) exposed to Gi specific stimuli [9]. The use of endophenotypes in complex diseases has the advantage of partitioning the genetic variation, thus giving a study greater power to detect a genetic effect. "Endophenotype" is a term that was first introduced to genetics in 1966 to describe "microscopic and internal" characteristics while phenotypes describe "obvious and external" characteristics [10, 11]. Endophenotypes are heritable traits that are representative of the molecular path from genes to the phenotype [12, 13]. In the course of identifying genes involved in the differential

signaling response occurring in IS, we identified *CHI3L1* as one of the genes that showed a significant differential expression among the different IS endophenotypes (Gorman *et al* 2017 manuscript submitted).

CHI3L1 gene encodes for the secretory factor YKL-40. YKL-40 is a member of the family "mammalian chitinase-like proteins," which are secreted glycoproteins that bind to heparin. YKL-40 was first discovered in 1989 were it was secreted *in vitro* by MG63 osteosarcoma cell lines in large amounts [14]. YKL-40 is expressed in many tissues and is secreted by several types of solid tumors. The exact function of YKL-40 in normal tissues and in certain pathological conditions remains unknown. YKL-40 acts as a growth factor in cells involved in tissue remodelling. YKL-40 may have a role in cancer cell proliferation, survival and their ability to invade surrounding tissue [15]. In addition, elevated serum levels of YKL-40 have also been observed in patients with non-malignant diseases of particular contexts of inflammation or tissue remodelling [16]. To see whether this differential expression of the *CHI3L1*, the gene encoding YKL-40, among the IS groups is reflected systematically, we analyzed YKL-40 plasma levels in a large cohort of IS patients which are distributed among the three biological endophenotypes. We also compared YKL-40 levels between IS patients and controls, and further sub-classified IS patients based on their severity and extended our comparison among non-severe and severe patients versus controls. Moreover we genotyped the same cohort for 12 SNPs to test the association of those SNPs to the different phenotypes and/or to the plasma levels of YKL-40 among the different sub classification of patients.

Materials and Methods

Study populations

A group of 787 French-Canadian IS patients and 239 age- and gender-matched controls were enrolled prospectively between January 2008 and December 2012 in three pediatric spine centers in Montreal and surrounding schools. All participants are residents of Quebec and of European descent. Each IS patient was clinically assessed by one of the orthopedic surgeons at the hospital. All healthy control subjects were screened by an orthopedic surgeon using the Adam's forward-bending test with a scoliometer. Any children with an apparent spinal curvature or family history of scoliosis were excluded from the control cohort. Plasma YKL-40 levels were measured in 710 patients and 227 controls.

IS patient biological endophenotypes were generated from primary osteoblasts or peripheral blood mononuclear cells (PBMCs) using cellular dielectric spectroscopy, as previously described [8, 9]. Through this classification, 145 patients were classified in the first biological endophenotype (FG1), 257 patients were classified in the second endophenotype (FG2) and 301 patients were classified in the third endophenotype (FG3). Upon classification of the patients based on their spinal deformity severity; 590 patients were considered as non-severe cases (Cobb 10°-39°) at the time of measuring the YKL-40 levels, while 120 patients were considered as severe cases (Cobb angle $\geq$ 40°). Among our cohort, 728 patients and 216 control subjects were successfully genotyped for 12 SNPs in the *CHI3L1* gene, which encodes for YKL-40. In this group, 146 patients were classified in endophenotype FG1, 240 in endophenotype FG2 and 286 in endophenotype FG3. Stratification in function of scoliosis severity phenotype was determined only in participants who has reached their skeletal maturity at their last visit, which allowed the selection in our prospective cohort of 132 patients as severe cases (Cobb angle $\geq$ 40°) and 227 patients as non-severe cases (Cobb angle 10°-39°).

Ethics statement

Informed written consent was obtained from parents or legal guardians of all participants, and minors gave their assent. The study was approved by the Institutional Review Board of Sainte-Justine University Hospital, The Montreal Children's Hospital, The Shriners Hospital for Children and McGill University as well as by The Affluent School Board and The English School Board of Montreal.

Measuring plasma YKL-40 and Ghrelin levels

Peripheral blood samples were collected in EDTA-treated tubes and then centrifuged. Plasma samples were collected then aliquoted and stored at $-80^{o}C$ until thawed and analyzed. The concentrations of plasma YKL-40 was measured by enzyme-linked immunosorbent assay kit (Quidel, San Diego, CA, USA) according to the protocol provided by the manufacturer. Unacylated ghrelin was measured in plasma of a subgroup of IS patient exhibiting a severe scoliosis (Cobb angle $\geq 40°$) and matched control subjects by an EIA kit (Cayman Chemicals, Ann Arbor, MI, USA) according to the manufacturer's specifications. Both assays were performed in duplicate and the mean values were taken for the analysis. The optical density was measured at 450nm using DTX880 microplate reader (Beckman Coulter, Brea, California, USA).

Genotyping of SNPs in the *CHI3L1* gene

Genomic DNA samples were derived from the peripheral blood of the subjects of the same cohort using PureLink® Genomic DNA kit (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Then they were genotyped for 12 SNPs in the region of *CHI3L1* gene. Part of the cohort was genotyped by the Illumina Human Omni 2.5M Bead Chip as part of a GWAS study previously done by our team at the McGill University and Genome Quebec Innovation Centre. We chose those SNPs due to the fact that their genotypes were already available for most of the cohort used for biochemical analysis.

Therefore the 12 SNPs which were present on the Illumina Human Omni 2.5M Bead Chip were further genotyped in the rest of the cohort using multiplex PCR at the McGill University and Genome Quebec Innovation. Multiplex PCR of the 12 SNPs was performed using standard procedures with 20ng of template genomic DNA and HotStarTaq DNA polymerase enzyme (QIAGEN). PCR reactions were run on the QIAxcel (QIAGEN) to assess the amplification, followed by the single base extension using iPlex Thermo Sequenase. Genotypes were determined by MALDI-TOF mass-spectrometry and data were analyzed using Mass ARRAY Typer Analyser software.

Cellular dielectric spectroscopy (CDS) assay

Functional effects of YKL-40 were investigated by a CDS assay as previously described [8, 9]. In brief, primary osteoblasts obtained from bone fragments obtained intraoperatively from IS patients and control subjects (trauma cases) were seeded into the CellKey^{TM} standard 96-well microplate at a density of 10 x 10^4 cells per well and incubated in standard conditions (37°C / 5% CO_2) with 0.5µg/ml of purified recombinant OPN (rOPN) or the vehicle (saline buffer) for 18h prior to stimulation. After overnight incubation, cells were directly stimulated with oxymetazoline (10µM) (Tocris Chemical Co. St. Louis, MO, USA), a specific ligand activating α1-adrenergic receptor normally coupled to Gi proteins. The same test was done with and without treatment the cells with recombinant YKL-40 (rYKL-40) to assess its effect.

Statistical analyses

For comparing patients versus controls and for comparing different sub classifications of patients with each other and controls, an Anova test was done with log YKL-40 as the dependant variable and the phenotype and gender as independent variables, with the age as a covariate. P-values were two sided t-test and $p<0.05$ was considered significant. The presented p-values were after Bonferroni adjustment

for pair wise comparisons. The prevalence of each genotype of each of the studied SNPs was calculated for the different sub-classification of patients and for controls. The genotypes of each SNP were compared between each two groups of patients and /or controls, and the significance is calculated using the Fisher exact test two-sided. Bonferroni adjustment was done for each separate multiple comparison. Statistical analysis was done using the software SPSS v.23. Analysis of the association of each SNP with the plasma YKL-40 levels in each group was performed using the 'q assoc' option in PLINK v1.09 software.

Results

Clinical and biochemical characteristics

A summary of demographic features, clinical profiles and plasma YKL-40 levels for our French-Canadian cohort is provided in **Table 1**. Plasma YKL-40 levels were analyzed in 710 patients and 227 controls. As expected, there were more females in patients than in controls (Fisher's exact test=0.001).

Table 1. Clinical and biochemical characteristics of participants.

Groups	Female Subjects			Male Subjects			All Subjects		
	Mean Age (Years)	Mean Cobb Angle (°)	YKL-40 (ng/ml)	Mean Age (Years)	Mean Cobb Angle (°)	YKL-40 (ng/ml)	Mean Age (Years)	Mean Cobb Angle (°)	YKL-40 (ng/ml)
All IS	13.7 ± 2.2 (6.9-26.0)	28 ± 16 (10-90)	32 ± 20 (3-326)	14.0 ± 2.1 (7.4-18.2)	22 ± 11 (10-72)	39 ± 36 (8-298)	13.8 ± 2.2 (6.9-26.0)	27 ± 16 (10-90)	33 ± 23 (3-326)
		N = 598			N = 112			N = 710	
Functional Group 1	13.5 ± 2.1 (6.9-18.7)	28 ± 15 (10-83)	30 ± 12 (5-69)	14.2 ± 1.6 (10.9-16.7)	27 ± 20 (10-76)	57 ± 62 (14-298)	13.6 ± 2.0 (6.9-18.7)	28 ± 16 (10-83)	34 ± 27 (5-298)
		N = 124			N = 21			N = 145	
Functional Group 2	13.8 ± 2.2 (7.3-19.1)	30 ± 17 (10-89)	32 ± 19 (3-183)	14.0 ± 2.6 (8.7-18.2)	24 ± 12 (10-56)	36 ± 39 (9-228)	13.9 ± 2.2 (7.3-19.1)	29 ± 17 (10-89)	32 ± 22 (3-228)
		N = 229			N = 28			N = 257	
Functional Group 3	13.6 ± 2.2 (7.7-26.0)	26 ± 15 (10-90)	33 ± 24 (4-326)	14.0 ± 2.3 (7.4-18.2)	22 ± 14 (10-66)	33 ± 16 (8-96)	13.7 ± 2.2 (7.4-26.0)	25 ± 14 (10-90)	33 ± 23 (4-326)
		N = 241			N = 60			N= 301	
Healthy Control Subjects	12.5 ± 3.3 (7.1-18.3)	NA	29 ± 13 (8-81)	12.4 ± 3.1 (7.2-17.6)	NA	28 ± 13 (4-90)	12.5 ± 3.2 (7.1-18.3)	NA	29 ± 13 (4-90)
		N = 124			N = 103			N = 227	

Association between plasma YKL-40 levels and IS biological endophenotypes

When all IS patients were compared to matched healthy controls, patients showed higher plasma YKL-40 levels than controls (p=0.002). Prior work has demonstrated that IS patients have a distinctive systemic signaling dysfunction for G inhibitory (Gi)-coupled receptors allowing their functional

classification in three distinct biological endophenotypes with respect to the impedance values: FG1 = [10-40 Ω], FG2 = [40-80 Ω], and FG3 = [80-120 Ω] (while healthy controls always exceed 120 Ω) [8]. We first tested the possible interaction between gender and endophenotype, which was found to be statistically significant (p=0.009). Therefore, we separated the analyses for females and males. By comparing females only, there was no significant difference in circulating YKL-40 levels among the different biological endophenotypes (FG1, FG2 and FG3). While upon comparing males only, there were significant differences among the biological endophenotypes (p=0.001). After Bonferroni adjusted pair wise comparisons, IS FG1 males showed higher levels than controls males (p=0.001) and IS FG3 males (p=0.042) respectively. So, changes observed in plasma YKL-40 levels replicate at the protein level our previous expression analyses using primary osteoblasts obtained from IS patients and matched healthy controls (Gorman *et al*. 2017 manuscript submitted).

Association between plasma YKL-40 levels and scoliosis severity phenotype

We further classified the IS patients into severe cases (Cobb $\geq 40^{\circ}$) or non-severe ones (Cobb $<40^{\circ}$) to assess for possible associations between plasma YKL-40 levels and scoliosis severity phenotype. The p-value for the interaction between gender and IS cases was not statistically significant. Bonferroni adjusted pair wise comparisons between both IS cases groups and controls shows a statistically significant elevation of plasma YKL-40 levels in the non-severe IS cases when compared with controls (p=0.003).

Association between plasma ghrelin and YKL-40 levels

Given that serum YKL-40 levels were previously reported to be inversly correlated with ghrelin [17] and the fact that a significantly higher circulating ghrelin levels was previously reported in IS [18], we measured plasma ghrelin levels in a subset of IS patients and matched healthy controls. Analysis of all IS patients compared to matched controls did not show a significant effect of circulating ghrelin

levels on plasma YKL-40 levels. However, when IS patients were stratified according to their biological endophenotypes, mean plasma ghrelin level was significantly lowered only in FG1 endophenotype samples (99.9 ± 44.9 pg/ml) when compared with the mean control value (162.8 ± 63.9 pg/ml; p= 0.028). Consistently, decreased ghrelin concentration in blood of AIS patients classified in FG1 endophenotype could contribute in part in the elevation of plasma YKL-40 in this AIS subgroup.

Genetic variants in *CHI3L1* gene and plasma YKL-40 levels

To determine whether the *CHI3L1* genotypes affected circulating YKL-40 levels, 12 SNPs were analyzed in the same cohort involved in plasma YKL-40 measurement. Our results showed that 8 SNPs were significantly associated with the plasma YKL-40 levels in IS patients (**Table 2**): rs55700740 ($p=3.85 \times 10^{-5}$), rs946259 ($p=3.87 \times 10^{-5}$), rs880633 ($p=3.82 \times 10^{-5}$), rs1538372 ($p=5.9 \times 10^{-6}$), rs4950881 ($p=6.12 \times 10^{-4}$), rs946261 (4.36×10^{-8}), rs10920579 ($p=1.09 \times 10^{-8}$) and the highest association was displayed by rs946262 ($p=6 \times 10^{-12}$). On the other hand, only two of these SNPs were associated with plasma YKL-40 levels in the healthy control group; rs1538372 ($p=5.7 \times 10^{-4}$) and rs946261 ($p=0.0018$). Distribution of those SNPs in function of biological endophenotypes and variation in plasma YKL-40 levels revealed no significant associations in FG1 endophenotype, while some SNPs showed significant associations in FG2 and FG3 endophenotypes; rs946261 ($p=0.0001$ and $p=0.00015$ respectively), rs946262 ($p=7 \times 10^{-6}$ and $p=1.48 \times 10^{-6}$ respectively), and rs10920579 ($p=0.0013$ and $p=2.2 \times 10^{-5}$ respectively). Other SNPs showed specific associations only with IS patients classified in FG2 endophenotype: rs55700740 ($p=0.00052$), rs946259 ($p=0.00069$), rs880633 ($p=0.000689$), rs1538372 ($p=0.000634$) (**Table 3**).

Table 2. Prevalence of the studied SNPs in *CHI3L1* gene and their associations with plasma YKL-40 levels in IS patients and healthy controls

SNP	Controls			IS		
	(N=216)	Mean YKL-40 (ng/ml)±SD (95% CI)	p VALUE	(N=728)	Mean YKL-40 (ng/ml)±SD (95% CI)	p VALUE
rs55700740			0.028			3.85×10^{-5}
CC	44(20.4%)	25.98±11.6(22.38-29.58)		167(23%)	26.93±12.2(24.93-28.93)	
CA	111(51.4%)	27.52±10.6(25.52-29.52)		384(52.9%)	33.8±28.4(30.8-36.8)	
AA	61(28.2%)	31.48±15.9(27.28-35.68)		175(24.1%)	38.3±19.6(35.1-41.5)	
rs7542294			0.11			0.952
GG	160(74.4%)	27.58±12.3(25.58-29.58)		494(67.9%)	33.23±22.9(31.23-35.23)	
GA	52(24.2%)	30.28±12.9(26.48-34.08)		212(29.1%)	33.5±26.9(29.5-37.5)	
AA	3(1.4%)	35.44±23.6(8.24-62.64)		22(3%)	31.65±11.1(26.85-36.45)	
rs946259			0.026			3.87×10^{-5}
GG	43(20%)	25.9±11.8(22.3-29.5)		164(22.6%)	26.75±12.2(24.75-28.75)	
GA	109(50.7%)	27.58±10.6(25.58-29.58)		386(53.1%)	33.8±28.3(30.8-36.8)	
AA	63(29.3%)	31.44±15.6(27.44-35.44)		177(24.3%)	38.18±19.5(35.18-41.18)	
rs880633			0.025			3.82×10^{-5}
GG	63(29.2%)	38.18±19.5(35.18-41.18)		177(24.3%)	31.44±15.6(27.44-35.44)	
GA	110(50.9%)	33.77±28.2(30.77-36.77)		387(53.2%)	27.48±10.6(25.48-29.48)	
AA	43(19.9%)	26.75±12.2(24.75-28.75)		164(22.5%)	25.9±11.8(22.3-29.5)	
rs1538372			**0.0005**			**5×10^{-6}**
GG	97(44.9%)	31.38±15(28.18-34.58)		318(44.2%)	37.6±25.4(34.6-40.6)	
GA	98(45.4%)	26.64±9.9(24.64-28.64)		327(45.4%)	31.27±23.5(28.47-34.07)	
AA	21(9.7%)	22.44±8.6(18.64-26.24)		75(10.4%)	24.19±12.3(21.19-27.19)	
rs4950881			0.029			**0.0006**
GG	12(5.6%)	25.62±9.2(20.22-31.02)		48(6.6%)	22.27±12(18.67-25.87)	
GA	80(37.2%)	26.13±9(20.73-31.53)		256(35.2%)	31.63±26.1(28.23-35.03)	
AA	123(57.2%)	30.14±14.7(27.34-32.94)		423(58.2%)	35.38±22.9(32.98-37.78)	
rs10399805			0.124			0.523
GG	164(76.3%)	27.53±12.3(25.53-29.53)		518(71.4%)	32.8±22.4(30.8-34.8)	
GA	49(22.8%)	31.34±13.4(27.34-35.34)		189(26.1%)	34.43±28.1(30.23-38.63)	
AA	2(0.9%)	23.93±3(19.73-28.13)		18(2.5%)	32.83±11.2(24.63-41.03)	
rs6691378			0.107			0.482
GG	166(77.2%)	27.73±12.5(25.73-29.73)		520(71.5%)	32.75±22.4(30.75-34.75)	
GA	47(21.9%)	29.9±12(26.3-33.5)		189(26%)	35±28.1(30.8-39.2)	
AA	2(0.9%)	42.02±28.6(1.61-82.41)		18(2.5%)	31.03±8.9(26.83-35.23)	
rs946261			0.036			**4.36×10^{-8}**
AA	79(38.2%)	30.44±12.5(27.64-33.24)		258(35.4%)	39.42±26.9(35.82-43.02)	
AG	107(51.7%)	26.44±10.2(24.4-28.4)		355(48.8)	31.5±23.3(28.9-34.1)	
GG	21(10.1%)	26.4±13.7(20.4-32.4)		115(15.8%)	25±11.1(23-27)	
rs946262			**0.002**			**6×10^{-12}**
GG	146(67.6%)	30.1±13.9(27.7-32.5)		477(65.8%)	37.8±26.7(35.19-40.39)	
GA	65(30.1%)	25.14±8.6(23.14-27.14)		224(30.9%)	25.92±13.8(23.92-27.92)	
AA	5(2.3%)	19.16±6.4(13.36-24.96)		24(3.3%)	15.83±9.5(11.83-19.83)	
rs116415868			0.816			0.645
GG	214(99.1%)	28.32±12.6(26.52-30.12)		711(97.8%)	33.33±24.1(31.33-35.33)	
GA	2(0.9%)	26.26±13.5(7.26-45.26)		16(2.2%)	30.44±8.5(26.04-34.84)	
rs10920576			0.01			**1.09×10^{-8}**
GG	158(73.5%)	29.68±13.6(27.48-31.88)		518(71.3%)	36.38±26(33.98-38.78)	
GA	53(24.7%)	25±9(22.4-27.6)		187(25.7%)	26.3±14.4(24.1-28.5)	
AA	4(1.9%)	21.6±3.9(7.6-35.6)		22(3%)	16.27±9.9(11.67-20.87)	

Table 3. Prevalence of the studied SNPs in *CHI3L1* gene and their associations with plasma YKL-40 levels in function of IS biological endophenotypes

SNP	FG1			FG2			FG3		
	(N=146)	Mean YKL-40 (ng/ml)±SD (95% CI)	p VALUE	(N=240)	Mean YKL-40 (ng/ml)±SD (95% CI)	p VALUE	(N=286)	Mean YKL-40 (ng/ml)±SD (95% CI)	p VALUE
rs55700740			0.186			**0.0005**			0.035
CC	27(18.5%)	27.92±13.6(22.52-33.32)		65(27.3%)	25.12±9.5(22.72-27.52)		66(23.1%)	28.33±14(24.73-31.93)	
CA	78(53.4%)	34.18±34.3(26.18-42.18)		123(51.7%)	32.98±26.5(28.18-37.78)		153(53.5%)	34.2±27.5(29.6-38.8)	
AA	41(28.1%)	37.47±18.6(31.67-43.27)		50(21%)	40.16±24.4(32.96-47.36)		67(23.4%)	36.88±16.5(32.88-40.88)	
rs7542294			0.613			0.68			0.98
GG	90(61.6%)	32.79±17.8(28.99-36.59)		170(70.8%)	32.45±21.8(29.05-35.85)		196(68.5)	33.66±26(29.86-37.46)	
GA	53(36.3%)	36.5±40(25.1-47.9)		60(25%)	32.02±27.9(24.62-39.42)		81(28.3%)	32.85±13.7(29.65-36.05)	
AA	3(2.1%)	28.26±8.4(18.66-37.86)		10(4.2%)	28.7±6.4(24.7-32.7)		9(3.1%)	36.07±14.9(26.07-46.07)	
rs946259			0.19			**0.0007**			0.03
GG	27(18.6%)	27.92±13.6(22.52-33.32)		65(27.1%)	25.1±9.6(22.7-27.5)		63(22%)	27.94±14(24.34-31.54)	
GA	77(53.1%)	34.33±34.5(26.33-42.33)		123(51.2%)	32.84±26.4(28.04-37.64)		156(54.5%)	34.24±27.2(29.84-38.64)	
AA	41(28.3%)	37.47±18.6(31.67-43.27)		52(21.7%)	39.69±24(32.89-46.49)		67(23.4%)	36.88±16.5(32.88-40.88)	
rs880633			0.19			**0.0007**			0.03
GG	41(28.1%)	37.47±18.6(31.47-43.47)		52(21.7%)	39.69±24(32.89-46.49)		67(23.4%)	36.88±16.5(32.88-40.88)	
GA	78(53.4%)	34.18±34.3(26.18-42.18)		123(51.2%)	32.84±26.4(28.04-37.64)		156(54.5%)	34.24±27.2(29.84-38.64)	
AA	27(18.5%)	27.92±13.6(22.52-33.32)		65(27.1%)	25.1±9.6(22.7-27.5)		63(22%)	27.94±14(24.34-31.54)	
rs1538372			0.045			**0.0006**			0.015
GG	79(54.5%)	38.53±34.3(30.53-46.53)		83(35.6%)	38.47±29.3(31.87-45.07)		125(43.7%)	36.35±15.3(33.55-39.15)	
GA	56(38.6%)	29.2±15.5(24.99-33.39)		123(52.8)	30.31±19.4(26.71-33.91)		128(44.8%)	32.88±29.8(27.48-38.28)	
AA	10(6.9%)	26.51±15.3(16.31-36.71)		27(11.6%)	22.04±9.4(18.44-25.64)		33(11.5%)	25.24±13.5(20.44-30.04)	
rs4950881			0.079			0.016			0.091
GG	4(2.7%)	16.66±9.2(7.46-25.86)		17(7.1%)	21.46±9.4(16.66-26.26)		23(8.1%)	23.82±14(18.02-29.62)	
GA	51(34.9%)	30.39±16.6(25.9-35.19)		94(39.2%)	30.14±21.5(25.54-34.74)		97(34%)	33.66±33.8(26.66-40.66)	
AA	91(62.3%)	36.79±32.2(26.79-43.79)		129(53.8%)	35.08±24.8(30.68-39.48)		165(57.9%)	34.53±14(32.33-36.73)	
rs10399805			0.436			0.93			0.591
GG	95(65.1%)	32.37±17.6(28.77-35.97)		177(74.1%)	32.15±21.5(28.75-35.55)		208(73.2%)	33.07±25.2(29.47-36.67)	
GA	49(33.6%)	37.8±41.5(25.4-50.2)		53(22.2%)	33.3±29(25.3-41.3)		69(24.3%)	33.16±14.3(29.76-36.56)	
AA	2(1.4%)	23.54±2.7(19.54-27.54)		9(3.8%)	28.4±6.8(23.8-33)		7(2.5%)	42.57±12.1(32.57-52.57)	
rs6691378			0.378			0.972			0.599
GG	95(65.1%)	32.19±17.6(26.19-38.19)		178(74.2%)	31.98±21.5(28.58-35.38)		209(73.3%)	33.25±25.2(29.65-36.85)	
GA	49(33.6%)	38.17±41.4(25.77-50.57)		52(21.7%)	33.53±29.3(25.53-41.53)		70(24.6%)	34.21±14.8(30.61-37.81)	
AA	2(1.4%)	23.54±2.7(19.74-27.34)		10(4.2%)	28.44±6.4(24.44-32.44)		6(2.1%)	39.2±9.9(30.4-48)	
rs946261			0.133			**0.0001**			**0.0001**
AA	54(37%)	36.22±18.3(31.22-41.22)		77(32.1%)	40.92±23.7(35.32-46.52)		106(37.1%)	39.89±32.8(33.5-46.3)	
AG	70(47.9%)	35.52±35.7(26.72-44.32)		115(47.9%)	29.41±24.8(24.6-34.2)		140(49%)	30.95±12.6(29-33)	
GG	22(15.1%)	23.12±9.9(18.72-27.52)		48(20%)	25.33±10.3(22.33-28.33)		40(14%)	25.42±12.9(21.22-29.62)	
rs946262			0.03			**7x10^{-6}**			**1.5x10^{-6}**
GG	103(70.5%)	36.95±30.7(31-43)		138(57.7%)	37.87±25.6(33.47-40.07)		195(68.7%)	38.01±25.8(36.11-39.91)	
GA	39(26.7%)	28±15.6(22.8-33.2)		93(38.9%)	25.48±16.6(22.08-28.88)		79(27.8%)	25.23±8.6(23.23-27.23)	
AA	4(2.7%)	14.9±8.5(6.5-23.3)		8(3.3%)	14.61±6(10.01-19.21)		10(3.5%)	17.06±12.3(13.16-20.96)	
rs116415868			0.865			0.977			0.583
GG	145(99.3%)	34.04±27.7(29.24-38.84)		235(97.9%)	32.17±23.2(29.17-35.17)		276(96.8%)	33.66±23.2(30.86-36.46)	
GA	1(0.7%)	29.31		5(2.1%)	32.5±7.7(25.7-39.3)		9(3.2%)	29.42±9.6(23.02-35.82)	
rs10920576			0.036			**0.0013**			**2.3x10^{-5}**

113

GG	109(74.7%)	36.55±30.1(30.75-42.35)	162(67.5%)	35.37±24.6(31.37-39.37)	206(72.3%)	36.83±25.3(33.23-40.43)
GA	33(22.6%)	27.63±16(21.83-33.43)	71(29.6%)	26.5±18.2(22.1-30.9)	70(24.6%)	25.27±8.6(23.27-27.27)
AA	4(2.7%)	14.9±8.5(6.4-23.4)	7(2.9%)	15.14±6.4(9.94-20.34)	9(3.2%)	17.64±12.8(9.04-26.24)

$p \leq 0.001$ is considered significant

Stratification of *CHI3L1* gene variants in function of scoliosis severity phenotype

Interestingly, stratification of the selected SNPs in function of scoliosis severity phenotype showed that the same 8 SNPs associated with plasma YKL-40 levels in IS patients were significantly associated with the non-severe cases: rs55700740 (p=0.000631), rs946259 (p=0.00064), rs880633 (p=0.00063), rs1538372 (p=1.94×10^{-5}), rs4950881 (p=0.00011), rs946261 (p=2.2×10^{-5}), rs10920579 (p=6.5×10^{-6}) and the same SNP rs946262 still showed the highest association (p=6.2×10^{-9}). While in the severe group only three SNPs showed marginal associations; rs55700740 (p=0.008), rs946259 (p=0.008), rs880633 (p=0.008) (**Table 4**).

114

Table 4. Prevalence of the studied SNPs in *CHI3L1* gene and their associations with plasma YKL-40 levels in function of scoliosis severity

SNP	IS severe (Cobb ≥ 40°) N=132	Mean YKL-40 (ng/ml)±SD (95% CI)	p VALUE	IS non-severe (Cobb < 40°) N=227	Mean YKL-40 (ng/ml)±SD (95% CI)	p VALUE
rs55700740			0.008			**0.0006**
CC	38(28.8%)	28.67±12(24.27-33.07)		57(25.1%)	24.71±10.3(21.91-27.51)	
CA	68(51.5%)	29.53±14.6(25.53-33.53)		107(47.1%)	33.12±14.3(30.12-36.12)	
AA	26(19.7)	40.17±11.8(35.17-45.17)		63(27.8%)	39.38±20.2(33.78-44.98)	
rs7542294			0.651			0.69
GG	93(70.5%)	31.87±14.7(28.67-35.07)		152(67%)	33.34±18.6(30.14-36.54)	
GA	39(29.5%)	30.26±11.7(25.66-34.86)		67(29.5%)	31.33±9.5(28.93-33.73)	
AA	0			8(3.5%)	29.4±6.8(24.6-34.2)	
rs946259			0.008			**0.0006**
GG	37(28%)	28.25±11.9(23.85-32.65)		57(25.2%)	24.7±10.3(21.9-27.5)	
GA	68(51.5%)	29.83±14.7(25.83-33.83)		106(46.9%)	33.22±14.4(30.22-36.22)	
AA	27(20.5%)	39.42±12.1(34.42-44.42)		63(27.9%)	39.38±20.2(33.78-44.98)	
rs880633			0.008			**0.0006**
GG	27(20.5%)	39.42±12.1(34.42-44.42)		63(27.8%)	39.38±20.2(33.78-44.98)	
GA	68(51.5%)	29.83±14.7(25.83-33.83)		107(47.1%)	33.12±14.3(30.12-36.12)	
AA	37(28%)	28.25±11.9(23.85-32.65)		57(25.1%)	24.7±10.3(21.9-27.5)	
rs1538372			0.012			**2×10^{-5}**
GG	50(37.9%)	37±12.8(33-41)		100(45%)	37.21±16.3(33.61-40.81)	
GA	59(44.7%)	28.33±14.2(24.33-32.33)		97(43.7%)	30.69±15(27.49-33.89)	
AA	23(17.4%)	28.68±13.1(22.88-34.48)		25(11.3%)	21.86±12.1(16.86-26.86)	
rs4950881			0.012			**0.0001**
GG	17(12.9%)	25.06±11.9(19.06-31.06)		15(6.6%)	18.7±11(12.9-24.5)	
GA	52(39.4%)	29.85±15.3(25-34.6)		80(35.2%)	29.94±15.7(26.34-33.54)	
AA	63(47.7%)	34.77±12.8(31.17-38.37)		132(58.1%)	36.02±15.8(33.02-39.02)	
rs10399805			0.695			0.353
GG	99(75.6%)	31.26±14.2(28.26-34.26)		165(72.7%)	33.1±17.9(30.09-36.09)	
GA	30(22.9%)	31.5±12.5(26.09-36.89)		56(24.7%)	31.35±10.2(28.35-34.35)	
AA	1(0.8%)	20.05		6(2.6%)	30.64±6.8(25.04-36.24)	
rs6691378			0.626			0.364
GG	100(75.8%)	31.60±14.4(28.61-34.61)		165(72.7%)	33.04±17.9(30.04-36.04)	
GA	30(22.7%)	31.49±12.5(26.09-36.89)		56(24.7%)	31.48±10.2(28.48-34.48)	
AA	1(0.8%)	20.05		6(2.6%)	30.64±6.8(25.04-36.24)	
rs946261			0.079			**2.2×10^{-5}**
AA	45(34.1%)	34.52±13.9(29.92-39.12)		87(38.3%)	40.3±19.7(35.7-44.9)	
AG	67(50.8%)	30.62±14.4(26.62-34.62)		102(44.9%)	29.4±10.8(27.2-31.6)	
GG	20(15.2%)	27.11±11.6(21.11-33.11)		38(16.7%)	23.79±11(19.99-27.59)	
rs946262			0.023			**6.23×10^{-9}**
GG	78(60%)	34.67±13.1(31.27-38.07)		149(65.9%)	37.46±16.8(34.46-40.46)	
GA	47(36.2%)	27.93±14.3(23.33-32.53)		69(30.5%)	25.5±9(23.3-27.7)	
AA	5(3.8%)	26.35±15.2(11.15-41.55)		8(3.5%)	9.13±3.7(7.73-10.53)	
rs116415868			0.921			0.619
GG	131(99.2%)	31.49±14.1(28.69-34.29)		223(98.2%)	32.7±16.2(30.3-35.1)	
GA	1(0.8%)	29.94		4(1.8%)	27.12±14.1(10.12-41.12)	
rs10920576			0.08			**6.5×10^{-6}**
GG	85(64.4%)	33.26±13.4(30.06-36.46)		164(72.2%)	35.67±17(32.87-38.47)	

GA	41(31.1%)	29.1±14.68(24.1-34.1)	57(25.1%)	26.23±8.7(23.83-28.63)
AA	6(4.5%)	23.49±14.7(10.29-36.69)	6(2.6%)	8.81±4.3(5.01-12.61)

p ≤ 0.002 is considered significant

Association of *CHI3L1* gene variants and different sub classifications of IS patients

None of the studied SNPs showed a significant association with the disease when IS cases were compared to the matched control group. However, rs1538372 showed a significant difference when IS biological endophenotypes were compared. Indeed, this variant is more strongly associated with IS patients classified in endophenotype FG1 when compared to IS cases classified in FG2 after Bonferroni correction (p=0.001) (**Supplementary Table S1**). All other SNPs did not show any significant difference among different IS endophenotypes and controls. Neither did any of the SNPs showed any significant difference in function of scoliosis severity (**Supplementary Table S2**). Upon gender separation two SNPs showed a significant difference between the severe males and control males; rs946262, rs10920576 (p=0.012, and p=0.005 respectively) after Bonferroni correction.

Functional assessment of the role of YKL-40 in IS pathogenesis

We have previously demonstrated the occurrence of a differential Gi-coupled receptor signalling dysfunction in primary osteoblasts and other cell types obtained from IS patients that led to the identification of three biological endophenotypes associated with IS as measured by CDS assay [8, 9]. To examine the possible functional impact of increased plasma YKL-40 levels, primary osteoblasts from three scoliotic patients were screened for their response to oxymetazoline (10μM), a GiPCR selective agonist activating α1-adrenergic receptor normally coupled to Gi proteins as cellular readout. In agreement with our previous results, exposure to rOPN induced a reduction of α1-adrenergic receptor signalling while treatment with purified YKL-40 rescued partially or completely the signalling dysfunction induced by rOPN suggesting that elevation of YKL-40 could attenuate scoliosis severity.
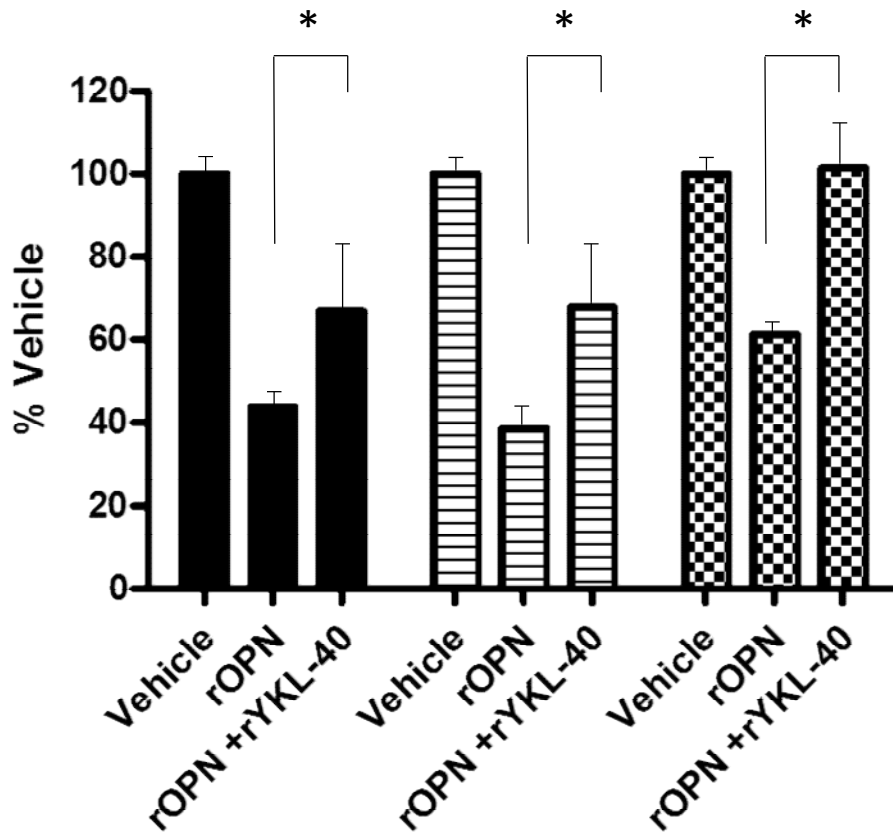
Figure 1. YKL-40 rescues Gi-coupled receptor signalling defect induced by rOPN. Primary osteoblasts obtained from three scoliotic patients were pre-treated with purified rOPN (0.5 µg/ml) with and without rYKL-40 (0.5 µg/ml) for 18 h prior to the stimulation with 10 µM of oxymethazolin. Error bars show SEM of independent experiments performed three times in duplicate. Data represent the percentage of the maximum impedance measured by CDS assay and were normalised to the response achieved in the absence of rOPN (vehicle only). * P < 0.01 based on one-way ANOVA followed by in post-hoc test of Dunnett.

Discussion

To the best of our knowledge, the present study is the first to show a significant association between plasma YKL-40 levels, *CHI3L1* gene single nucleotide polymorphisms and reduced susceptibility to the development of severe spinal deformities in the context of IS. A recent global expression analysis performed with primary osteoblasts obtained from IS patients showed a significant overexpression of *CHI3L1* gene, encoding for the circulating factor YKL-40, in a subgroup of IS patients (biological endophenotype FG1), which drew our attention given that the IS patients classified into the two other biological endophenotypes were more prone to develop a severe scoliosis. By comparing plasma YKL-40 levels of each biological endophenotype of IS patient versus controls, and since the interaction between gender and group was statistically significant, we separated the analyses for females and males. Surprisingly, for females no significant difference between the groups was observed, whereas in males there were significant differences among the groups (p=0.001). Specifically, males classified in biological endophenotype FG1 showed significantly higher plasma YKL-40 levels than controls and the two other IS endophenotypes. Thus, these results replicate and confirmed our recent expression analyses using primary human osteoblasts (Gorman *et al*., 2017 manuscript submitted). Furthermore, when IS patients were classified based on their spinal deformity severity, the non-severe group showed statistically significant higher levels of YKL-40 than controls. To recapitulate, there are higher circulating YKL-40 levels in non-severe cases of IS as well as in male IS patients classified in the biological endophenotype FG1, which is the IS subgroup least likely to develop a severe curvature when compared to the other endophenotypes [19]. It is also widely known that males are less likely to develop severe forms of the disease when compared to females. Collectively our data strongly suggest that YKL-40 could play a protective role in the context of IS.

Previous studies on SNPs regulating circulating YKL-40 levels have demonstrated that genetic variations of *CHI3L1* gene have an impact on plasma YKL-40 levels, both in healthy subjects as well as in patients suffering different diseases from asthma [20] to rheumatoid arthritis [21]. Indeed, 8 of the 12 studied SNPs were associated with plasma YKL-40 levels in IS patients while only 2 of them were associated with YKL-40 plasma levels in healthy controls. Interestingly, the same 8 SNPs showed significant associations with YKL-40 plasma levels in the non-severe group. Most of those SNPs were shown in previous studies to be associated with YKL-40 levels and/or *CHI3L1* expression. One limitation of our study is that we missed two important SNPs that were shown to be associated with circulating YKL-40 levels; rs4950928 and rs10399931 [22, 23].

In conclusion, we found a positive correlation of plasma YKL-40 levels with non-severe form of scoliosis as well as with male patients classified in IS endophenotype FG1, who are less prone to spinal deformity progression. We also found associations between certain SNPs of *CHI3L1* gene and plasma YKL-40 levels. Positive associations between SNPs of *CHI3L1* gene and some IS endophenotypes (FG2 and FG3) were observed although none of the SNPs tested could explain the higher YKL-40 levels in patients classified in IS endophenotype FG1. Further studies must be undertaken to elucidate this. Our functional *in vitro* analysis strongly suggests that elevation of YKL-40 could reduce the severity of scoliosis by interfering with Gi-signaling dysfunction induced by osteopontin (OPN) in IS. The role of OPN in scoliosis development in humans and different animal models has more recently been reported by different groups [24-27]. Collectively our findings suggest that YKL-40 could act as a protecting disease-modifying factor in the context of IS.

Acknowledgements

Competing interests

This work led to a patent application (pending) own by Sainte-Justine University Hospital (CHU Sainte-Justine). The authors declare no other potential conflicts of interest.

References

1. Cheng JC, Castelein RM, Chu WC, Danielsson AJ, Dobbs MB, Grivas TB, et al. PRIMER. NATURE. 2015.
2. Donzelli S, Zaina F, Lusini M, Minnella S, Negrini S. In favour of the definition" adolescents with idiopathic scoliosis": juvenile and adolescent idiopathic scoliosis braced after ten years of age, do not show different end results. SOSORT award winner 2014. Scoliosis. 2014;9(1):7.
3. Asher MA, Burton DC. Adolescent idiopathic scoliosis: natural history and long term treatment effects. Scoliosis. 2006;1(1):2.
4. Tang NL, Yeung HY, Hung VW, Di Liao C, Lam TP, Yeung HM, et al. Genetic epidemiology and heritability of AIS: A study of 415 Chinese female patients. Journal of orthopaedic research. 2012;30(9):1464-9.
5. Grauers A, Rahman I, Gerdhem P. Heritability of scoliosis. European Spine Journal. 2012;21(6):1069-74.
6. Gorman KF, Julien C, Moreau A. The genetic epidemiology of idiopathic scoliosis. European Spine Journal. 2012;21(10):1905-19.
7. Azeddine B, Letellier K, Moldovan F, Moreau A. Molecular determinants of melatonin signaling dysfunction in adolescent idiopathic scoliosis. Clinical orthopaedics and related research. 2007;462:45-52.

8. Akoume M-Y, Azeddine B, Turgeon I, Franco A, Labelle H, Poitras B, et al. Cell-based screening test for idiopathic scoliosis using cellular dielectric spectroscopy. Spine. 2010;35(13):E601-E8.

9. Akoume M-Y, Franco A, Moreau A. Cell-based assay protocol for the prognostic prediction of idiopathic scoliosis using cellular dielectric spectroscopy. JoVE (Journal of Visualized Experiments). 2013;(80):e50768-e.

10. John B, Lewis KR. Chromosome variability and geographic distribution in insects. Science. 1966;152:711-21.

11. Ertekin-Taner N. Gene expression endophenotypes: a novel approach for gene discovery in Alzheimer's disease. Molecular neurodegeneration. 2011;6(1):31.

12. Gottesman II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. American Journal of Psychiatry. 2003;160(4):636-45.

13. Chan RC, Gottesman II. Neurological soft signs as candidate endophenotypes for schizophrenia: a shooting star or a Northern star? Neuroscience & Biobehavioral Reviews. 2008;32(5):957-71.

14. Johansen JS. Studies on serum YKL-40 as a biomarker in diseases with inflammation, tissue remodelling, fibroses and cancer. Dan Med Bull. 2006;53(2):172-209.

15. Johansen JS, Schultz NA, Jensen BV. Plasma YKL-40: a potential new cancer biomarker? Future Oncology. 2009;5(7):1065-82.

16. Huang K, Wu L. YKL-40: a potential biomarker for osteoarthritis. Journal of International Medical Research. 2009;37(1):18-24.

17. Kyrgios I, Galli-Tsinopoulou A, Stylianou C. Ghrelin–leptin network influences serum chitinase 3-like protein 1 (YKL-40) levels in obese prepubertal children. Regulatory peptides. 2013;183:69-73.

18. de Gauzy JS, Gennero I, Delrous O, Salles J-P, Lepage B, Accadbled F. Fasting total ghrelin levels are increased in patients with adolescent idiopathic scoliosis. Scoliosis. 2015;10(1):33.

19. Julien C, Gorman KF, Akoume M-Y, Moreau A. Towards a comprehensive diagnostic assay for scoliosis. Personalized Medicine. 2013;10(1):97-103.

20. Ober C, Tan Z, Sun Y, Possick JD, Pan L, Nicolae R, et al. Effect of variation in CHI3L1 on serum YKL-40 level, risk of asthma, and lung function. New England Journal of Medicine. 2008;358(16):1682-91.

21. Nielsen KR, Steffensen R, Boegsted M, Baech J, Lundbye-Christensen S, Hetland ML, et al. Promoter polymorphisms in the chitinase 3-like 1 gene influence the serum concentration of YKL-

40 in Danish patients with rheumatoid arthritis and in healthy subjects. Arthritis research & therapy. 2011;13(3):R109.

22. Kjaergaard A, Johansen J, Bojesen S, Nordestgaard B. Role of inflammatory marker YKL-40 in the diagnosis, prognosis and cause of cardiovascular and liver diseases. Critical reviews in clinical laboratory sciences. 2016;53(6):396-408.

23. Zheng JL, Lu L, Hu J, Zhang RY, Zhang Q, Chen QJ, et al. Genetic polymorphisms in chitinase 3-like 1 (CHI3L1) are associated with circulating YKL-40 levels, but not with angiographic coronary artery disease in a Chinese population. Cytokine. 2011;54(1):51-5.

24. Sun G, Lam T, Bobby N, Yim P, Lee K, Moreau A, et al. High osteopontin plasma level associated with abnormal cortical bone mineral density in girls with adolescent idiopathic scoliosis. Stud Health Technol Inform. 2012;176:457.

25. Yadav MC, Huesa C, Narisawa S, Hoylaerts MF, Moreau A, Farquharson C, et al. Ablation of osteopontin improves the skeletal phenotype of Phospho1−/− mice. Journal of Bone and Mineral Research. 2014;29(11):2369-81.

26. Xie N, Li M, Wu T, Liu J, Wang B, Tang F. Does elevated osteopontin level play an important role in the development of scoliosis in bipedal mice? The Spine Journal. 2015;15(7):1660-4.

27. Hoo C, Gatam L. Role of Remodelling in Adolescent Idiopathic Scoliosis: an Evaluation of Osteopontin Level. The Journal of Indonesian Orthopaedic. 2012;40(2).

Supplementary tables

Table S1. **Comparison between each two separate biological endophenotypes groups for the associations of each of the 12 studied SNPs (fisher exact test two sided)**

SNP

	Group1 VS Control	Group2 VS Control	Group3 VS Control	Group1 VS Group2	Group1 VS Group3	Group2 VS Group3
rs55700740	P=0.899	P=0.099	P=0.437	P=0.089	P=0.415	P=0.509
rs7542294	P=0.027	P=0.195	P=0.248	P=0.042	P=0.232	P=0.606
rs946259	P=0.918	P=0.082	P=0.338	P=0.108	P=0.485	P=0.411
rs880633	P=0.908	P=0.084	P=0.349	P=0.107	P=0.486	P=0.411
rs1538372	P=0.186	P=0.133	P=0.832	**P=0.001**	P=0.072	P=0.15
rs4950881	P=0.368	P=0.706	P=0.485	P=0.099	P=0.09	P=0.472
rs10399805	P=0.058	P=0.153	P=0.434	P=0.028	P=0.112	P=0.626
rs6691378	P=0.027	P=0.094	P=0.432	p=0.017	P=0.142	P=0.319
rs946261	P=0.368	P=0.014	P=0.444	P=0.388	P=0.95	P=0.153
rs946262	P=0.773	P=0.097	P=0.672	P=0.036	P=0.902	P=0.025
rs116415868	P=0.644	P=0.27	P=0.08	P=0.267	P=0.095	P=0.315
rs10920576	P=0.798	P=0.347	P=0.729	P=0.306	P=0.895	P=0.451

Table S2. Comparison between each two separate groups stratified based on severity for the associations of each of the 12 studied SNPs (fisher exact test two sided)

SNP	Severe VS Control	Non-severe VS Control	Severe VS non-severe
rs55700740	p=0.09	p=0.472	p=0.231
rs7542294	p=0.256	p=0.139	p=0.084
rs946259	p=0.094	p=0.427	p=0.304
rs880633	p=0.095	p=0.425	p=0.308
rs1538372	p=0.091	p=0.853	p=0.19
rs4950881	p-0.037	p=0.859	p-0.057
rs10399805	p=0.888	p=0.371	p=0.767
rs6691378	p=0.817	p=0.308	p=0.768
rs946261	p=0.348	p=0.107	p=0.574
rs946262	p=0.302	p=0.774	p=0.508
rs116415868	p=0.678	p=0.366	p=0.394
rs10920576	p=0.12	p=0.896	p=0.241

CHAPTER III

DISCUSSION

The main goal of this thesis was to gain more knowledge and understanding of the genetic basis of IS. IS represents a serious and chronic health condition affecting around seven million patients in the US and over 350000 in Canada. IS is a common complex disease, where a genetic component is widely recognized. Since the 1930s, familial incidence of scoliosis has been reported [162]. The hereditary nature of the disease is inferred from both family and twin studies, which reported high risk to relatives of affected individuals together with high rate of concordance among twins [165, 225, 226]. During the last decade, several studies have been undertaken to uncover the genetic architecture of IS. The mode of inheritance appears to be heterogeneous. Indeed, reports of IS inheritance include autosomal dominant, autosomal recessive, X-linked and multifactorial. The most prevalent types of genetic studies that have been performed so far are linkage analysis family studies or hypothesis driven candidate gene studies. Linkage studies were not particularly successful for the detection of genetic loci in IS. That is the reason why after 2006 most studies preferentially used case-control association methods [172]. In spite of the large number of studies involving candidate gene association studies, however poor replication results have been reported [172]. The poor results of hypothesis-driven candidate gene studies might be due to the limited state of knowledge regarding the underlying mechanisms associated with IS onset and/or spinal deformity progression, which are still not completely understood.

Since the emerging of GWAS in 2005, they have become one of the most powerful approaches used for studying the genetic basis of most complex diseases [227, 228]. These studies have led to the discovery of thousands of common variants associated with hundreds of common complex diseases (http://www.ebi.ac.uk/gwas/docs). Nevertheless, most of these discovered variances are of small to moderate effects and only explain a small proportion of

the heritability of these diseases [229]. The first GWAS in scoliosis was done in 2011. Since then multiple studies have been published with many suggestive associated loci, two of which (*LBX1* and *GPR126*) have been replicated by multiple groups [42, 193-197, 199, 200]. The most robust locus lies in the vicinity of the *LBX1* gene [42]. However, these two loci only explain approximately1% of the estimated 38% heritability of this disorder [17, 199]. One of the most important topics now in the field of genetics of common complex diseases is the missing heritability [229-231]. Common Disease Common Variant (CDCV) hypothesis is the main hypothesis on which GWAS were based on [232]. Because GWAS so far did not explain most of the heritability of most common diseases, the Common Disease Rare Variant (CDRV) hypothesis [233] was proposed. CDRV suggests that rare high penetrance variants might explain a part of this missing heritability. Hence, studying rare variants might uncover genetic and biological mechanisms underlying common complex diseases. Currently, due to the enormous advances in high-throughput sequencing technologies (NGS) and associated bioinformatics algorithms, it has become possible to identify all genetic variants in individual genomes, including those with both common and rare frequencies. In order to detect rare variants associated with IS, the most comprehensive approach is to sequence the whole genome in a large number of patients. Given that whole genome sequencing is still very expensive and because of the difficulty of interpretation of variants which lie away from genes, whole exome sequencing (WES) represents a more cost-effective alternative.

Gene expression profiling is another approach that yields massive and high-throughput information. One can refine this information and wisely choose which to investigate more deeply to identify those disease causing or modifiers, which play a role in the etiopathogenesis of the disease.

In this thesis, we adopted different approaches towards studying the genetic basis of IS. Our first study was based on the CDRV hypothesis wherein we used whole exome sequencing with a combined two-stage case/control and multiplex family model to study genes enriched with rare variants in IS patients. In the second study we performed a standard case/control GWAS of common variants. The full results of the GWAS analysis are not the scope of this thesis, but association of the known *LBX1* locus was analyzed. This locus has not been tested before in a French Canadian population. In the third study, we assessed the contribution of YKL-40, a secreted glycoprotein encoded by *CHI3L1* gene, as a potential protecting disease-modifier in the context of IS.

1. Rare variants in the *FAT3* gene are associated with familial and isolated Idiopathic Scoliosis

Due to the issue of missing heritability in complex traits in general, some investigators have shifted their research towards studying rare variants involved in disease susceptibility. Deleterious variants are more likely to be rare [234], because high penetrance variants which are disease causing are more rapidly removed from populations by selection, and thus will end up with very low frequency. Exome sequencing offers a comprehensive study of exonic variation and hence, was our method of choice to investigate rare variants in IS. Due to the great success of WES in mapping genes involved in Mendelian diseases, it has become very promising for utility in complex disorders as well [235]. WES has limitations being difficult to accurately detect indels and not suitable for the detection of chromosomal rearrangements and structural variants. Moreover, the methods used for capturing the exons are not perfect where the capture is uneven and incomplete [236]. This is beside the fact that WES can miss

important variants in the non-coding regulatory regions like promoters or enhancers. However, WES is a cost-effective strategy; in spite of occupying < 1.5% of the genome, coding exons account for 85% of the known variation discovered in Mendelian diseases [235]. The advent of NGS has revolutionized the research in genetics of diseases. It has made possible to perform sequencing projects much easily in the laboratories. However, the costs of NGS are still relatively expensive to sequence large cohort sizes. Hence, to perform our study we implemented a two-stage design. First, a discovery cohort was exome sequenced to identify candidate genes, followed by a second phase sequencing of these candidate genes in an independent replication cohort. In this study, we also implemented a family study beside the two stages case/control study. Family studies in complex diseases have demonstrated subgroups of patients with Mendelian inheritance and there are several successful examples in the literature for identifying genes in several complex diseases using family studies examples: *BRCA1* and *BRCA2* in breast cancer, *APC* in colon cancer and LDL receptor gene in heart disease [180, 237]. We did not anticipate that both studies will lead to identification of the same gene. Hence, the two studies strongly support each other. We successfully identified a strong novel association between the gene *FAT3* and IS. Our study brought new evidence of the importance of rare variants in the susceptibility to complex genetic disorders like IS. The high odds ratio (OR=13.6) obtained when statistical analysis was performed on protein altering rare variants (MAF ≤1%) in the replication cohort reflect the high effect of these rare variants, much higher than those obtained from previous GWAS studies, supporting the CDRV hypothesis. Due to the very low frequency of these rare variants to yield sufficient statistical power individually, we employed a collapsing gene burden test in which we compared genes enriched with rare variants in patients versus controls. One

limitation of this statistical approach is the assumption that all the variants are acting in the same direction with the same magnitude, which might not accurately be the case. However, being in the coding region and protein altering variants, chance is very high that they might be deleterious. Hence, we used three different algorithms to predict the functional consequences of these non-synonymous SNVs: SIFT, Polyphen-2 and Mutation Taster-2. Two out of the 26 reported non-synonymous variants were predicted to be pathogenic by all three software tools and thirteen variants were predicted to be pathogenic by two of them, while one variant was a frame shift mutation expected also to be deleterious. Experimental functional studies in the future will reveal the functional consequences of these variants on FAT3 protein function. We would like to note here that recently we noticed that the pipeline used by our Bioinformatics team missed some rare variants that were present in controls only, consequently we did a revised analysis in which rare variants in the controls increased but it was still over presented in cases versus controls.

Our study is among few studies, which performed WES approach to study rare variants in IS. Most other studies looked for individual rare variants rather than collapsing them by gene [205, 221]. Other studies were family based which subsequently investigated their genes in independent cohorts [218, 219]. There are only two studies, which were similar to ours in terms of being case/control and employing a gene burden analysis, although neither identified *FAT3*. One of them performed WES for a discovery cohort, and then followed up the most strongly associated gene and a related gene in their replication cohort [217], while in our case we chose a larger number (twenty-four) of the highest associated genes to follow up in the replication cohort. In the second study [220], the definition of rare variant was more stringent than ours, as they were only looking for novel variants absent in the public databases.

Consequently, it is unclear to us whether *FAT3* would have been identified by these two studies if they had followed the same strategic analysis like ours.

FAT3 is a very interesting gene candidate contributing to either the occurrence or progression (or both) of IS. It is a member of the *FAT* gene family homologous to the *Drosophila* gene *Fat* [238] regulating planar cell polarity (PCP) in the Drosophila wing [239]. Members of the FAT cadherin subfamily have conserved structures from flies to vertebrates [240]. We suggest that the involvement of *FAT3* gene in the susceptibility to IS is through the PCP pathway. Hayes et al demonstrated that a targeted mutation in the zebrafish *D. rerioptk7* gene lead to the occurrence of scoliosis and also led to the disruption of both planar cell polarity (PCP) and Wnt/ß-catenin signaling [241]. Guo et al demonstrated that overexpression of human *LBX1* (the most robustly confirmed gene to be associated with IS by GWAS) and its homologs in zebrafish led to the formation of curvature in the zebrafish body axis through affecting the Wnt/PCP signalling. Recently, another study reported a mutation of the *VANGL1* gene, one of the PCP genes, in IS patients [242]. Moreover, Le Pabic et al proposed a model whereby FAT3 through PCP signaling coordinates the polarity and differentiation of chondrocytes affecting skeletal morphology [243]. All these studies present evidence which supports our suggested hypothesis of the involvement of *FAT3* in IS susceptibility through the PCP pathway.

2. A Replication Study for Association of *LBX1* locus with Adolescent Idiopathic Scoliosis in French-Canadian Population

GWAS is a powerful approach for the identification of genetic variants contributing to complex multifactorial traits. Since 2011, GWAS has been increasingly performed to study

the genetic architecture of IS which is considered among the most common complex genetic diseases affecting the musculoskeletal system. It is generally agreed upon that GWAS results should be confirmed in different populations with different ethnic backgrounds. The most important locus which has been identified by GWAS in IS is the *LBX1* locus. The association of this region with IS was discovered by a Japanese group and confirmed in multiple Asian populations. This association has also been confirmed by other cohorts of European descent. However, results of a GWAS performed in the United States on 491 IS families of which 81% were white non-Hispanic, did not contain rs11190870 (the most validated SNP in this locus) among the top 100 associated SNPs [244]. French-Canadians although of European origin, are genetically different from other cohorts of European descent. Hence, confirmation of the *LBX1* locus association in a French-Canadian population was important because this might help eventually in management of French Canadian IS patients if a different genetic architecture implied different pathogenesis mechanisms.

In this study, we confirmed the association of the *LBX1* locus with our French-Canadian IS population. Moreover, the severity showed an impact on the *LBX1* association and the p-values were more significant. Furthermore, we reported both at risk and at resistance to risk haplotypes that have the potential for use in prediction of IS risk in clinical and public health settings. Functional studies of *LBX1* in the zebrafish model have shown that overexpression of *LBX1* or any of the zebrafish homolog genes causes scoliosis like deformation in the body axis at different stages of development. Moreover, they showed that this effect is through dysregulation of the Wnt/PCP pathway. These results are in agreement with our discovered gene in the first part of the study: *FAT3* is an important player in the PCP pathway. Our results strongly suggest the involvement of the PCP pathway in the

etiopathogenesis of IS which warrants further investigation in future studies. Our hypothesis is further supported by other evidences from the literature which we mentioned before (summarized in **figure 7**.)
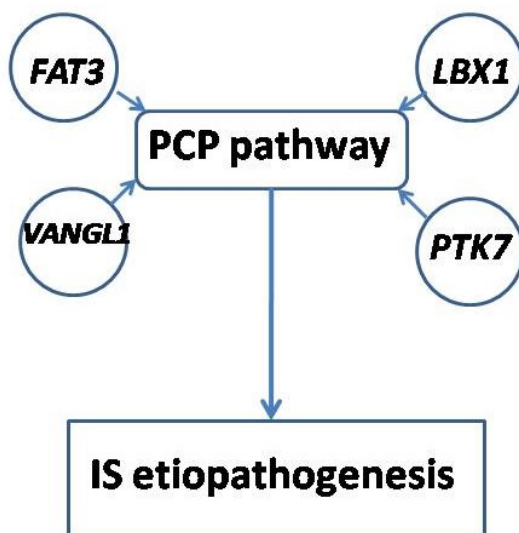
Figure 7: PCP pathway might be involved in the etiopathogenesis of IS. *LBX1* and *PTK7* have been shown by functional studies to be related to the scoliotic phenotype through affecting the PCP pathway, *FAT3* and *VANGL1* are players of PCP pathway and have been shown to be associated with IS patients where functional studies are needed.

3. Association of Circulating YKL-40 Levels and *CHI3L1* Gene Variants with Idiopathic Scoliosis and the Risk of Spinal Deformity Progression

This study is a candidate gene study, in which we studied the gene *CHI3L1*, encoding the secretory glycoprotein YKL-40. This gene was identified by a previous study in our lab (Gorman et al, manuscript submitted), to be overexpressed in IS patients classified in the biological endophenotype FG1 compared to the other biological endophenotypes (FG2 and FG3), which are more prone to develop severe scoliosis. These endophenotypes were

previously identified by our team, where IS patients were classified based on a cellular assay in which maximum Gi-protein signaling responses are measured [151]. This new IS patients' stratification, led to a better understanding of the variability associated with IS. Hence, we decided to study the biochemical correlates of circulating YKL-40 and to analyze the *CHI3L1* gene variants in a large cohort of IS patients. Our results showed that plasma YKL-40 levels were significantly associated with males patients classified in endophenotype FG1. Of note, FG1 patients are the least subgroup likely to develop a severe curvature [175]. It is also known that males are much less prone to develop severe scoliosis than females. In addition, when IS patients were classified according to their severity, the non-severe group showed statistically significant higher levels of YKL-40 than controls. The association of high plasma YKL-40 levels with the non-severe forms of IS, prompted us to hypothesize that YKL-40 might play a protective role in this disorder. Some of the tested SNPs in *CHI3L1* gene showed association with plasma YKL-40 levels in IS cases specially the non-severe cases. However, none of the tested SNPs could explain the higher YKL-40 levels in patients classified in endophenotype FG1 and further studies are needed to elucidate this.

We also performed a functional *invitro* analysis to test for the effect of YKL-40 on Gi-signaling dysfunction, which was shown by our team to be induced by osteopontin in IS patients [157]. The role of OPN in scoliosis development in humans and different animal models has more recently been reported by different groups [245-248]. Indeed, our results showed that YKL-40 counteracts the effect of osteopontin and alleviate the Gi-signaling dysfunction which could consequently reduce the severity of scoliosis.

Our results suggest that YKL-40 is a potential modifying factor that might have a protective role in IS.

CHAPTER IV

FUTURE PERSPECTIVES AND CONCLUSIONS

Future perspectives

Gene discovery is an important first step towards understanding the genetic basis and biological mechanisms underlying IS. Before the advent of GWAS, there was only limited success in the genetic dissection of complex diseases. Since the first published GWAS in 2005, hundreds of GWAS were performed and lead to the detection of thousands of moderately effective SNPs associated with many common complex diseases (http://www.ebi.ac.uk/gwas/docs). In the field of scoliosis, few strong candidates have been suggested by this approach. GWAS still have the potential to uncover more genes. Meta-analysis studies and cooperation between different groups of researchers should increase the power of the studies and detect variants with smaller effect sizes.

However, as a learned lesson from other complex traits where GWAS have been extensively done, only a small proportion of the heritability of these diseases were explained by common variants identified by GWAS. This raised arguments in favour of the importance of other factors to explain this missing heritability. This includes rare variants, structural variants, gene-gene interactions and gene-environment interactions. These fields of research are still in their infancy and needed to be thoroughly addressed. The currently available WES and targeted sequencing technologies have the limitation of screening the coding regions and adjacent splice sites preferentially. In the near future, with improved sequencing technologies and Bioinformatics tools, it is anticipated that WGS will become more affordable and practically applied. WGS of large populations is the most extensive approach to uncover variants contributing to IS.

NGS technologies and microarray based techniques have enabled the identification of structural variants including small and large indels, CNVs, and other chromosomal rearrangements >50bp [249]. Structural variants are very important as they account for >70% of the nucleotides alterations inspite of comprising only 20% of all genetic variation [250]. Studying structural variation in IS is needed and promising and expected to give valuable results.

Gene-gene interactions are likely to be important in multifactorial complex traits like IS. Timberlake et al documented genetic interaction between common and rare variants of two functionally related genes involved in the same phenotype (unrelated to IS) [251]. Morocz et al studied the association of selected SNPs in candidate genes and IS. They reported that while individual SNPs did not show associations with IS, there were associations with IS when some of these SNPs were combined together, indicating epistasis between those genes [252]. Gene-gene interactions are important to be investigated and would give deeper insights into the pathological mechanisms of the disease.

IS is a multifactorial disorder, hence both genetic and environmental factors are likely to contribute to its risk. In a MZ twin pair study, twins were diagnosed with IS at different ages, with different apical levels and different degree of severity which stresses the significance of presumptive non-genetic factors in the etiopathogenesis of the disorder [253]. The environmental component of IS has been estimated to be 62% [17]. The environmental risk factors are both internal and external, from intrauterine life onwards. Possible environmental risk factors that might contribute to IS include[69]; nutrition [254], physical activity [255], geographic latitude [256], maternal age and socio-economic level [257, 258], velocity of trunk growth and hormones [259], and asymmetric pressure of the intervertebral

disc [260, 261]. Indeed, individuals sharing a similar genetic predisposition and exposed to the same environment are more likely to develop the disease if this environment is enriched of risk factors, while in absence of those environmental factors, they may be less affected or could even remain asymptomatic. Conversely, individuals subjected to the same risk environmental factors but lacking the genetic predisposition are more likely to remain unaffected. Consequently, the penetrance of a genetic variant is likely to be affected by its cross talk with environmental factors. Hence, gene-environment interactions are needed to be assessed in IS patients. These studies will be challenged by the need of very large sample sizes. It has been estimated that for a study to detect interactions, the study will need four times the sample size needed for studying each factor individually [262]. Genetic-environmental interactions were extensively studied in human growth, behavioural conditions, early-onset conditions and gastrointestinal disorders [69].

An important intermediate which has the potential to fill the gap between genetics and environmental factors is epigenetics [263]. Petronis suggested a modification to the equation: P (phenotype)= G (genes) + E (environment) to P= G + E + EpiG (epigenetics) [264]. Epigenetics are heritable information during cell division without affecting the DNA sequence [265]. It includes DNA methylation, histone modifications and nucleosome positioning. Epigenetic mechanisms control the expression and repression of targeted sequences in the genome. Environmental and random factors can lead to malfunction of this complex system, known as epimutations, giving rise to diseases [69]. Sex is an important factor which affects epigenetic processes [266]. Research in epigenetics is complicated by the fact that it differs according to the cell type, sex, age, environmental exposures, developmental stage and other factors [267]. During the last decade extensive epigenetic

studies have been performed in other diseases including those of developmental origin and chronic non-communicable diseases [69]. To our knowledge, epigenetics has not been studied in IS and represents an unexplored frontier. Epigenetics research is promising for both disease prevention and treatment because epigenetic processes are reversible giving the possibility of medical intervention [69]. Disrupted epigenetic pathways could be targeted by medication, which could both modify the effects of malfunctioning genes and the effects of environmental factors on phenotypic plasticity [69]. Hence, epigenetic research in IS is definitely warranted and the understanding of epigenetic mechanisms have great promises in identifying diagnostic tests and targeted therapies.

Another important field of research that should be considered in the study of IS, are the contribution of small non-coding micro RNAs, which may modify gene expression through multiple mechanisms [268]. Micro RNAs are single stranded and composed of ~22 nucleotides [269]. Micro RNAs are believed to play important roles in the regulation of many biological processes including development, growth, differentiation and cell death by targeting genes involved in these processes and are involved in some diseases like cancer [268]. Micro RNAs of plant origin can enter the body and bind to some targeted mRNAs and amplify the expression of some genes and thus can be used in management of some diseases [69]. There are available expression arrays to help investigators to study the level of expression of miRNAs together with mRNAs in different tissues. It is highly anticipated to witness in the future studies reporting mutations that affect miRNA-mediated regulation of genes, which increase disease risk.

Gene expression profiling is another important approach for the detection of genes whose expressions are dysregulated in disease. Few IS studies using this approach have been

performed [175, 223, 224]. There are multiple challenges for investigators who choose this type of approach. The choice of which tissue to study is difficult, since IS is a multi-systemic disease involving multiple tissues. In addition, the optimal time of sample collection is problematic as IS is a developmental disorder. Whether or not the patients are using braces may be a confounding factor, as it has been reported that braces induce mechanical stress and consequently affect gene expression [180, 270]. Another challenge for this type of study in IS is the difficulty of obtaining control samples from the same anatomical site as from patients because usually these samples are specimens from bones, muscles, or intervertebral discs which are taken from patients during surgeries performed on their spine. It is important in gene expression studies to compare between cases and control samples from the same anatomical site to have accurate comparisons [180].

By identifying genes associated with IS we could gain a better understanding of the underlying pathology and we would be able to develop better treatment approaches. Gene discovery is only the first step towards understanding the disease biology. Follow-up of these genes and studying their actual role in the disease and thus clarifying the biochemical mechanisms would be the real discovery. Hence, genetic studies for IS in the future should be more directed to functional studies. Functional studies take the gene from just a name to signalling pathways and biological mechanisms that well clarifies the disease. More specifically future studies suggested from this thesis would be to do functional experiments to study the role of *FAT3*, *LBX1*, YKL-40 and the PCP pathway in the pathophysiology of IS.

Conclusions

Idiopathic scoliosis is a common complex multifactorial disorder, which is still poorly understood and poorly managed from a patient perspective. To fully understand this disease is like solving a big complicated puzzle. One should first have in his hands the different pieces of this puzzle to be able to put them together and have the full clear picture. Our thesis represents some pieces of this big puzzle, which together with other studies will help clarify the full picture of the disease. In this thesis, we were able to identify genetic and molecular factors implicated in the pathophysiology of the disease. *FAT3* is a new genetic factor that presents a strong candidate associated with IS patients, and *LBX1* is a newly confirmed genetic factor for its association with population of French-Canadian ethnicity. Planar cell polarity is a plausible pathway to play a role in the biology of the disease, as suggested by evidence from the literature and by the involvement of both *FAT3* and *LBX1* in this pathway. Finally, YKL-40 is a new biochemical factor that is associated with the non-severe scoliosis forms suggesting a role as a protecting disease-modifying factor. Collectively, our findings add significant new information to the field.

Reference List

1.   Nussbaum RL, McInnes RR, Willard HF. Thompson & Thompson genetics in medicine: Elsevier Health Sciences; 2015.

2.   Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature. 2001;409(6822):928-33.

3.   Gibbs RA, Belmont JW, Hardenbol P, Willis TD, Yu F, Yang H, et al. The international HapMap project. Nature. 2003;426(6968):789-96.

4.   Consortium IH. A haplotype map of the human genome. Nature. 2005;437(7063):1299-320.

5.   Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007;449(7164):851-61.

6.   Consortium GP. A map of human genome variation from population-scale sequencing. Nature. 2010;467(7319):1061-73.

7.   Keane MG, Pyeritz RE. Medical management of Marfan syndrome. Circulation. 2008;117(21):2802-13.

8.   Walker FO. Huntington's disease. The Lancet. 2007;369(9557):218-28.

9.   Salvatore D, Buzzetti R, Baldo E, Furnari ML, Lucidi V, Manunza D, et al. An overview of international literature from cystic fibrosis registries. Part 4: update 2011. Journal of Cystic Fibrosis. 2012;11(6):480-93.

10.   Lubs HA, Stevenson RE, Schwartz CE. Fragile X and X-linked intellectual disability: four decades of discovery. The American Journal of Human Genetics. 2012;90(4):579-90.

11.   Craig J. Complex diseases: Research and applications. Nature Education. 2008;1(1):184.

12.   Merlo CA, Boyle MP. Modifier genes in cystic fibrosis lung disease. Journal of Laboratory and Clinical Medicine. 2003;141(4):237-41.

13.   Nagel RL. Pleiotropic and epistatic effects in sickle cell anemia. Current opinion in hematology. 2001;8(2):105-10.

14. Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA. Linkage of early-onset familial breast cancer to chromosome 17q21. Science. 1990;250(4988):1684.

15. Wooster R, Bignell G, Lancaster J, Swift S. Identification of the breast cancer susceptibility gene BRCA2. Nature. 1995;378(6559):789.

16. Falconer DS. The inheritance of liability to certain diseases, estimated from the incidence among relatives. Annals of human genetics. 1965;29(1):51-76.

17. Grauers A, Rahman I, Gerdhem P. Heritability of scoliosis. European Spine Journal. 2012;21(6):1069-74.

18. MacGregor AJ, Snieder H, Schork NJ, Spector TD. Twins: novel uses to study complex traits and genetic diseases. Trends in Genetics. 2000;16(3):131-4.

19. So HC, Gui AH, Cherny SS, Sham PC. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. Genetic epidemiology. 2011;35(5):310-7.

20. Ikegawa S. Genomic study of adolescent idiopathic scoliosis in Japan. Scoliosis and Spinal Disorders. 2016;11(1):5.

21. Ku CS, Magnusson PK, Chia KS, Pawitan Y. Research on rare variants for complex diseases. eLS. 2010.

22. Elbein S, Corsetti L, Ullrich A, Permutt M. Multiple restriction fragment length polymorphisms at the insulin receptor locus: a highly informative marker for linkage analysis. Proceedings of the National Academy of Sciences. 1986;83(14):5223-7.

23. Huang K, Weng X, Liu J, Zhu Z, Xiao B, Yan M, et al. Amplified fragment length polymorphism analysis on D5S436 locus and its application to linkage analysis in gene diagnosis of asthma. Chinese medical journal. 1999;112(2):112-4.

24. Selmer KK, Brandal K, Olstad OK, Birkenes B, Undlien DE, Egeland T. Genome-wide linkage analysis with clustered SNP markers. Journal of biomolecular screening. 2009;14(1):92-6.

25. Teo Y-Y, Small KS, Kwiatkowski DP. Methodological challenges of genome-wide association analysis in Africa. Nature Reviews Genetics. 2010;11(2):149-60.

26. Marker PD. Cystic fibrosis locus defined by genetically linked polymorphic DNA marker. 1985.

27. Knowlton RG, Cohen-Haguenauer O, Van Cong N, Frézal J, Brown VA, Barker D, et al. A polymorphic DNA marker linked to cystic fibrosis is located on chromosome 7. 1985.

28. Barrett JH, Sheehan NA, Cox A, Worthington J, Cannings C, Teare M. Family based studies and genetic epidemiology: theory and practice. Human heredity. 2007;64(2):146-8.

29. Risch NJ. Searching for genetic determinants in the new millennium. Nature. 2000;405(6788):847-56.

30. Rioux JD, Xavier RJ, Taylor KD, Silverberg MS, Goyette P, Huett A, et al. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. Nature genetics. 2007;39(5):596-604.

31. Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nature genetics. 2009;41(6):703-7.

32. Collins FS, Morgan M, Patrinos A. The Human Genome Project: lessons from large-scale biology. Science. 2003;300(5617):286-90.

33. Thorisson GA, Smith AV, Krishnan L, Stein LD. The international HapMap project web site. Genome research. 2005;15(11):1592-3.

34. Klein RJ, Zeiss C, Chew EY, Tsai J-Y, Sackler RS, Haynes C, et al. Complement factor H polymorphism in age-related macular degeneration. Science. 2005;308(5720):385-9.

35. Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. science. 2006;314(5804):1461-3.

36. Burton PR, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A, et al. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447(7145):661-78.

37. Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V, et al. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nature genetics. 2007;39(7):857-64.

38. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature. 2007;445(7130):881-5.

39. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science. 2007;316(5826):889-94.

40. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nature genetics. 2008;40(5):638-45.

41. Franke A, McGovern DP, Barrett JC, Wang K, Radford-Smith GL, Ahmad T, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nature genetics. 2010;42(12):1118-25.

42. Takahashi Y, Kou I, Takahashi A, Johnson TA, Kono K, Kawakami N, et al. A genome-wide association study identifies common variants near LBX1 associated with adolescent idiopathic scoliosis. Nature genetics. 2011;43(12):1237-40.

43. Esposito F, Sorosina M, Ottoboni L, Lim ET, Replogle JM, Raj T, et al. A pharmacogenetic study implicates SLC9a9 in multiple sclerosis disease activity. Annals of neurology. 2015;78(1):115-27.

44. Goyal R, Lee S, Wang T, Trucco M, Haagenson M, Spellman S, et al. Novel HLA-DP region susceptibility loci associated with severe acute GvHD. Bone Marrow Transplantation. 2016.

45. Price AL, Zaitlen NA, Reich D, Patterson N. New approaches to population stratification in genome-wide association studies. Nature Reviews Genetics. 2010;11(7):459-63.

46. Campbell CD, Ogburn EL, Lunetta KL, Lyon HN, Freedman ML, Groop LC, et al. Demonstrating stratification in a European American population. Nature genetics. 2005;37(8):868-72.

47. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nature Reviews Genetics. 2010;11(6):415-25.

48. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proceedings of the national academy of sciences. 1977;74(12):5463-7.

49. Langaee T, Ronaghi M. Genetic variation analyses by Pyrosequencing. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis. 2005;573(1):96-102.

50. Cock PJ, Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic acids research. 2010;38(6):1767-71.

51. Mardis ER. The impact of next-generation sequencing technology on genetics. Trends in genetics. 2008;24(3):133-41.

52. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC genomics. 2012;13(1):341.

53. Albert TJ, Molla MN, Muzny DM, Nazareth L, Wheeler D, Song X, et al. Direct selection of human genomic loci by microarray hybridization. Nature methods. 2007;4(11):903-5.

54. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, et al. Exome sequencing identifies the cause of a mendelian disorder. Nature genetics. 2010;42(1):30-5.

55. Maxmen A. Exome sequencing deciphers rare diseases. Cell. 2011;144(5):635-7.

56. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. The American Journal of Human Genetics. 2008;83(3):311-21.

57. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. Nature Reviews Genetics. 2010;11(10):685-96.

58. Cheng JC, Castelein RM, Chu WC, Danielsson AJ, Dobbs MB, Grivas TB, et al. PRIMER. NATURE. 2015.

59. Yaman O, Dalbayrak S. Idiopathic scoliosis. Turkish neurosurgery. 2013;24(5):646-57.

60. Konieczny MR, Senyurt H, Krauspe R. Epidemiology of adolescent idiopathic scoliosis. Journal of children's orthopaedics. 2013;7(1):3-9.

61. Kamtsiuris P, Atzpodien K, Ellert U, Schlack R, Schlaud M. [Prevalence of somatic diseases in German children and adolescents. Results of the German Health Interview and Examination Survey for Children and Adolescents (KiGGS)]. Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz. 2006;50(5-6):686-700.

62. Suh S-W, Modi HN, Yang J-H, Hong J-Y. Idiopathic scoliosis in Korean schoolchildren: a prospective screening study of over 1 million children. European spine journal. 2011;20(7):1087-94.

63. Nery LS, Halpern R, Nery PC, Nehme KP, Tetelbom Stein A. Prevalence of scoliosis among school students in a town in southern Brazil. Sao Paulo Medical Journal. 2010;128(2):69-73.

64. Daruwalla J, Balasubramaniam P, Chay S, Rajan U, Lee H. Idiopathic scoliosis. Prevalence and ethnic distribution in Singapore schoolchildren. Bone & Joint Journal. 1985;67(2):182-4.

65. Wong H-K, Hui JH, Rajan U, Chia H-P. Idiopathic scoliosis in Singapore schoolchildren: a prevalence study 15 years into the screening program. Spine. 2005;30(10):1188-96.

66. Cilli K, Tezeren G, Taş T, Bulut O, Oztürk H, Oztemur Z, et al. [School screening for scoliosis in Sivas, Turkey]. Acta orthopaedica et traumatologica turcica. 2008;43(5):426-30.

67. Soucacos PN, Soucacos PK, Zacharis KC, Beris AE, Xenakis TA. School-screening for scoliosis. A prospective epidemiological study in northwestern and central Greece. J Bone Joint Surg Am. 1997;79(10):1498-503.

68. Fong DYT, Lee CF, Cheung KMC, Cheng JCY, Ng BKW, Lam TP, et al. A meta-analysis of the clinical effectiveness of school scoliosis screening. Spine. 2010;35(10):1061-71.

69. Burwell RG, Dangerfield PH, Moulton A, Grivas TB. Adolescent idiopathic scoliosis (AIS), environment, exposome and epigenetics: a molecular perspective of postnatal normal spinal growth and the etiopathogenesis of AIS with consideration of a network approach and possible implications for medical therapy. Scoliosis. 2011;6(1):1.

70. Ueno M, Takaso M, Nakazawa T, Imura T, Saito W, Shintani R, et al. A 5-year epidemiological study on the prevalence rate of idiopathic scoliosis in Tokyo: school screening of more than 250,000 children. Journal of Orthopaedic Science. 2011;16(1):1-6.

71. Souza FId, Ferreira RBD, Labres D, Elias R, Sousa APMd, Pereira RE. Epidemiology of adolescent idiopathic scoliosis in students of the public schools in Goiânia-Go. Acta ortopedica brasileira. 2013;21(4):223-5.

72. Luk KD, Lee CF, Cheung KM, Cheng JC, Ng BK, Lam TP, et al. Clinical effectiveness of school screening for adolescent idiopathic scoliosis: a large population-based retrospective cohort study. Spine. 2010;35(17):1607-14.

73. Lee C, Fong DY, Cheung KM, Cheng JC, Ng BK, Lam T, et al. A new risk classification rule for curve progression in adolescent idiopathic scoliosis. The Spine Journal. 2012;12(11):989-95.

74. Deurloo J, Verkerk P. To screen or not to screen for adolescent idiopathic scoliosis? A review of the literature. Public health. 2015;129(9):1267-72.

75. Adobor RD, Joranger P, Steen H, Navrud S, Brox JI. A health economic evaluation of screening and treatment in patients with adolescent idiopathic scoliosis. Scoliosis. 2014;9(1):1.

76. Labelle H, Richards SB, De Kleuver M, Grivas TB, Luk KD, Wong HK, et al. Screening for adolescent idiopathic scoliosis: an information statement by the scoliosis research society international task force. Scoliosis. 2013;8(1):1.

77. Rigo M. Patient evaluation in idiopathic scoliosis: Radiographic assessment, trunk deformity and back asymmetry. Physiotherapy theory and practice. 2011;27(1):7-25.

78. El-Hawary R, Chukwunyerenwa C. Update on evaluation and treatment of scoliosis. Pediatric clinics of North America. 2014;61(6):1223-41.

79. Lonstein JE, Carlson J. The prediction of curve progression in untreated idiopathic scoliosis during growth. J Bone Joint Surg Am. 1984;66(7):1061-71.

80. Grivas TB, Dangas S, Samelis P, Maziotou C, Kandris K. Lateral Spinal Profile in School-screening Referrals with and without Late Onset Idiopathic Scoliosis 10^ o-20^ o. Studies in health technology and informatics. 2002:25-31.

81. Bunnell WP. The natural history of idiopathic scoliosis before skeletal maturity. Spine. 1986;11(8):773-6.

82. Negrini S, Aulisa AG, Aulisa L, Circo AB, de Mauroy JC, Durmala J, et al. 2011 SOSORT guidelines: orthopaedic and rehabilitation treatment of idiopathic scoliosis during growth. Scoliosis. 2012;7(1):1.

83. Kotwicki T, Durmała J, Czaprowski D, Głowacki M, Kołban M, Snela S, et al. Conservative management of idiopathic scoliosis--guidelines based on SOSORT 2006 Consensus. Ortopedia, traumatologia, rehabilitacja. 2008;11(5):379-95.

84. Mordecai SC, Dabke HV. Efficacy of exercise therapy for the treatment of adolescent idiopathic scoliosis: a review of the literature. European Spine Journal. 2012;21(3):382-9.

85. Romano M, Minozzi S, Zaina F, Saltikov JB, Chockalingam N, Kotwicki T, et al. Exercises for adolescent idiopathic scoliosis: a Cochrane systematic review. Spine. 2013;38(14):E883-E93.

86. Lenssinck M-LB, Frijlink AC, Berger MY, Bierma-Zeinstra SM, Verkerk K, Verhagen AP. Effect of bracing and other conservative interventions in the treatment of idiopathic scoliosis in adolescents: a systematic review of clinical trials. Physical therapy. 2005;85(12):1329-39.

87. Negrini S, Fusco C, Minozzi S, Atanasio S, Zaina F, Romano M. Exercises reduce the progression rate of adolescent idiopathic scoliosis: results of a comprehensive systematic review of the literature. Disability and rehabilitation. 2008;30(10):772-85.

88. Weiss H-R, Weiss G, Petermann F. Incidence of curvature progression in idiopathic scoliosis patients treated with scoliosis in-patient rehabilitation (SIR): an age-and sex-matched controlled study. Pediatric rehabilitation. 2003;6(1):23-30.

89. Weiss H-R, Weiss G. Curvature progression in patients treated with scoliosis in-patient rehabilitation--a sex and age matched controlled study. Studies in health technology and informatics. 2001;91:352-6.

90. Otman S, Kose N, Yakut Y. The efficacy of Schroth s 3-dimensional exercise therapy in the treatment of adolescent idiopathic scoliosis in Turkey. Saudi medical journal. 2005;26(9):1429-35.

91. Dobosiewicz K, Durmala J, Czernicki K, Jendrzejek H. Pathomechanic basics of conservative treatment of progressive idiopathic scoliosis according to Dobosiewicz method based upon radiologic evaluation. Studies in health technology and informatics. 2001;91:336-41.

92. Negrini S, Negrini A, Romano M, Verzini N, Negrini A, Parzini S. A controlled prospective study on the efficacy of SEAS. 02 exercises in preparation to bracing for idiopathic scoliosis. Studies in health technology and informatics. 2006;123:519.

93. Weiss H. Influence of an in-patient exercise program on scoliotic curve. Italian journal of orthopaedics and traumatology. 1991;18(3):395-406.

94. Negrini S, Zaina F, Atanasio S. BRACE MAP, a proposal for a new classification of braces. Studies in health technology and informatics. 2007;140:299-302.

95. Kotwicki T, Chowanska J, Kinel E, Czaprowski D, Tomaszewski M, Janusz P. Optimal management of idiopathic scoliosis in adolescence. Adolesc Health Med Ther. 2013;4:59-73.

96. Harrington PR. Treatment of scoliosis. J Bone Joint Surg Am. 1962;44(4):591-634.

97. Luque ER. Segmental spinal instrumentation for correction of scoliosis. Clinical orthopaedics and related research. 1982;163:192-8.

98. Veldhuizen A, Wever D, Webb P. The aetiology of idiopathic scoliosis: biomechanical and neuromuscular factors. European Spine Journal. 2000;9(3):178-84.

99. Herman R, MacEwen G, Zorab P. Idiopathic scoliosis: a visuovestibular disorder of the central nervous system. Scoliosis Academic Press, New York. 1979:61-99.

100. Herman R, Maulucci R, Stuyck J. Development and plasticity of visual and vestibular generated eye movements. Experimental brain research. 1982;47(1):69-78.

101. Sahlstrand T, Petruson B, örtengren R. Vestibulospinal reflex activity in patients with adolescent idiopathic scoliosis: postural effects during caloric labyrinthine stimulation recorded by stabilometry. Acta Orthopaedica Scandinavica. 1979;50(3):275-81.

102. SAHLSTRAND T, LIDSTRÖM J. Equilibrium factors as predictors of the prognosis in adolescent idiopathic scoliosis. Clinical orthopaedics and related research. 1980;152:232-6.

103. Sahlstrand T. An analysis of lateral predominance in adolescent idiopathic scoliosis with special reference to convexity of the curve. Spine. 1980;5(6):512-8.

104. Tezuka A. Development of scoliosis in cases with congenital organic abnormalities of the brain-stem. A report of 7 cases. The Tokushima journal of experimental medicine. 1971;18:49.

105. Dubousset J, Queneau P, Thillard M. Experimental scoliosis induced by pineal and diencephalic lesions in young chickens: its relation with clinical findings. Orthop Trans. 1983;7(7):4.

106. HERMAN R, MIXON J, FISHER A, MAULUCCI R, STUYCK J. Idiopathic Scoliosis and the Central Nervous System: A Motor Control Problem: The Harrington Lecture, 1983 Scoliosis Research Society. Spine. 1985;10(1):1-14.

107. Petersén I, Sahlstrand T, Sellden U. Electroencephalographic investigation of patients with adolescent idiopathic scoliosis. Acta Orthopaedica Scandinavica. 1979;50(3):283-93.

108. Chu WC, Lam WW, Chan Y-l, Ng BK, Lam T-p, Lee K-m, et al. Relative shortening and functional tethering of spinal cord in adolescent idiopathic scoliosis?: study with multiplanar reformat magnetic resonance imaging and somatosensory evoked potential. Spine. 2006;31(1):E19-E25.

109. Porter RW. Idiopathic scoliosis: the relation between the vertebral canal and the vertebral bodies. Spine. 2000;25(11):1360-6.

110. Porter RW. Can a short spinal cord produce scoliosis? European Spine Journal. 2001;10(1):2-9.

111. Porter RW. The pathogenesis of idiopathic scoliosis: uncoupled neuro-osseous growth? European Spine Journal. 2001;10(6):473-81.

112. Roth M. Idiopathic scoliosis and Scheuermann's disease: essentially identical manifestations of neuro-vertebral growth disproportion. Radiologia diagnostica. 1981;22(3):380.

113. Dayer R, Haumont T, Belaieff W, Lascombes P. Idiopathic scoliosis: etiological concepts and hypotheses. Journal of children's orthopaedics. 2013;7(1):11-6.

114. Liu T, Chu WC, Young G, Li K, Yeung BH, Guo L, et al. MR analysis of regional brain volume in adolescent idiopathic scoliosis: neurological manifestation of a systemic disease. Journal of Magnetic Resonance Imaging. 2008;27(4):732-6.

115. BARRACK RL, WHITECLOUD III TS, BURKE SW, COOK SD, HARDING AF. Proprioception in idiopathic scoliosis. Spine. 1984;9(7):681-5.

116. Hoogmartens M, Basmajian J. Postural tone in the deep spinal muscles of idiopathic scoliosis patients and their siblings. An etiologic study based on vibration-induced electromyogram. Electromyography and clinical neurophysiology. 1975;16(1):93-114.

117. Low W, Chew E, Kung L, Hsu L, Leong J. Ultrastructures of nerve fibers and muscle spindles in adolescent idiopathic scoliosis. Clinical orthopaedics and related research. 1983;174:217-21.

118. Matthews P. Muscle spindles and their motor control. Physiological Reviews. 1964;44(2):219-88.

119. Whitecloud T, Cook S, Burke S, Leinhardt T, Barrack R, editors. Upper extremity proprioceptive deficit in idiopathic scoliosis. Nineteenth Annual Scoliosis Research Society Meeting, Orlando, Florida; 1984.

120. Yekutiel M, Robin G, Yarom R. Proprioceptive function in children with adolescent idiopathic scoliosis. Spine. 1981;6(6):560-6.

121. Fidler M, Jowett R, Troup J. Histochemical study of the function of multifidus in scoliosis. Scoliosis and Muscle. 1974:184-92.

122. Fidler M, Jowett R. Muscle imbalance in the aetiology of scoliosis. Bone & Joint Journal. 1976;58(2):200-1.

123. Ford D, Bagnall K, McFadden K, Greenhill B, Raso V. Paraspinal muscle imbalance in adolescent idiopathic scoliosis. Spine. 1984;9(4):373-6.

124. Spencer G, Eccles M. Spinal muscle in scoliosis: Part 2. The proportion and size of type 1 and type 2 skeletal muscle fibres measured using a computer-controlled microscope. Journal of the neurological sciences. 1976;30(1):143-54.

125. Yarom R, Robin GC. Studies on spinal and peripheral muscles from patients with scoliosis. Spine. 1979;4(1):12-21.

126. Butterworth Jr T, James C. Electromyographic studies in idiopathic scoliosis. Southern medical journal. 1969;62(8):1008-10.

127. Redford JB, Butterworth T, Clements Jr E. Use of electromyography as a prognostic aid in the management of idiopathic scoliosis. Archives of physical medicine and rehabilitation. 1969;50(8):433-8.

128. Riddle H, Roaf R. Muscle imbalance in the causation of scoliosis. The Lancet. 1955;265(6877):1245-7.

129. Zetterberg C, Björk R, Örtengren R, Andersson GB. Electromyography of the paravertebral muscles in idiopathic scoliosis: measurements of amplitude and spectral changes under load. Acta Orthopaedica Scandinavica. 1984;55(3):304-9.

130. Arkin A. The mechanism of the structural changes in scoliosis; preliminary report. New York state journal of medicine. 1949;49(5):495.

131. Deane G, Duthie RB. A new projectional look at articulated scoliotic spines. Acta Orthopaedica Scandinavica. 1973;44(4-5):351-65.

132. Millner P, Dickson R. Idiopathic scoliosis: biomechanics and biology. European Spine Journal. 1996;5(6):362-73.

133. Anderson M, Hwang S-C, Green WT. Growth of the normal trunk in boys and girls during the second decade of life. J Bone Joint Surg Am. 1965;47(8):1554-64.

134. Calvo I. Observations on the growth of the female adolescent spine and its relation to scoliosis. Clinical orthopaedics. 1957;10:40-7.

135. Duthie RB. The significance of growth in orthopaedic surgery. Clinical Orthopaedics. 1959;14:7-19.

136. Duval-Beaupere G, Lamireau T. Scoliosis at Less than 30 [degrees]: Properties of the Evolutivity (Risk of Progression). Spine. 1985;10(5):421-4.

137. Willner S. A study of growth in girls with adolescent idiopathic structural scoliosis. Clinical orthopaedics and related research. 1974;101:129-35.

138. Buric M, Momcilovic B. Growth pattern and skeletal age in school girls with idiopathic scoliosis. Clinical orthopaedics and related research. 1982;170:238-42.

139. Leong J, Low W, Mok C, Kung L, Yau A. Linear growth in southern Chinese female patients with adolescent idiopathic scoliosis. Spine. 1982;7(5):471-5.

140. Nordwall A, Willner S. A study of skeletal age and height in girls with idiopathic scoliosis. Clinical orthopaedics and related research. 1975;110:6-10.

141. Dickson RA, Lawton J, Archer I, Butt W. The pathogenesis of idiopathic scoliosis. Biplanar spinal asymmetry. Bone & Joint Journal. 1984;66(1):8-15.

142. Smith R, Dickson R. Experimental structural scoliosis. Bone & Joint Journal. 1987;69(4):576-81.

143. Drevelle X, Lafon Y, Ebermeyer E, Courtois I, Dubousset J, Skalli W. Analysis of idiopathic scoliosis progression by using numerical simulation. Spine. 2010;35(10):E407-E12.

144. Hadley-Miller N, Mims B, Milewicz DM. The potential role of the elastic fiber system in adolescent idiopathic scoliosis. J Bone Joint Surg Am. 1994;76(8):1193-206.

145. Sanders JO, Browne RH, McConnell SJ, Margraf SA, Cooney TE, Finegold DN. Maturity assessment and curve progression in girls with idiopathic scoliosis. J Bone Joint Surg Am. 2007;89(1):64-73.

146. Machida M, Dubousset J, Yamada T, Kimura J. Serum melatonin levels in adolescent idiopathic scoliosis prediction and prevention for curve progression–a prospective study. Journal of pineal research. 2009;46(3):344-8.

147. Kulis A, Zarzycki D, Jaśkiewicz J. Concentration of estradiol in girls with idiophatic scoliosis. Ortopedia, traumatologia, rehabilitacja. 2006;8(4):455-9.

148. de Gauzy JS, Gennero I, Delrous O, Salles J-P, Lepage B, Accadbled F. Fasting total ghrelin levels are increased in patients with adolescent idiopathic scoliosis. Scoliosis. 2015;10(1):33.

149. Qiu Y, Sun X, Qiu X, Li W, Zhu Z, Zhu F, et al. Decreased circulating leptin level and its association with body and bone mass in girls with adolescent idiopathic scoliosis. Spine. 2007;32(24):2703-10.

150. Moreau A, Franco A, Azeddine B, Rompré PH, Bagnall K, Poitras B, et al. High Circulating Levels Of Osteopontin Are Associated With Idiopathic Scoliosis Onset And Spinal Deformity Progression. The Journal of Bone and Joint Surgery-british Volume. 2010;92:430.

151. Akoume M-Y, Azeddine B, Turgeon I, Franco A, Labelle H, Poitras B, et al. Cell-based screening test for idiopathic scoliosis using cellular dielectric spectroscopy. Spine. 2010;35(13):E601-E8.

152. Wang B, Chen Z, Qiu Y, Liu W. Decreased circulating matrilin-1 levels in adolescent idiopathic scoliosis. Zhonghua wai ke za zhi [Chinese journal of surgery]. 2009;47(21):1638-41.

153. Kulis A, Goździalska A, Drąg J, Jaśkiewicz J, Knapik-Czajka M, Lipik E, et al. Participation of sex hormones in multifactorial pathogenesis of adolescent idiopathic scoliosis. International orthopaedics. 2015;39(6):1227-36.

154. Lowe T, Lawellin D, Smith D, Price C, Haher T, Merola A, et al. Platelet calmodulin levels in adolescent idiopathic scoliosis: do the levels correlate with curve progression and severity? Spine. 2002;27(7):768-75.

155. Leboeuf D, Letellier K, Alos N, Edery P, Moldovan F. Do estrogens impact adolescent idiopathic scoliosis? Trends in Endocrinology & Metabolism. 2009;20(4):147-52.

156. Xia Z, Storm DR. Calmodulin-regulated adenylyl cyclases and neuromodulation. Current opinion in neurobiology. 1997;7(3):391-6.

157. Akoume M-Y, Franco A, Moreau A. Cell-based assay protocol for the prognostic prediction of idiopathic scoliosis using cellular dielectric spectroscopy. JoVE (Journal of Visualized Experiments). 2013(80):e50768-e.

158. Zhao D, Qiu Gx, Wang Yp, Zhang Jg, Shen Jx, Wu Zh, et al. Association of calmodulin1 gene polymorphisms with susceptibility to adolescent idiopathic scoliosis. Orthopaedic surgery. 2009;1(1):58-65.

159. Qiu XS, Tang NL, Yeung HY, Lee K-M, Hung VW, Ng BK, et al. Melatonin receptor 1B (MTNR1B) gene polymorphism is associated with the occurrence of adolescent idiopathic scoliosis. Spine. 2007;32(16):1748-53.

160. Chen Z, Tang NL, Cao X, Qiao D, Yi L, Cheng JC, et al. Promoter polymorphism of matrilin-1 gene predisposes to adolescent idiopathic scoliosis in a Chinese population. European Journal of Human Genetics. 2009;17(4):525-32.

161. Burwell RG, Aujla RK, Grevitt MP, Dangerfield PH, Moulton A, Randell TL, et al. Pathogenesis of adolescent idiopathic scoliosis in girls-a double neuro-osseous theory involving disharmony between two nervous systems, somatic and autonomic expressed in the spine and trunk: possible dependency on sympathetic nervous system and hormones with implications for medical therapy. Scoliosis. 2009;4(1):24.

162. Garland HG. Hereditary scoliosis. British medical journal. 1934;1(3823):686.

163. Lowe TG, Edgar M, Margulies JY, Miller NH, Raso VJ, Reinker KA, et al. Etiology of idiopathic scoliosis: current trends in research. J Bone Joint Surg Am. 2000;82(8):1157-.

164. RISEBOROUGH EJ, WYNNE-DAVIES R. A genetic survey of idiopathic scoliosis in Boston, Massachusetts. J Bone Joint Surg Am. 1973;55(5):974-82.

165. Kesling KL, Reinker KA. Scoliosis in Twins: A Meta-analysis of the Literature and Report of Six Cases. Spine. 1997;22(17):2009-14.

166. Cowell HR, Hall JN, MacEven GD. Genetic Aspects of Idiopathic Scoliosis: A Nicholas Andry Award Essay, 1970. Clinical orthopaedics and related research. 1972;86:121-31.

167. Alden KJ, Marosy B, Nzegwu N, Justice CM, Wilson AF, Miller NH. Idiopathic scoliosis: identification of candidate regions on chromosome 19p13. Spine. 2006;31(16):1815-9.

168. Edery P, Margaritte-Jeannin P, Biot B, Labalme A, Bernard J-C, Chastang J, et al. New disease gene location and high genetic heterogeneity in idiopathic scoliosis. European Journal of Human Genetics. 2011;19(8):865-9.

169. Ocaka L, Zhao C, Reed JA, Ebenezer ND, Brice G, Morley T, et al. Assignment of two loci for autosomal dominant adolescent idiopathic scoliosis (AIS) to chromosomes 9q31. 2-q34. 2 and 17q25. 3-qtel. Journal of medical genetics. 2007.

170. Ogilvie JW, Braun J, Argyle V, Nelson L, Meade M, Ward K. The search for idiopathic scoliosis genes. Spine. 2006;31(6):679-81.

171. Burwell RG, Clark EM, Dangerfield PH, Moulton A. Adolescent idiopathic scoliosis (AIS): a multifactorial cascade concept for pathogenesis and embryonic origin. Scoliosis and Spinal Disorders. 2016;11(1):8.

172. Gorman KF, Julien C, Moreau A. The genetic epidemiology of idiopathic scoliosis. European Spine Journal. 2012;21(10):1905-19.

173. Moreau A, Forget S, Azeddine B, Angeloni D, Fraschini F, Labelle H, et al. Melatonin signaling dysfunction in adolescent idiopathic scoliosis. Spine. 2004;29(16):1772-81.

174. Chan RC, Gottesman II. Neurological soft signs as candidate endophenotypes for schizophrenia: a shooting star or a Northern star? Neuroscience & Biobehavioral Reviews. 2008;32(5):957-71.

175. Julien C, Gorman KF, Akoume M-Y, Moreau A. Towards a comprehensive diagnostic assay for scoliosis. Personalized Medicine. 2013;10(1):97-103.

176. Miller NH, Justice CM, Marosy B, Doheny KF, Pugh E, Zhang J, et al. Identification of candidate regions for familial idiopathic scoliosis. Spine. 2005;30(10):1181-7.

177. Clough M, Justice CM, Marosy B, Miller NH. Males with familial idiopathic scoliosis: a distinct phenotypic subgroup. Spine. 2010;35(2):162.

178. Gao X, Gordon D, Zhang D, Browne R, Helms C, Gillum J, et al. CHD7 gene polymorphisms are associated with susceptibility to idiopathic scoliosis. The American Journal of Human Genetics. 2007;80(5):957-65.

179. Tilley MK, Justice CM, Swindle K, Marosy B, Wilson AF, Miller NH. CHD7 gene polymorphisms and familial idiopathic scoliosis. Spine. 2013;38(22).

180. Gorman KF, Julien C, Oliazadeh N, Tang Q, Moreau A. Genetics of Idiopathic Scoliosis. eLS. 2014.

181. Justice CM, Miller NH, Marosy B, Zhang J, Wilson AF. Familial idiopathic scoliosis: evidence of an X-linked susceptibility locus. Spine. 2003;28(6):589-94.

182. Raggio CL, Giampietro PF, Dobrin S, Zhao C, Dorshorst D, Ghebranious N, et al. A novel locus for adolescent idiopathic scoliosis on chromosome 12p. Journal of Orthopaedic Research. 2009;27(10):1366-72.

183. Aulisa AG, Guzzanti V, Mastantuoni G, Giordano M, Poggiaroni A, Aulisa L. The familiarity of idiopathic scoliosis: statistical analysis and clinical considerations. European journal of orthopaedic surgery & traumatology: orthopedie traumatologie. 2013;23(7):781-4.

184. Salehi L, Mangino M, De Serio S, De Cicco D, Capon F, Semprini S, et al. Assignment of a locus for autosomal dominant idiopathic scoliosis (IS) to human chromosome 17p11. Human genetics. 2002;111(4-5):401-4.

185. Chan V, Fong GC, Luk KD, Yip B, Lee M-K, Wong M-S, et al. A genetic locus for adolescent idiopathic scoliosis linked to chromosome 19p13. 3. The American Journal of Human Genetics. 2002;71(2):401-6.

186. Marosy B, Justice CM, Vu C, Zorn A, Nzegwu N, Wilson AF, et al. Identification of susceptibility loci for scoliosis in FIS families with triple curves. American Journal of Medical Genetics Part A. 2010;152(4):846-55.

187. Montanaro L, Parisini P, Greggi T, Di Silvestre M, Campoccia D, Rizzi S, et al. Evidence of a linkage between matrilin-1 gene (MATN1) and idiopathic scoliosis. Scoliosis. 2006;1(1):21.

188. Carr AJ, Ogilvie DJ, Wordsworth BP, Priestly LM, Smith R, Sykes B. Segregation of structural collagen genes in adolescent idiopathic scoliosis. Clinical orthopaedics and related research. 1992;274:305-10.

189. Miller N, Sponseller P, Mims B, Child A, Milewicz D, Blanton S. Genetic analysis of structural elastic fiber and collagen genes in familial adolescent idiopathic scoliosis. Journal of orthopaedic research. 1996;14(6):994-9.

190. Zorkoltseva I, Lubinsky O, Sharipov R, Zaidman A, Axenovich T, Dimshits G. Analysis of polymorphism of the number of tandem repeats in the aggrecan gene exon G3 in the families with idiopathic scoliosis. Russian Journal of Genetics. 2002;38(2):196-200.

191. Marosy B, Justice CM, Nzegwu N, Kumar G, Wilson AF, Miller NH. Lack of association between the aggrecan gene and familial idiopathic scoliosis. Spine. 2006;31(13):1420-5.

192. Morcuende JA, Minhas R, Dolan L, Stevens J, Beck J, Wang K, et al. Allelic variants of human melatonin 1A receptor in patients with familial adolescent idiopathic scoliosis. Spine. 2003;28(17):2025-8.

193. Fan Y-H, Song Y-Q, Chan D, Takahashi Y, Ikegawa S, Matsumoto M, et al. SNP rs11190870 near LBX1 is associated with adolescent idiopathic scoliosis in southern Chinese. Journal of human genetics. 2012;57(4):244-6.

194. Gao W, Peng Y, Liang G, Liang A, Ye W, Zhang L, et al. Association between common variants near LBX1 and adolescent idiopathic scoliosis replicated in the Chinese Han population. PloS one. 2013;8(1):e53234.

195. Jiang H, Qiu X, Dai J, Yan H, Zhu Z, Qian B, et al. Association of rs11190870 near LBX1 with adolescent idiopathic scoliosis susceptibility in a Han Chinese population. European Spine Journal. 2013;22(2):282-6.

196. Liang J, Xing D, Li Z, Chua S, Li S. Association between rs11190870 polymorphism near LBX1 and susceptibility to adolescent idiopathic scoliosis in East Asian population: a genetic meta-analysis. Spine. 2014;39(11):862-9.

197. Londono D, Kou I, Johnson TA, Sharma S, Ogura Y, Tsunoda T, et al. A meta-analysis identifies adolescent idiopathic scoliosis association with LBX1 locus in multiple ethnic groups. Journal of medical genetics. 2014;51(6):401-6.

198. Guo L, Yamashita H, Kou I, Takimoto A, Meguro-Horike M, Horike S-i, et al. Functional investigation of a non-coding variant associated with adolescent idiopathic scoliosis in Zebrafish: elevated expression of the ladybird homeobox gene causes body axis deformation. PLoS Genet. 2016;12(1):e1005802.

199. Kou I, Takahashi Y, Johnson TA, Takahashi A, Guo L, Dai J, et al. Genetic variants in GPR126 are associated with adolescent idiopathic scoliosis. Nature genetics. 2013;45(6):676-9.

200. Xu J-F, Yang G-h, Pan X-H, Zhang S-J, Zhao C, Qiu B-S, et al. Association of GPR126 gene polymorphism with adolescent idiopathic scoliosis in Chinese populations. Genomics. 2015;105(2):101-7.

201. Miyake A, Kou I, Takahashi Y, Johnson TA, Ogura Y, Dai J, et al. Identification of a susceptibility locus for severe adolescent idiopathic scoliosis on chromosome 17q24. 3. PloS one. 2013;8(9):e72802.

202. Ogura Y, Kou I, Miura S, Takahashi A, Xu L, Takeda K, et al. A functional SNP in BNC2 is associated with adolescent idiopathic scoliosis. The American Journal of Human Genetics. 2015;97(2):337-42.

203. Sharma S, Londono D, Eckalbar WL, Gao X, Zhang D, Mauldin K, et al. A PAX1 enhancer locus is associated with susceptibility to idiopathic scoliosis in females. Nature communications. 2015;6.

204. Grauers A, Einarsdottir E, Gerdhem P. Genetics and pathogenesis of idiopathic scoliosis. Scoliosis and Spinal Disorders. 2016;11(1):45.

205. Grauers A, Wang J, Einarsdottir E, Simony A, Danielsson A, Åkesson K, et al. Candidate gene analysis and exome sequencing confirm LBX1 as a susceptibility gene for idiopathic scoliosis. The Spine Journal. 2015;15(10):2239-46.

206. Aulisa L, Papaleo P, Pola E, Angelini F, Aulisa AG, Tamburrelli FC, et al. Association between IL-6 and MMP-3 gene polymorphisms and adolescent idiopathic scoliosis: a case-control study. Spine. 2007;32(24):2700-2.

207. Suh KT, Eun I-S, Lee JS. Polymorphism in vitamin D receptor is associated with bone mineral density in patients with adolescent idiopathic scoliosis. European Spine Journal. 2010;19(9):1545-50.

208. Wang H, Wu Z, Zhuang Q, Fei Q, Zhang J, Liu Y, et al. Association study of tryptophan hydroxylase 1 and arylalkylamine N-acetyltransferase polymorphisms with adolescent idiopathic scoliosis in Han Chinese. Spine. 2008;33(20):2199-203.

209. Inoue M, Minami S, Nakata Y, Kitahara H, Otsuka Y, Isobe K, et al. Association between estrogen receptor gene polymorphisms and curve severity of idiopathic scoliosis. Spine. 2002;27(21):2357-62.

210. Wu J, Qiu Y, Zhang L, Sun Q, Qiu X, He Y. Association of estrogen receptor gene polymorphisms with susceptibility to adolescent idiopathic scoliosis. Spine. 2006;31(10):1131-6.

211. Zhang H-Q, Lu S-J, Tang M-X, Chen L-Q, Liu S-H, Guo C-F, et al. Association of estrogen receptor β gene polymorphisms with susceptibility to adolescent idiopathic scoliosis. Spine. 2009;34(8):760-4.

212. Peng Y, Liang G, Pei Y, Ye W, Liang A, Su P. Genomic polymorphisms of G-protein estrogen receptor 1 are associated with severity of adolescent idiopathic scoliosis. International orthopaedics. 2012;36(3):671-7.

213. Yeung H, Tang N, Lee K, Ng B, Hung V, Kwok R, et al. Genetic association study of insulin-like growth factor-I (IGF-I) gene with curve severity and osteopenia in adolescent idiopathic scoliosis. Studies in health technology and informatics. 2005;123:18-24.

214. Zhou S, Qiu X-s, Zhu Z-z, Wu W-f, Liu Z, Qiu Y. A single-nucleotide polymorphism rs708567 in the IL-17RC gene is associated with a susceptibility to and the curve severity of adolescent idiopathic scoliosis in a Chinese Han population: a case-control study. BMC musculoskeletal disorders. 2012;13(1):181.

215. Mao S, Xu L, Zhu Z, Qian B, Qiao J, Yi L, et al. Association between genetic determinants of peak height velocity during puberty and predisposition to adolescent idiopathic scoliosis. Spine. 2013;38(12):1034-9.

216. Ryzhkov II, Borzilov EE, Churnosov MI, Ataman AV, Dedkov AA, Polonikov AV. Transforming growth factor beta 1 is a novel susceptibility gene for adolescent idiopathic scoliosis. Spine. 2013;38(12):E699-E704.

217. Buchan JG, Alvarado DM, Haller GE, Cruchaga C, Harms MB, Zhang T, et al. Rare variants in FBN1 and FBN2 are associated with severe adolescent idiopathic scoliosis. Human molecular genetics. 2014;23(19):5271-82.

218. Baschal EE, Wethey CI, Swindle K, Baschal RM, Gowan K, Tang NL, et al. Exome sequencing identifies a rare HSPG2 variant associated with familial idiopathic scoliosis. G3: Genes| Genomes| Genetics. 2015;5(2):167-74.

219. Patten SA, Margaritte-Jeannin P, Bernard J-C, Alix E, Labalme A, Besson A, et al. Functional variants of POC5 identified in patients with idiopathic scoliosis. The Journal of clinical investigation. 2015;125(3):1124-8.

220. Haller G, Alvarado D, Mccall K, Yang P, Cruchaga C, Harms M, et al. A polygenic burden of rare variants across extracellular matrix genes among individuals with adolescent idiopathic scoliosis. Human molecular genetics. 2016;25(1):202-9.

221. Li W, Li Y, Zhang L, Guo H, Tian D, Li Y, et al. AKAP2 identified as a novel gene mutated in a Chinese family with adolescent idiopathic scoliosis. Journal of medical genetics. 2016:jmedgenet-2015-103684.

222. O'Brien M, Costin B, Miles M. Using genome-wide expression profiling to define gene networks relevant to the study of complex traits: from RNA integrity to network topology. International review of neurobiology. 2012;104:91.

223. Fendri K, Patten SA, Kaufman GN, Zaouter C, Parent S, Grimard G, et al. Microarray expression profiling identifies genes with altered expression in Adolescent Idiopathic Scoliosis. European Spine Journal. 2013;22(6):1300-11.

224. Zhuang Q, Mao W, Xu P, Li H, Sun Z, Li S, et al. Identification of Differential Genes Expression Profiles and Pathways of Bone Marrow Mesenchymal Stem Cells of Adolescent Idiopathic Scoliosis Patients by Microarray and Integrated Gene Network Analysis. Spine. 2016;41(10):840-55.

225. Hadley MN. Spine update: genetics of familial idiopathic scoliosis. Spine. 2000;25(18):2416-8.

226. Ward K, Ogilvie J, Argyle V, Nelson L, Meade M, Braun J, et al. Polygenic inheritance of adolescent idiopathic scoliosis: a study of extended families in Utah. American journal of medical genetics Part A. 2010;152(5):1178-88.

227. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nature Reviews Genetics. 2005;6(2):95-108.

228. Hardy J, Singleton A. Genomewide association studies and human disease. New England Journal of Medicine. 2009;360(17):1759-68.

229. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461(7265):747-53.

230. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, et al. Missing heritability and strategies for finding the underlying causes of complex disease. Nature Reviews Genetics. 2010;11(6):446-50.

231. Maher B. The case of the missing heritability. Nature. 2008;456(7218):18.

232. Reich DE, Lander ES. On the allelic spectrum of human disease. TRENDS in Genetics. 2001;17(9):502-10.

233. Iyengar SK, Elston RC. The genetic basis of complex traits: rare variants or "common gene, common disease"? Linkage Disequilibrium and Association Mapping: Analysis and Applications. 2007:71-84.

234. Gorlov IP, Gorlova OY, Sunyaev SR, Spitz MR, Amos CI. Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. The American Journal of Human Genetics. 2008;82(1):100-12.

235. Wang Z, Liu X, Yang B-Z, Gelernter J. The role and challenges of exome sequencing in studies of human diseases. Frontiers in genetics. 2013;4:160.

236. Parla JS, Iossifov I, Grabill I, Spector MS, Kramer M, McCombie WR. A comparative analysis of exome capture. Genome biology. 2011;12(9):R97.

237. Baron M. The search for complex disease genes: fault by linkage or fault by association? Molecular psychiatry. 2001;6(2):143.

238. Lin D-C, Hao J-J, Nagata Y, Xu L, Shang L, Meng X, et al. Genomic and molecular characterization of esophageal squamous cell carcinoma. Nature genetics. 2014;46(5):467-73.

239. Katoh Y, Katoh M. Comparative integromics on FAT1, FAT2, FAT3 and FAT4. International journal of molecular medicine. 2006;18(3):523.

240. Sadeqzadeh E, Bock CE, Thorne RF. Sleeping giants: emerging roles for the fat cadherins in health and disease. Medicinal research reviews. 2014;34(1):190-221.

241. Hayes M, Gao X, Lisa XY, Paria N, Henkelman RM, Wise CA, et al. ptk7 mutant zebrafish models of congenital and idiopathic scoliosis implicate dysregulated Wnt signalling in disease. Nature communications. 2014;5.

242. Andersen MR, Farooq M, Koefoed K, Kjaer KW, Simony A, Christensen ST, et al. Mutation of the planar cell polarity gene VANGL1 in adolescent idiopathic scoliosis. Spine. 2017.

243. Le Pabic P, Ng C, Schilling TF. Fat-Dachsous signaling coordinates cartilage differentiation and polarity during craniofacial development. PLoS Genet. 2014;10(10):e1004726.

244. Sharma S, Gao X, Londono D, Devroy SE, Mauldin KN, Frankel JT, et al. Genome-wide association studies of adolescent idiopathic scoliosis suggest candidate susceptibility genes. Human molecular genetics. 2011;20(7):1456-66.

245. Sun G, Lam T, Bobby N, Yim P, Lee K, Moreau A, et al. High osteopontin plasma level associated with abnormal cortical bone mineral density in girls with adolescent idiopathic scoliosis. Stud Health Technol Inform. 2012;176:457.

246. Yadav MC, Huesa C, Narisawa S, Hoylaerts MF, Moreau A, Farquharson C, et al. Ablation of osteopontin improves the skeletal phenotype of Phospho1−/− mice. Journal of Bone and Mineral Research. 2014;29(11):2369-81.

247. Xie N, Li M, Wu T, Liu J, Wang B, Tang F. Does elevated osteopontin level play an important role in the development of scoliosis in bipedal mice? The Spine Journal. 2015;15(7):1660-4.

248. Hoo C, Gatam L. Role of Remodelling in Adolescent Idiopathic Scoliosis: an Evaluation of Osteopontin Level. The Journal of Indonesian Orthopaedic. 2012;40(2).

249. Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. Nature Reviews Genetics. 2011;12(5):363-76.

250. Frazer KA, Murray SS, Schork NJ, Topol EJ. Human genetic variation and its contribution to complex traits. Nature Reviews Genetics. 2009;10(4):241-51.

251. Timberlake AT, Choi J, Zaidi S, Lu Q, Nelson-Williams C, Brooks ED, et al. Two locus inheritance of non-syndromic midline craniosynostosis via rare SMAD6 and common BMP2 alleles. Elife. 2016;5:e20125.

252. Mórocz M, Czibula Á, Grózer ZB, Szécsényi A, Álmos PZ, Raskó I, et al. Association study of BMP4, IL6, Leptin, MMP3, and MTNR1B gene promoter polymorphisms and adolescent idiopathic scoliosis. Spine. 2011;36(2):E123-E30.

253. Hermus JP, van Rhijn LW, van Ooij A. Non-genetic expression of adolescent idiopathic scoliosis: a case report and review of the literature. European Spine Journal. 2007;16(3):338-41.

254. Worthington V, Shambaugh P. Nutrition as an environmental factor in the etiology of idiopathic scoliosis. Journal of manipulative and physiological therapeutics. 1992;16(3):169-73.

255. McMaster M, Lee A, Burwell R, editors. Physical activities of patients with adolescent idiopathic scoliosis (AIS) compared with a control group: implications for etiology and possible prevention. Orthopaedic Proceedings; 2006: Orthopaedic Proceedings.

256. Grivas TB, Vasiliadis E, Mouzakis V, Mihas C, Koufopoulos G. Association between adolescent idiopathic scoliosis prevalence and age at menarche in different geographic latitudes. Scoliosis. 2006;1(1):9.

257. Ryan M, Nachemson A. Thoracic adolescent idiopathic scoliosis: perinatal and environmental aspects in a Swedish population and their relationship to curve severity. Journal of Pediatric Orthopaedics. 1987;7(1):72-7.

258. De George FV, Fisher RL. Idiopathic scoliosis: genetic and environmental aspects. Journal of medical genetics. 1967;4(4):251.

259. Smyrnis T, Antoniou D, Valavanis J, Zachariou C. Idiopathic scoliosis: characteristics and epidemiology. Orthopedics. 1987;10(6):921-6.

260. Bibby S, Meir A, Fairbank J, Urban J. Cell viability and the physical environment in the scoliotic intervertebral disc. Studies in health technology and informatics. 2001;91:419-21.

261. Meir A, McNally D, Fairbank J, Jones D, Urban J. The internal pressure and stress environment of the scoliotic intervertebral disc—a review. Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine. 2008;222(2):209-19.

262. Smith P, Day N. The design of case-control studies: the influence of confounding and interaction effects. International journal of epidemiology. 1984;13(3):356-65.

263. Butcher LM, Beck S. Future impact of integrated high-throughput methylome analyses on human health and disease. Journal of Genetics and Genomics. 2008;35(7):391-401.

264. Petronis A. Epigenetics and twins: three variations on the theme. Trends in Genetics. 2006;22(7):347-50.

265. Bjornsson HT, Fallin MD, Feinberg AP. An integrated epigenetic and genetic approach to common human disease. TRENDS in Genetics. 2004;20(8):350-8.

266. Jessen HM, Auger AP. Sex differences in epigenetic mechanisms may underlie risk and resilience for mental health disorders. Epigenetics. 2011;6(7):857-61.

267. Murrell A, Rakyan VK, Beck S. From genome to epigenome. Human molecular genetics. 2005;14(suppl 1):R3-R10.

268. Silahtaroglu A, Stenvang J. MicroRNAs, epigenetics and disease. Essays in biochemistry. 2010;48:165-85.

269. Felekkis K, Touvana E, Stefanou C, Deltas C. microRNAs: a newly described class of encoded molecules that play a role in health and disease. Hippokratia. 2010;14(4):236-40.

270. Chiquet M. Regulation of extracellular matrix gene expression by mechanical stress. Matrix biology. 1999;18(5):417-26.

ANNEX   I

Review Article

Biochemistry of Idiopathic Scoliosis

From Discovery to Diagnostic Biomarkers

Dina Nada MSc [1,2] and Alain Moreau PhD [1,2,3,4] [*]

[1] Viscogliosi Laboratory in Molecular Genetics of Musculoskeletal Diseases, Sainte-Justine University Hospital Research Center, Montreal, Qc, Canada

[2] Department of Biomedical Sciences, Faculty of Medicine, Université de Montréal, Montreal, Qc, Canada;

[3] Department of Stomatology, Faculty of Dentistry, Université de Montréal, Montreal, Qc, Canada;

[4] Department of Biochemistry and Molecular Medicine, Faculty of Medicine, Université de Montréal, Montréal, Montreal, Qc, Canada

*Corresponding author: Alain Moreau PhD

Email address: alain.moreau@recherche-ste-justine.qc.ca

Keywords: Idiopathic scoliosis, Scoliotic etiopathogenesis, Biochemical factors

Abstract:

Idiopathic scoliosis (IS) is a condition of the spine where the spinal curve is deformed. The condition of IS appears at the onset of puberty and is most often seen in women. Current scientific literature reveals potential biomarkers of the onset and progression of IS, which include hormones, systemic factors, hematological factors, and bone metabolism factors. Hormones that include insulin like growth factor-1, melatonin, estrogen, ghrelin, and leptin, have been implicated as potential diagnostic biomarkers in IS. Systemic factors, such as osteopontin and Gi proteins; proteins involved in bone metabolism, such as matrilin-1, cartilage oligomeric matrix protein (COMP), osteocalcin; and the hematological protein calmodulin, are the other potential candidates for diagnosing IS. These diagnostic biomarkers do not all act in isolation but are interconnected through signaling pathways. There is crosstalk between melatonin and estrogen; estrogen and growth hormones; or between melatonin and calmodulin. Similarly, increased levels of osteopontin regulate Gi protein-coupled receptor (GiPCR) dysfunction. Other studies implicate single nucleotide polymorphisms (SNPs) in calmodulin, melatonin receptors, or matrilin-1, as potential risk factors of increasing the spinal curvature. While theories, such as the double neuro-osseous theory implicate increased sensitivity of the hypothalamus to leptin in the development of IS, the downregulation of leptin has also been indicated in the etiopathogenesis of IS. Although data on the potential candidates for diagnosis appears promising, it is acknowledged that larger sample studies with greater statistical power is required to confirm the diagnostic potential of blood biomarkers.

Introduction

Idiopathic scoliosis (IS) indicates a deformed spinal curve that has no known cause or etiology [1]. Scoliosis has been predominantly observed in females during puberty and affects 1% to 4% of children. Nearly 80% of idiopathic scoliosis is in the form of adolescent idiopathic scoliosis (AIS) [1, 2]. Research into the causes of IS have unearthed evidence for potential biomarkers [1]. So far, the diagnosis of IS is based on measuring the Cobb angle of the spine. When the value of the Cobb angle is ≥10º, the patient is diagnosed with scoliosis. However, physicians arrive at this diagnosis only after ruling out all other possible disorders [3]. Prior to the progression of the spinal curvature, it would be useful to identify markers in the body that could indicate the onset or the progression of IS. In this chapter, we will cover the current research data on the biochemical factors associated with IS. We have classified the biomarkers into four main groups: hormones, systemic factors, hematological factors, and bone metabolism factors. Studies that cover all the key aspects of these factors have been included. Hence, gene association studies, protein function analyses, and protein interactions have been recorded. We will discuss the validity of the studies and the contradictory data for each factor. Through this process, we will understand the potential role of each factor in the pathogenesis of IS. We will also elaborate on the hypotheses associated with the biochemical factors and their potential relevance in the pathogenesis of IS.

1. Hormonal factors:

1.1. Growth hormone:

The vast majority of IS cases occur in adolescence during the pubertal growth spurt. The association of abnormal growth patterns with the development of scoliotic curves is

well known [4]. A number of studies have confirmed that adolescent idiopathic scoliosis (AIS) patients are generally taller than age-matched controls [5, 6]. A study that involved 598 IS patients, compared the growth pattern as well as other anthropometric parameters of the IS patients to normal controls at both the prepubertal and the postpubertal stages [7]. Prior to puberty, the length of the different body segments was significantly shorter in the IS patients compared with the normal controls. In contrast, following puberty, the length of the different body segments was significantly longer in the IS patients, and all the anthropometric parameters were greatly associated with the severity of the curve. This indicated that the abnormality in the growth rate of IS patients is specifically associated with the pubertal stage. The abnormal growth pattern in IS patients was further confirmed by another study that included 611 IS girls and 296 age-matched controls [8]. The study observed that the severe cases of spinal curve progression displayed late menarche and higher skeletal growth rate between the ages of 12 and 16 years.

Growth hormone (GH) is the hormone responsible for growth stimulation. The correlation of GH with IS has been studied since the 1970s, where Willner et al observed significantly higher plasma levels of GH in IS girls compared with age-matched controls [9]. It has been shown that treatment with GHs could lead to scoliosis [10]. However, this observation was contradicted by other studies [11] and the results of other stimulating tests were inconclusive[9]. Later, Willner's group reported an increase in the morning fasting levels of the growth hormone in IS girls compared with the healthy controls [12]. Another study that compared the GH profile of scoliotic girls with matched controls for 24 hrs showed that only scoliotic girls exhibited higher levels of the hormone in the early stages of puberty, suggesting an early growth spurt in IS girls [13].

A genetic study that involved 30 IS patients and 30 matched controls identified a single nucleotide polymorphism (SNP) in the GH receptor gene as a predisposing factor to IS [14]. This was, however, contradicted by a large study of 510 IS patients and 363 controls, which investigated the GH receptor and found no association of GH receptor variations with IS [15].

Growth hormone itself is very variable and hard to assess but its activity is correlated with insulin-like growth factor-1 (IGF-1) that in turn is associated with puberty [16, 17]. Insulin-like growth factor is known to play an important role in bone growth. One study suggested increased levels of IGF-1 and estrogen in IS individuals after they achieved a growth spurt and recorded the maximum growth rate [18]. Another study demonstrated that IGF-1 and IGF binding protein-3 (IGFBP-3) were significantly associated with an accelerated rate of spinal curve progression in females with IS[19]. Yeung et al studied two SNPs (rs5742612 and rs2288377) in the *IGF-1* promoter in a Chinese cohort of 506 IS girls and 227 healthy girls. They observed that the two SNPs affected the progression of the spinal curve but did not initiate or predispose individuals to IS [20]. Similarly, a Korean study reported an association between rs5742612 that was present upstream of *IGF-1* and the onset and progression of IS [21]. However, there were research data that failed to find any association between IGF-1 and IS. One study observed that SNPs found in the GH receptor and *IGF-1* were not associated with IS in a Han Chinese population [22]. This was confirmed by another report that failed to find an association between *IGF-1* and IS in a Japanese population [23]. Liu et al reported in a study of 200 patients and 200 controls that SNPs in *IGF-1R* were not associated with either IS susceptibility or the severity of curve progression [24]. This conflict between different reports may warrant additional large-scale

genetic studies that have sufficient statistical power, in different populations to validate the involvement of *IGF-1* and its receptor genes in IS.

3.1.2. Estrogens:

The occurrence of scoliosis in females is almost double that of males, and this ratio reaches 10:1 for Cobb angles $\geq 30^{o}$ [25]. Taking into consideration the high prevalence of this disorder in females and the concurrent onset of scoliosis with sexual maturation, it is plausible to hypothesize the involvement of estrogen and/or its receptors in the pathophysiology of the disease. Indeed, some studies report that estrogens and estrogen antagonists have a direct effect on the onset of scoliosis [26, 27]. Estrogen is critical for puberty [28]. Some researchers have studied blood estrogen levels in IS but their results are inconclusive. Kulis et al reported a reduction of plasma estrogen levels in females with scoliosis [29]. This was supported by another study that demonstrated lower average estrogen and progesterone levels in IS females compared with healthy controls of the same age [30]. It was also reported that low levels of estrogen increase the susceptibility of ballet dancers to scoliosis [31]. It was suggested that the effect of estrogen on IS was not due to its concentration but due to its interaction with target cells [32], specifically bone cells [33]. Interestingly, one study reported that there were elevated levels of testosterone in IS girls relative to age-matched controls [34] .

As mentioned earlier, most IS cases occur at puberty. This stage involves significant bone remodeling [35]. The involvement of estrogen in bone remodeling is well known [36]. Estrogen plays a role in the formation of bones; it enhances the activity of alkaline phosphatase, the synthesis of collagen, and the deposition of calcium in the extracellular

matrix [37]. It also affects bone growth, maturation, and turnover. Reduced estrogen levels lead to a reduced differentiation of osteoblasts, which affects both bone mineralization as well as its mechanical properties [32]. Estrogen plays a role in bone metabolism by affecting both bone cell types: osteoblasts and osteoclasts [38]. Estrogen affects the secretion of cytokines needed for the formation of osteoclasts [39] and inhibits IL-6 cytokine. The cytokine IL-6 enhances bone resorption and activates osteoclasts [32].

The association between osteopenia and IS is widely recognized [40] but it has been mainly demonstrated in different Asian pediatric populations. It has been reported by many researchers that osteopenia occurs at a high frequency in IS [41, 42]. Moreover, osteopenia or osteoporosis has been shown to be involved in the progression of the spinal curvature [42]. A persistent lower bone mineral density was reported in IS patients compared with healthy controls [43]. The severity of scoliosis has been inversely correlated with bone mineral density and bone mineral content [44]. Defects in estrogen function or reduced estrogen levels should be taken into account as driving factors for osteopenia and reduced bone mineralization in IS girls [32].

Letellier et al demonstrated the important role of estrogen in the functioning of osteoblasts. The interaction between estrogen and melatonin hormones and their subsequent effect on osteoblasts were established [33]. The involvement of melatonin in the pathogenesis of IS has been thoroughly investigated and will be discussed in detail later. Briefly, melatonin exerts its functions through Gi-protein-coupled receptors with the subsequent inhibition of cAMP. The occurrence of a differential melatonin signaling dysfunction in the osteoblasts isolated from IS patients resulted in an intracellular elevation of cAMP levels [45]. In a certain subset of IS patients, this effect was due to a switch in the

coupling of melatonin receptors from Gi to Gs [33]. Interestingly, osteoblasts obtained from this IS subgroup and treated with 17β-estradiol together with melatonin resulted in a significant reduction of cAMP levels compared with melatonin treatment alone. These results can be explained by the fact that Gi proteins are also known to be regulated by estrogen activity [46]. Indeed, exposure to 17β -estradiol is known to increase Gi and Gq alpha proteins and decreased Gs alpha subunit expression.

The interaction of estrogen with the melatonin signaling pathway was supported by the works from Acaroglu's group [47]. In this study, they treated melatonin-deficient mice models that were rendered bipedal, with both tamoxifen and raloxifen, which are selective estrogen receptor modulators (SERM) and have estrogen agonistic effects. Treatment of mice with tamoxifen or raloxifen did not prevent the occurrence of scoliosis but interfered with the aggravation of the scoliotic curves. Previous studies have shown the effectiveness of tamoxifen in reducing the severity of scoliotic curves in experimental mice models [48]. However, it is unclear whether tamoxifen's effects were solely due to its role as estrogen agonist or through off-target effects given the fact that tamoxifen acts also as a non-specific calmodulin inhibitor [49], where calmodulin was previously implicated in scoliosis severity [50, 51].

Estrogen stimulates the maturation of the physis, which leads to a linear growth spurt [28, 52, 53]. This action is mediated through its receptors; estrogen receptor α (ERα) and estrogen receptor β (ERβ), which are expressed in the growth plate [54]. At low levels, estrogen stimulates growth by enhancing the secretion of GH, while at high levels, estrogen interferes with GH secretion and subsequently terminates longitudinal growth [32]. This

interplay with GH could affect IS children who are known to be taller than healthy subjects of the same age.

Polymorphisms in the gene encoding ERα (*ESR1*) were shown to be associated with bone mineral density that is known to be abnormal in many IS cases [55]. Skeletal abnormalities have been shown to be associated with mutations in both *ESR1* and *CYP19* [56, 57]. The *CYP19* encodes the aromatase enzyme involved in estrogen synthesis. During the last decade there has been extensive research on the polymorphisms of genes encoding estrogen receptors in an effort to detect a genetic correlation between these receptors and IS. However, there is a high discrepancy between the published studies, and this genetic correlation is still debated. A summary of these studies is presented in Table 5.1.

There has been an extended effort by Fendri et al to uncover the role of estrogen in IS. Osteoblasts from IS patients were treated with estrogen, which was followed by microarray analysis. Fendri et al identified a list of genes that were differentially expressed in IS cells versus normal cells and played a role in IS. Some of the genes are regulated by estrogen [58]. In general, it has been suggested that estrogens may not be involved in the etiology of IS but may play an important role in the progression or the repression of the disease [32]. This is due to the interaction of estrogen with many factors that are involved in the etiopathogenesis of IS, bone modeling and remodeling, osteopenia, growth factors, the melatonin signaling pathway, and platelet regulatory factors (eg, calmodulin that is discussed below) [32].

Table 1. Summary of the genetic studies performed on ERs and their correlation to IS

Reference	Study	Conclusions
Inoue et al (2002) [59]	SNPs PvuII and XbaI in *ERα* were studied in 304 IS girls	An association between the SNP XbaI and curve progression was observed
Wu et al (2006) [60]	The polymorphism XbaI in *ERα* was studied in 202 IS patients versus 174 controls.	This SNP was shown to be correlated with IS susceptibility
Tang et al (2006) [61]	SNPs PvuII, XbaI in *ERα* were studied in a Chinese cohort of 540 IS patients compared with 260 controls	The two SNPs did not show any correlation with IS or its severity
Zhang et al (2009)[62]	Polymorphism in the exon ØK of *ERβ* was studied in 218 IS patients versus 140 normal subjects	The studied polymorphism appeared to be correlated with IS predisposition and severity of the curves
Esposito et al (2009)[30]	Detected 4 SNPs in the exons that encode the steroid binding domain, and 2 SNPs in the exons that encode the transactivation domain of *ERα*	Association with clinical manifestations of IS patients
Zhao et al (2009) [63]	rs2234693 (Pvu II) and rs9340799 (Xba I) in *ERα* were studied in 67 Chinese IS patients and 100 controls	rs2234693 (Pvu II) was found to be associated with the double curve pattern, the thoracic curve, and the severity of IS
Takahashi et al (2011) [64]	rs9340799 and rs1256120 in the estrogen receptor genes were studied in a Japanese cohort of 798 IS patients vs 637 controls	Neither of the two SNPs was found to be associated with either susceptibility or progression of IS
Xu et al (2011)	SNPs in the *ERα* and *ERβ*	rs9340799 in *ERα* appeared to play

Reference	Study	Conclusions
[65]	together with other genes were tested as to whether they can be used as markers for predicting the outcome of brace treatment. This study included 312 IS patients.	a role in predicting the efficiency of brace treatment
Peng et al (2012) [66]	16 SNPs in the G protein-coupled estrogen receptor 1 (*GPER*) were studied in a total of 389 IS cases and 338 controls	3 SNPs (rs3808351, rs10269151 and rs426655s3) were found to be associated with curve progression in IS patients
Ogura et al (2013) [67]	rs3808351, rs10269151, and rs4266553 in the G protein-coupled estrogen receptor (*GPER)* were studied in a Japanese cohort of 2117 IS patients	No association was observed between the studied SNPs and curve progression severity
Janusz et al (2013) [68]	rs934099 (XbaI) and rs2234693 (PvuII) in the *ERα* were studied in 287 Caucasian females and 182 controls	No association was found between either of the studied SNPs and severity of IS
Yang et al (2014) [69]	This meta-analysis includes the results of six previous articles that studied the polymorphisms XbaI and PvuII in *ERα*	No association between the studied SNPs and IS
Chen et al (2014) [70]	This meta-analysis included the results of four previous studies that studied the SNP rs9340799 of *ERα*	No association between the studied SNP and predisposition to IS

Reference	Study	Conclusions
Kotwicki et al (2014) [71]	rs1256120, rs4986938 and rs1256049 in *ERβ* were studied in 248 Caucasian females versus 243 healthy females	None of the studied SNPs showed association with IS susceptibility
Janusz et al (2014) [72]	rs9340799 (XbaI) and rs2234693 (PvuII) in *ERα*, and rs4986938 (AluI) and rs1256049 (RasI) in *ERβ* were studied in 208 IS Caucasian females to test their association with age at menarche.	None of the studied SNPs showed any association with age at menarche

3.1.3. Leptin:

Leptin is a hormone produced by adipose cells and plays a role in regulating the energy balance. Leptin also plays an important function in bone metabolism. Due to its role in the crosstalk between energy and bone metabolism, it was hypothesized that it is most likely involved in the pathogenesis of scoliosis. Reduced leptin serum levels in IS patients are reported in many studies and are associated with a reduced bone mass [73-75], reduced bone strength [76], and abnormal growth parameters [75]. Moreover, the levels of sOB-R (soluble leptin receptor) are shown to be elevated in IS patients compared with controls [73, 74], which reflects abnormal bioavailability of leptin and leptin signaling malfunction in those patients. This abnormal signaling was also manifested in a report where the osteoblasts of IS patients were significantly less responsive to leptin when compared with those of controls [77]. It was suggested that reduced serum leptin levels may be a result of the reduced ability of adipogenesis in IS [78].

A double neuro-osseous theory was proposed to explain the role of leptin in the pathogenesis of scoliosis [79]. This theory states that there is a distinct contribution of the autonomic as well as the somatic nervous systems to the development of IS. In the autonomic nervous system, due to the enhanced sensitivity of the hypothalamus to circulating leptin levels, the sympathetic nervous system causes a dissymmetry of the axial skeleton during growth, which affects the width of the trunk. That effect is enforced by the somatotropic (IGF/growth hormone) axis. In the somatic nervous system, a dysfunction in the postural mechanism is another way of inducing spinal deformities. The postural mechanisms of the trunk and the spine are affected by the somatic nervous system [79].

On the other hand, other researchers suggest that the reduced levels of circulating leptin may result from a larger metabolic defect. It has been shown that melatonin signaling can lead to a reduction in the levels of leptin [80]. Where another study showed that the levels of leptin are reduced in pinealectomized rats and are elevated with melatonin administration [81]. On the other hand, it has been proven that cytosolic cAMP inhibits leptin synthesis [82]. This finding is further strengthened by the fact that there is a generalized Gi-signaling dysfunction in IS patients, which leads to the accumulation of cAMP [45, 83, 84]. These findings together can help explain the reduced circulating levels of leptin that are observed in IS patients.

3.1.4. Ghrelin:

Ghrelin is another hormone that plays a role in regulating energy as well as hunger. The ghrelin hormone is secreted by the ghrelin cells in the gastrointestinal tract [85]. Ghrelin also plays a role in bone metabolism [86]. The involvement of ghrelin in the pathogenesis of

IS has recently been suggested. A study reported high circulating levels of ghrelin in IS patients compared with healthy subjects. The data of the study hypothesized the involvement of ghrelin in the pathophysiology of IS [87].

3.1.5. Melatonin:

Melatonin is a hormone produced mostly by the pineal gland. It mainly functions in the regulation of the circadian rhythm and in the treatment of some sleep disorders. It plays a role in reproduction, aging, some cancers, and bone metabolism. It is also a potent antioxidant that helps in the protection of the DNA. Melatonin is produced as a response to darkness and is inhibited by light. Upon the onset of darkness, the receptors of the retina secrete norepinephrine that increases the number of adrenergic receptors of the pineal gland. This stimulates the activity of the enzyme arylalkylamine *N*-acetyltransferase (AANAT), the key regulator of melatonin synthesis from tryptophan [88, 89]. Melatonin is mainly metabolized in the liver producing 6-sulfa-oxymelatonin that is excreted in the urine. The level of 6-sulfatoxymelatonin in the urine is correlated to serum melatonin levels [89]. Melatonin has two groups of receptors; melatonin receptor 1 (ML1) (high affinity) and melatonin receptor 2 (ML2) (low affinity) receptors [90]. There are two subtypes of ML1 receptors known as melatonin receptor 1A (MTNR1A) and melatonin receptor 1B (MTNR1B) receptors [91, 92]. Both receptors are Gi-protein-coupled receptors that inhibit adenylate cyclase activity [89].

Pinealectomized animal models have been proposed as good models to study IS. In the late 1950s and early 1960s, Thillard et al were the first to provide evidence on the development of deformities in the vertebral column in chickens by pinealectomy [93]. This

was followed by the work of Dubousset et al in the early 1980s where scoliosis was induced in young chickens by provoking pineal and diencephalic damages [94]. This field of research was strongly supported thereafter by the work of Machida et al. The group of Machida et al showed that if pinealectomy is performed shortly after hatching, chickens will consistently develop scoliosis similar to that in human IS. Moreover, these researchers showed that scoliosis could be prevented by intramuscular implantation of the pineal body in pinealectomized chickens [95]. They suggested that melatonin deficiency plays a role in the abnormal development of the asymmetry of the proprioceptive system, including the spine and the paraspinal muscles [96]. They then demonstrated that treating pinealectomized chickens with serotonin, the precursor of melatonin, leads to the development of scoliosis in 73.3% of the chickens, while treatment with melatonin leads to the development of scoliosis in only 20% of the chickens [96]. The ineffectiveness of serotonin to prevent scoliosis could be explained by the fact that serotonin is unable to cross the blood-brain barrier [97]. Thus, a precursor of serotonin, 5-hydroxytryptophan that has the ability of crossing the blood-brain barrier was found to be more effective in scoliosis prevention in the chickens [98]. Machida et al were also able to reinforce their findings by demonstrating that not only chickens but also all bipedal pinealectomized rats developed scoliosis [99]. Moreover, the same group was able to demonstrate scoliosis in all the C57BL/6 mice (a strain of mice genetically deficient in melatonin) that were rendered bipedal [100]. Machida et al also suggested that deficiency of melatonin has an effect in the prognosis of IS [101]. Bagnall et al reinforced the observations of Machida et al by demonstrating that the development of scoliosis in pinealectomized chickens is not due to an artifact of the surgical procedure [102].

On the other hand, the data presented by other researchers contradicted the work of Machida et al [103-107]. The development of scoliosis in all pinealectomized chickens could not be reproduced in subsequent studies. Instead, the frequency of scoliosis was found to range from 50% to 60% [103, 104, 107]. Consistently, similar results were obtained in C57BL/6 mice and rat studies [48, 108]. A 2-year study that tried to reproduce the results in primates used 18 pinealectomized monkeys and showed that none of them developed scoliosis [109]. This raised the question of the appropriateness of using avian species and rodents as models for human idiopathic scoliosis. Furthermore, experiments that attempted to administer melatonin or transplant the pineal gland to inhibit the occurrence of scoliosis in chickens, were unsuccessful [105, 106]. This conflict of data extended to human studies, where researchers disagreed with Machida's group [110, 111]. One of the studies observed elevated serum melatonin levels in IS patients in the early stage of the disease when compared with matched controls [111]. Also to mimic the animal studies that involved pinealectomy in humans, the only way was to examine patients who have pineal gland dysfunction, like tumors or any other lesions [112]. In a study that included children with pineal lesions, only two out of 48 had scoliosis [113].

As a result, the hypothesis that melatonin deficiency produces scoliosis is not convincing. Yet, it is still accepted that melatonin does have a significant role in the pathogenesis of IS. Deciphering its involvement is still the focus of attention for scoliotic researchers.

Qiu et al studied the mRNA expression of melatonin receptors MTNR1A and MTNR1B in the paravertebral muscles of patients with IS, congenital scoliosis, and controls. They observed higher expression of MTNR1B mRNA on the concave side of the

paravertebral muscle than on the convex side in IS patients and no difference in expression of MTNR1A between both sides of the paravertebral muscle in IS patients [114]. They explained this asymmetry as a secondary effect. The expression of melatonin receptors in the osteoblasts of IS patients was also studied [115]. There was a reduced expression of MTNR1B receptors at the mRNA level and the protein level in those patients when compared with controls. This abnormal expression was associated with an abnormally long arm span.

The genetic regulation of melatonin synthesis and receptors in IS was also studied. Genetic association studies have been conducted on *TPH1* (Tryptophan-Hydoxylase-1) and *AANAT* (Arylalkylamine N-Acetyltransferase), the two main enzymes involved in the biosynthesis of melatonin. Only polymorphisms in *TPH1* were found to be associated with the incidence of IS [116]. Qiu et al studied the genetic association of the two melatonin receptors; *MTNR1A* and *MTNR1B*. They observed that a polymorphism in the promoter of *MTNR1B* is associated with IS [117] while no association was detected for *MTNR1A* with IS [118]. However, these results were contradicted, thereafter, by another meta-analysis study that collected the data of five different studies that contained a total of 2395 cases and 3645 controls and reanalyzed *MTNR1B* polymorphisms and their correlation with IS. They found no association in both the Asian and Caucasian populations [119].

Moreau et al suggested the possibility of the involvement of a defect in the melatonin signaling pathway in the pathogenesis of scoliosis [45]. This will be discussed in detail below in the section on Gi proteins. The hypothesis of the presence of an abnormal melatonin signaling pathway was confirmed by the work of Wang et al. This group demonstrated there was no effect of melatonin on the proliferation and the differentiation of

growth plate chondrocytes of IS patients as opposed to normal cells indicating a melatonin signaling impairment in IS derived chondrocytes [120]. This effect was suggested to be correlated with a defect in endochondral ossification in IS patients.

3.2. Systemic factors:

3.2.1. G inhibitory proteins:

G proteins are a family of guanine nucleotide-binding proteins that play a role in signal transduction of another big family of membrane receptors known as GPCRs (G protein-coupled receptors). They act as molecular switches that help transmit signals from external stimuli to cells. G proteins are classified as monomeric small GTPases and heterotrimeric G-proteins. The latter consists of three subunits; alpha (α), beta (β), and gamma (γ) [121]. The heterotrimeric G proteins are classified according to their subunits, Gαi/o (Gi), Gαs, Gαq, and Gα12/13. The G inhibitory (Gi) proteins have three different isoforms Gi1, Gi2, and Gi3. When an external stimulus interacts with one of the GPCRs, G proteins are activated and initiate a series of signaling events that ends with an altered cellular function. Cyclic adenosine mono phosphate (cAMP) is a second messenger involved in these signaling events. When Gi proteins are activated, the formation of cAMP is normally inhibited. Gi proteins inhibit adenylate cyclases (AC) that catalyzes the conversion of ATP (Adenosine Tri Phosphate) to cAMP, and thus inhibits the intracellular accumulation of cAMP [122].

We previously mentioned that Moreau et al suggested the role of a defect in the melatonin signaling pathway in the pathogenesis of IS. This group studied the effect of melatonin on osteoblasts isolated from IS patients. Owing to the fact that melatonin

receptors are GPCRs that are conjugated to Gi proteins, they should be able to naturally prevent the accumulation of cAMP. In this study, it was demonstrated that melatonin failed to inhibit the accumulation of cAMP in all the osteoblasts isolated from IS patients compared with controls [45]. It is hypothesized that there is a defect in the downstream melatonin signaling pathway that may involve posttranslational modifications of Gi proteins. In another study that focused on Gi proteins, Azeddine et al suggested an increased phosphorylation of serine residues in these proteins in IS cells and pointed to distinct protein-protein interactions that include those between MTNR1B melatonin receptors and PKC delta (protein kinase C delta) or the receptor for activated protein C kinase 1 and PKC delta [83]. In addition, this group was able to classify IS patients into three distinct groups based on their cellular response to melatonin [45, 83]. The same authors confirmed their findings by performing a cell-based screening test using the cellular dielectric spectroscopy technique that serves to directly measure the cellular response to Gi protein stimulation [123]. In their study, they confirmed their previous findings by obtaining the same patterns of reduced responses of IS cells (they used both osteoblasts and PBMCs) to both melatonin and its derivative iodomelatonin. This indicates that the Gi protein hypofunction is the key factor that causes the observed defect in the melatonin signal transduction pathway [84]. The evidence to classify IS patients into three functional groups according to their differential responses to melatonin, was also verified [84].

The same group further reported similar responses in different cell types isolated from IS patients, as observed with melatonin. They used osteoblasts, skeletal myoblasts, and PBMCs (peripheral blood mononuclear cells) from the same set of IS patients and controls. They concluded that the Gi protein hypofunction was not restricted to downstream

melatonin receptors, but in fact there was a generalized and systemic impairment of Gi protein-mediated receptor signaling. All IS cells showed Gi-coupled receptor hypofunction, but to three different extents, which confirmed their previous classification into three functional groups or biological endophenotypes (FG1, FG2, and FG3) [84, 124]. They further studied the cause of this differential hypofunctionality and found distinctive phosphorylation patterns occurring at serine residues of the Gi protein isoforms. Unlike the phenotypic variations that are observed in the measurement and site of the spinal curve in scoliotic patients, endophenotypes represent a conserved heritable trait. These endophenotypes can be used to stratify patients, and serve as clinical prognostic indicators, that can be associated with the molecular signatures of a condition [125]. It has been observed that severe cases of IS individuals account for 60% of the FG2 endophenotype in the French-Canadian population indicating an association with high risk of severe scoliosis in such individuals. In contrast, only 13% and 27% of IS individuals with the FG1 and FG3 endophenotypes, respectively, have severe scoliosis [125].

Owing to the fact that Gi proteins are spread all over the body in many different types of tissues, and act as an adaptor through which different stimuli, molecules, hormones, and peptides, perform their functions, Gi protein hypofunction represents a systemic defect that would explain the larger picture of metabolic abnormalities in IS rather than restricting them to a few molecules. For example, Alonso-Vale et al showed that cAMP was able to inhibit leptin synthesis [82]. Therefore, the diminution of leptin, which has been documented in IS may be correlated to the previously reported Gi signaling abnormality [45, 83, 84].

Additionally, the biggest challenge in studying IS is the complex nature of the disease and the high level of genotypic and phenotypic heterogeneity. The use of endophenotypes

improves the stratification of IS patients at a molecular level since molecular signatures may be utilized to predict the risk of development of the condition or any associated symptoms.

3.2.2. Osteopontin:

Osteopontin (OPN) is a multifunctional glycoprotein expressed by many different cell types and occurs throughout most of the biological systems in the body. OPN can exist either as a cytokine, which is expressed in all body fluids or as an extracellular matrix protein in mineralized tissues [126]. It has a wide variety of biological functions in both normal and pathological conditions [127-130].

Moreau's group previously reported high circulating plasma levels of OPN in patients with IS, which was associated with curve severity [131]. They further analyzed the role of OPN in scoliosis. They used C57BL/6 mice, a well-established model for studying scoliosis, which normally develop spinal deformities at a rate of approximately 60% following 40 weeks of bipedal ambulation, as mentioned previously [132]. They observed that the genetic depletion of OPN prevented spinal deformation in these mice [131] and enhanced the response to GiPCR stimulation in their osteoblasts [124].

The involvement of OPN in the pathogenesis of scoliosis was further supported by Yadav et al who showed that OPN plasma levels were increased, and the expression was enhanced in the vertebrae of Phospho1 KO (knockout) mice. Increased levels of OPN were associated with scoliosis and other skeletal abnormalities [133]. Furthermore, when OPN was genetically depleted from these mice, an improvement in their skeletal phenotype was observed. More evidence was provided by a recent study that studied OPN involvement in the susceptibility and progression of scoliosis [134]. They used a new mouse model

C3H/HeJ with elevated levels of OPN. The mice were divided into three groups: the first group was rendered bipedal and was treated with OPN, the second group was also rendered bipedal when treated with saline, and the third group was kept quadrupedal and treated with saline to serve as a control. 92.5% of the first group developed scoliosis with an average Cobb angle of 29.8°, 52.5% of the second group got scoliosis with an average cobb angle of 20.9° while the percentage of the development of scoliosis in the third group was 12.5 (average cobb angle 17.5°). Hence, we can conclude that there is consistent evidence for the potential involvement of OPN in the onset and progression of scoliosis.

3.3. Hematological factors (Platelet regulatory factors):

3.3.1. Calmodulin:

Calmodulin (CaM - calcium modulated), is a protein that binds to calcium and acts as a messenger that transmits calcium signals and hence regulates the function of different enzymes [135, 136]. It plays a role in the the contraction of muscles and platelets through calcium that regulates the interaction of actin and myosin. Since the early 1980s, a number of anomalies in platelets, calmodulin, and Ca transport in IS have been documented in the literature.

Elevated levels of intracellular phosphorous, and calcium in platelets together with a defective structure of myosin chains [137, 138], reduction in the ATPase activity in the cytosol suggesting a dysfunctional calcium transport system [137], reduction in the functionality of the contractile system inside the cell [139], and also reduction in the aggregation of platelets [140] in IS patients have been reported. Another report demonstrated aberrant platelet maturation in IS patients and found differential expression of

various Ca^{2+} ATPases in different groups of patients with different curve patterns [141]. These findings were challenged by two reports that demonstrated normal platelet behavior with respect to the different parameters that were previously studied [142, 143].

Platelet calmodulin levels were elevated in association with curve progression while calmodulin levels in stabilized moderate curves and controls were similar [50]. These results were confirmed by another longitudinal study that performed a follow-up on calmodulin levels over time and demonstrated that there were elevated calmodulin levels in curve progression cases while its levels did not change in stable cases [51]. Moreover, double curved and severe cases showed higher levels of calmodulin than moderate cases. These levels were reduced in most of the cases following brace treatment and surgical interventions, which suggested calmodulin as a potential biomarker for curve progression [51]. A Chinese group measured the protein expression levels of calmodulin in the paravertebral muscles of the curve apex of IS patients, and observed significantly lower expression on the convex side compared with either the concave side or normal muscles [144]. The asymmetric distribution of calmodulin in paravertebral muscles was also shown by another study. However, in contrast, when the protein levels on the convex side were higher than the concave side of the scoliotic curve, there was no difference in the concentration of calmodulin in the platelets compared with the controls [145].

During the past few years, many genetic studies were performed to detect a genetic association between *CALM1* (the gene encoding for calmodulin) and IS. A study that involved 30 IS Chinese patients and 30 matched controls, suggested a genetic correlation of a SNP in the promoter of *CALM1* and a predisposition to IS [14]. Another study that included 100 Chinese IS patients and 100 matched controls showed that different SNPs in

CALM1 may be correlated to different curve patterns; for example, *CALM1* rs12885713 was associated with double curves and lumbar curves, while rs5871 was associated with thoracic curves [146]. These findings were supported by another Chinese study involving 67 IS patients and 100 matched controls, where they confirmed the association of rs12885713 with double curve cases, and the association of rs5871with thoracic curve cases [63]. Recently, another Chinese study was performed on 146 IS patients and 146 controls, where they reported an association between 3 SNPs in *CALM1* (rs2300496, rs2300500, and rs3231718) and IS susceptibility [147]. None of the 12 studied SNPs was correlated with the severity of curve progression.

Akel et al analyzed the effect of calmodulin antagonists tamoxifen and trifluoperozine on the onset and progression of scoliosis in experimental animal models. In the first study that was performed on chickens that underwent pinealectomy, both the compounds were effective in preventing the deformity, tamoxifen was also effective in reducing the progression of the curve [49]. In the second study, they used C57BL/6 mice, where they confirmed the efficiency of calmodulin antagonists in preventing scoliotic deformities [48]. It should be mentioned here that tamoxifen has an estrogen agonistic effect too and its action may be due to that effect. The idea of testing calmodulin antagonists came from the fact that melatonin acts as a calmodulin antagonist.

As we previously mentioned, melatonin's involvement in IS is widely accepted since most of the pinealectomized experimental animal models that were bipedal develop scoliosis due to either fluctuations in the melatonin levels or a dysfunction in its signaling pathway. The interplay between melatonin and calmodulin is very complicated. It has been suggested that calmodulin is involved in regulating melatonin synthesis through adenylyl cyclases

[148]. Melatonin also affects the activity of calmodulin by inhibiting calmodulin kinase II [149]. Moreover, downstream melatonin PKCα (Protein Kinase Cα) is activated, which induces calmodulin phosphorylation and subsequent inhibition [149]. Both melatonin and calmodulin directly interact together and melatonin exerts most of its effects through this interaction [149, 150]. As mentioned above, calmodulin plays a role in regulating the contractile system in muscles and platelets. It was suggested that calmodulin may be involved in regulating the contractile proteins of the spinal cord too and thus affects the growth of the neural cells [151]. This may correlate with the hypothesis that there is an imbalance in the growth of the spinal cord and the vertebrae, which may lead to the deformity observed in scoliosis [151]. The role of melatonin as a free-radical scavenger is important in this hypothesis. If melatonin fails to combat the free radicals produced during stretching, damage to neural cells is probable and consequently growth of the spinal cord will be affected [37, 151].

In summary, calmodulin is a potential modifying factor in IS. Calmodulin is not only involved in a crosstalk with melatonin but also plays a regulatory role in the contractile system of muscles and consequently the paraspinal muscle tone as well as the spinal cord contractile system. This is in addition to the regulation of Ca^{2+} transport. It has been hypothesized that there may be a generalized defect in the calcium pump of scoliotic patients[151].

3.4. Bone metabolism factors:

3.4.1. Receptor activator of nuclear factor kappa-B ligand (RANKL):

Receptor activator of nuclear factor kappa-B ligand (RANKL) is also known as osteoprotegerin ligand (OPGL) or osteoclast differentiation factor (ODF). The protein

RANKL is a member of the tumor necrosis factor (TNF) super family. It is well known that RANKL plays an important role in controlling bone resorption and remodeling [152]. The RANKL/RANK interaction induces the formation of osteoclasts from their precursor cells, maintains their survival, and helps in the activation of osteoclasts. The OPG hinders bone resorption by binding to RANKL and blocks its binding to RANK. Hence, the RANKL/OPG ratio is an important measurement that reflects bone quality [152].

The presence of osteopenia and the reduction of bone mass in the axial and the peripheral skeleton of IS patients, have been consistently reported. Therefore, a number of researchers have studied the interrelationship between low bone mineral density (BMD) and its determinants in IS [153-155]. Suh et al [153] reported an elevation of serum RANKL and the RANKL/OPG ratio together with the reduction of BMD in 72 IS patients compared with 64 matched controls. This was confirmed by another study that demonstrated higher levels of RANKL in both pre- and post-menarcheal girls with IS when compared with corresponding control groups [156]. Another study involving 15 IS patients and 8 matched controls, reported a significant elevation in the levels of RANKL and an imbalance in the RANKL/OPG ratio, suggesting that this imbalance plays a role in bone remodeling [157]. In agreement with these findings, Zhou et al showed an elevation of both mRNA and protein levels of RANKL and an increased RANKL/OPG ratio in osteoblasts isolated from 20 IS patients with low BMD as compared with 8 age-matched controls [155]. Another study reported an association between a polymorphism in OPG gene and LSBMD (lumbar spine BMD) in IS girls [154]. Lu at al suggested that anti-osteoporotic treatment could effectively restore bone strength as well as the balance of the OPG/RANK/RANKL system [158].

3.4.2. Osteocalcin (OST):

Osteocalcin is a non-collagenous protein present in mineralized tissues. It is a hormone that is secreted by osteoblasts, odontoblasts, and hypertrophic chondrocytes [159]. Osteocalcin is an important biochemical marker for bone formation, in the case of uncoupled bone formation and resorption. In addition, osteocalcin is a marker for bone turnover in cases of high rates of bone resorption and formation [159]. It was observed that there was a significant elevation of osteocalcin levels in patients with IS compared with matched controls [156, 157]. This elevation reflects the high rate of bone remodeling and abnormal bone metabolism.

3.4.3. Matrilin 1:

Matrilin-1 or MATN1 is a protein known as cartilage matrix protein that is encoded by *MATN1*. The MATN1 protein is a member of a family of matrix proteins that forms filamentous structures in the matrices of cartilage and many other types of tissues [160]. Matrilin helps in the organization of the extracellular matrix (ECM). The ECM is an important structure that supports the spinal column and a defect in the ECM has been hypothesized as a possible cause in the pathogenesis of scoliosis [161]. A number of reports have studied the genetic association between IS and *MATN1*. However, the results of the reports are inconsistent. Montanaro et al studied 81 trios (parents and their child with IS) of IS and reported that microsatellite polymorphisms in *MATN1* were associated with familial IS [162]. Chen et al studied a Chinese cohort involving a total of 419 IS cases and 750 controls and reported a significant association between a polymorphism in the promoter region of *MATN1* and the incidence as well as progression of scoliosis [163]. In another

study that included a Korean cohort comprising 166 IS cases and 126 controls, a polymorphism in *MATN1* was found to be associated with a double major curve pattern[164].

On the other hand, other studies failed to reproduce the association of *MATN1* gene polymorphisms with IS susceptibility and/or curve progression [23, 165]. As a result of this discrepancy between different studies, a meta-analysis was conducted recently to assess the association of the rs1149048 SNP in *MATN1* with IS in different ethnic backgrounds [166]. This study included many studies involving a total of 1436 AIS patients and 1,879 controls. They found that the association of this SNP in *MATN1* was restricted to the Asian cohort and not the Caucasian cohort.

Furthermore, Wang et al reported a significant reduction in the plasma levels of matrilin-1 in IS patients compared to controls. This study involved 25 IS patients and 25 controls. The plasma matrilin 1 levels were significantly lower in severe curve progression cases compared with non-severe cases, suggesting matrilin-1 as a potential biomarker for IS[167].

Single nucleotide polymorphisms as well as reduced expression levels of matrilin-1 have been found to be associated with the progression of IS. Besides being a potential biomarker, matrilin-1 also appears to be a selective marker for IS based on an ethnic background.

3.4.4. Cartilage Oligomeric Matrix Protein (COMP):

Cartilage oligomeric matrix protein (COMP) is an extracellular matrix protein whose exact function is yet to be determined. It is present in the cartilage, ligament, tendon [168],

bone, synovium [169], intervertebral discs [170] and in the growth plate [171]. At the cellular level, COMP is found in chondrocytes and osteoblasts [171]. It has been suggested that COMP interacts with ECM1 (extracellular matrix protein 1), another protein in the extracellular matrix in the growth plate. This interaction may affect endochondral ossification [172]. The expression of *COMP* is elevated during the development of the skeleton and the growth of long bones [173]. It has been reported that mutations in *COMP* are associated with two skeletal disorders; pseudoachondroplasia (PSACH) and multiple epiphyseal dysplasia (MED) [174, 175].

One of the clinical features of PSACH is the development of scoliosis [175]. Defective COMP protein was found to accumulate inside chondrocytes and failed to be transported outside the cells, leading to the early death of these chondrocytes and interfering with the normal growth of the bones [176]. Serum COMP levels were observed to be correlated with the rate of growth in juvenile idiopathic arthritic patients [177]. In a microarray analysis that was performed on osteoblasts from IS patients, it was reported that *COMP* is one of the most downregulated genes in IS osteoblasts where the expression of *COMP* is 4-fold lower than that in the controls [58].

These findings were further supported by a recent study that involved 105 IS patients and 103 age-matched healthy controls [178]. The COMP serum levels were found to be significantly lower in IS cases versus controls. They also reported that COMP levels were associated with the speed of growth in IS patients. Hence, COMP is a potential biomarker that needs to be further investigated as of its usefulness in the diagnostic screening of scoliosis.

Conclusion

Identifying potential biomarkers is an essential exercise in improving the diagnostic efficiency of a condition. It has been a challenge to identify significant biomarkers for idiopathic scoliosis. Considerable research has been conducted in different aspects of cell physiology to detect and confirm these biomarkers. In this review, we consolidated the relevant studies for each molecule and provided a perspective on the potential role of these factors in diagnosing IS.

What is interesting is the crosstalk between some of the candidate biomarkers. There is a crosstalk between proteins, for example, between calmodulin and melatonin; Gi proteins and osteopontin, which provide multiple clues to detect IS. Given the dearth of confirmed biomarkers for IS, identifying a connection between some of the candidate factors is an affirmation of the direction in which IS research is headed.

Although IS is observed in first-degree relatives, the genetic aspect of the condition has not been clearly understood. In such a scenario, it has been useful to identify single nucleotide polymorphisms (SNPs) in calmodulin and matrilin-1, which influence the incidence of the spinal curve. Microsatellite polymorphisms in the matrilin-1 gene have been shown to be associated with familial IS while the SNP rs1149048 is associated with IS in the Asian population. Similarly, different SNPS in the calmodulin gene appear to regulate the development of curves in different sections of the spine (thoracic spine, lumbar spine).

The influence of hormones on the spinal curve progression has provided interesting candidates, one of which is melatonin. Pinealectomized animals develop scoliosis due to either a defect in the signaling pathway or variation in the concentration of melatonin. The

role of melatonin is diverse as it interacts and inhibits the action of calmodulin on platelets and muscle cells. Melatonin also acts on reactive oxygen species preventing their damaging action on neural cells. Downregulation of this hormone may cause neural cell damage and subsequently the potential development of IS.

Several diagnostic candidates are involved in bone metabolism. Osteocalcin, RANKL, COMP, and Matrilin-1, have been identified as potential diagnostic biomarkers of IS. While increased levels of osteopontin were associated with IS, reduced concentration of *MATN1* and *COMP* were implicated in IS.

It is clear that some studies are in their preliminary stages and additional studies are required to confirm the status of these candidates as diagnostic biomarkers. The data on growth hormones and estrogen have been inconclusive. While these hormones appear to be relevant due to their effects on growth at puberty, growth hormones and estrogen have to be validated and confirmed as potential diagnostic biomarkers with a number of large-scale clinical studies.

Identifying these factors is only half the job done. There are multiple facets to each factor that have to be considered. For example, calmodulin is a potential diagnostic marker for many reasons. It is up regulated in IS patients. Hence, elevated levels of calmodulin appear to be an indicator of spinal curve progression. There are single nucleotide polymorphisms within the calmodulin gene that appear to be correlated with different patterns of spinal curvature. And in addition, calmodulin regulates the synthesis of melatonin. In separate studies, melatonin has been identified as a potential diagnostic candidate of IS. Melatonin, in turn, interacts with the hormone leptin that is a diagnostic

candidate for IS. It is interesting to observe that the data of independent research studies have provided candidates that are interconnected with each other much like the pieces of a puzzle.

Acknowledgement:

References:

1. Cheng JC, Castelein RM, Chu WC, Danielsson AJ, Dobbs MB, Grivas TB, et al. PRIMER. NATURE. 2015.

2. Noshchenko A, Hoffecker L, Lindley EM, Burger EL, Cain CM, Patel VV, et al. Predictors of spine deformity progression in adolescent idiopathic scoliosis: A systematic review with meta-analysis. World journal of orthopedics. 2015;6(7):537.

3. Weinstein SL, Dolan LA, Cheng JC, Danielsson A, Morcuende JA. Adolescent idiopathic scoliosis. The Lancet. 2008;371(9623):1527-37.

4. Loncar-Dusek M, Pecina M, Prebeg Z. A longitudinal study of growth velocity and development of secondary gender characteristics versus onset of idiopathic scoliosis. Clin Orthop Relat Res. 1991(270):278-82.

5. Archer IA, Dickson RA. Stature and idiopathic scoliosis. A prospective study. J Bone Joint Surg Br. 1985;67(2):185-8.

6. Skogland LB, Steen H, Trygstad O. Spinal deformities in tall girls. Acta Orthop Scand. 1985;56(2):155-7.

7. Siu King Cheung C, Tak Keung Lee W, Kit Tse Y, Ping Tang S, Man Lee K, Guo X, et al. Abnormal peri-pubertal anthropometric measurements and growth pattern in adolescent idiopathic scoliosis: a study of 598 patients. Spine (Phila Pa 1976). 2003;28(18):2152-7.

8. Yim AP, Yeung HY, Hung VW, Lee KM, Lam TP, Ng BK, et al. Abnormal skeletal growth patterns in adolescent idiopathic scoliosis--a longitudinal study until skeletal maturity. Spine (Phila Pa 1976). 2012;37(18):E1148-54.

9. Willner S, Nilsson KO, Kastrup K, Bergstrand CG. Growth hormone and somatomedin A in girls with adolescent idiopathic scoliosis. Acta Paediatr Scand. 1976;65(5):547-52.

10. Dymling JF, Willner S. Progression of a structural scoliosis during treatment with growth hormone. A case report. Acta Orthop Scand. 1978;49(3):264-8.

11. de Lind van Wijngaarden RF, de Klerk LW, Festen DA, Duivenvoorden HJ, Otten BJ, Hokken-Koelega AC. Randomized controlled trial to investigate the effects of growth hormone treatment on scoliosis in children with Prader-Willi syndrome. J Clin Endocrinol Metab. 2009;94(4):1274-80.

12. Willner S, Johnell O. Study of biochemical and hormonal data in idiopathic scoliosis in girls. Arch Orthop Trauma Surg. 1981;98(4):251-5.

13. Ahl T, Albertsson-Wikland K, Kalen R. Twenty-four-hour growth hormone profiles in pubertal girls with idiopathic scoliosis. Spine (Phila Pa 1976). 1988;13(2):139-42.

14. Zhuang Q, Wu Z, Qiu G. Is polymorphism of CALM1 gene or growth hormone receptor gene associated with susceptibility to adolescent idiopathic scoliosis? Zhonghua yi xue za zhi. 2007;87(31):2198-202.

15. Qiu XS, Tang NL, Yeung HY, Qiu Y, Cheng JC. Genetic association study of growth hormone receptor and idiopathic scoliosis. Clin Orthop Relat Res. 2007;462:53-8.

16. Hesse V, Jahreis G, Schambach H, Vogel H, Vilser C, Seewald H, et al. Insulin-like growth factor I correlations to changes of the hormonal status in puberty and age. Experimental and clinical endocrinology. 1993;102(4):289-98.

17. Juul A, Dalgaard P, Blum WF, Bang P, Hall K, Michaelsen KF, et al. Serum levels of insulin-like growth factor (IGF)-binding protein-3 (IGFBP-3) in healthy infants, children, and adolescents: the relation to IGF-I, IGF-II, IGFBP-1, IGFBP-2, age, sex, body mass index, and pubertal maturation. J Clin Endocrinol Metab. 1995;80(8):2534-42.

18. Sanders JO, Browne RH, Cooney TE, Finegold DN, McConnell SJ, Margraf SA. Correlates of the peak height velocity in girls with idiopathic scoliosis. Spine. 2006;31(20):2289-95.

19. Sanders JO, Browne RH, McConnell SJ, Margraf SA, Cooney TE, Finegold DN. Maturity assessment and curve progression in girls with idiopathic scoliosis. J Bone Joint Surg Am. 2007;89(1):64-73.

20. Yeung HY, Tang NL, Lee KM, Ng BK, Hung VW, Kwok R, et al. Genetic association study of insulin-like growth factor-I (IGF-I) gene with curve severity and osteopenia in adolescent idiopathic scoliosis. Stud Health Technol Inform. 2006;123:18-24.

21. Moon ES, Kim HS, Sharma V, Park JO, Lee HM, Moon SH, et al. Analysis of single nucleotide polymorphism in adolescent idiopathic scoliosis in Korea: for personalized treatment. Yonsei Med J. 2013;54(2):500-9.

22. Yang Y, Wu Z, Zhao T, Wang H, Zhao D, Zhang J, et al. Adolescent idiopathic scoliosis and the single-nucleotide polymorphism of the growth hormone receptor and IGF-1 genes. Orthopedics. 2009;32(6):411.

23. Takahashi Y, Matsumoto M, Karasugi T, Watanabe K, Chiba K, Kawakami N, et al. Lack of association between adolescent idiopathic scoliosis and previously reported single nucleotide polymorphisms in MATN1, MTNR1B, TPH1, and IGF1 in a Japanese population. Journal of Orthopaedic Research. 2011;29(7):1055-8.

24. Liu W, Qiu Y, Chen Z, Liu Z, Shu H, Wang X, et al. Association between insulin-like growth factor-1 receptor gene polymorphisms and with susceptibility to adolescent idiopathic scoliosis. Zhonghua wai ke za zhi [Chinese journal of surgery]. 2009;47(23):1813-6.

25. Moldovan F, Letellier K, Azeddine F, Lacroix G, Wang D, Turgeon I, et al. The role of estrogens and estrogen receptors in the pathogenesis of adolescent idiopathic scoliosis (AIS). Journal of Bone & Joint Surgery, British Volume. 2008;90(SUPP III):431-.

26. Iwamuro S, Sakakibara M, Terao M, Ozawa A, Kurobe C, Shigeura T, et al. Teratogenic and anti-metamorphic effects of bisphenol A on embryonic and larval Xenopus laevis. Gen Comp Endocrinol. 2003;133(2):189-98.

27. Boudreau M, Courtenay SC, Maclatchy DL, Bérubé CH, Parrott JL, Van Der Kraak GJ. Utility of morphological abnormalities during early-life development of the estuarine mummichog, Fundulus heteroclitus, as an indicator of estrogenic and antiestrogenic endocrine disruption. Environmental toxicology and chemistry. 2004;23(2):415-25.

28. Grumbach MM. Estrogen, bone, growth and sex: a sea change in conventional wisdom. J Pediatr Endocrinol Metab. 2000;13 Suppl 6:1439-55.

29. Kulis A, Zarzycki D, Jaśkiewicz J. Concentration of estradiol in girls with idiophatic scoliosis. Ortopedia, traumatologia, rehabilitacja. 2006;8(4):455-9.

30. Esposito T, Uccello R, Caliendo R, Di Martino G, Carnevale UG, Cuomo S, et al. Estrogen receptor polymorphism, estrogen content and idiopathic scoliosis in human: a possible genetic linkage. The Journal of steroid biochemistry and molecular biology. 2009;116(1):56-60.

31. Warren MP, Brooks-Gunn J, Hamilton LH, Warren LF, Hamilton WG. Scoliosis and fractures in young ballet dancers. Relation to delayed menarche and secondary amenorrhea. N Engl J Med. 1986;314(21):1348-53.

32. Leboeuf D, Letellier K, Alos N, Edery P, Moldovan F. Do estrogens impact adolescent idiopathic scoliosis? Trends in Endocrinology & Metabolism. 2009;20(4):147-52.

33. Letellier K, Azeddine B, Parent S, Labelle H, Rompre PH, Moreau A, et al. Estrogen cross-talk with the melatonin signaling pathway in human osteoblasts derived from adolescent idiopathic scoliosis patients. J Pineal Res. 2008;45(4):383-93.

34. Raczkowski JW. The concentrations of testosterone and estradiol in girls with adolescent idiopathic scoliosis. Neuro endocrinology letters. 2007;28(3):302-4.

35. Slemenda CW, Reister TK, Hui SL, Miller JZ, Christian JC, Johnston CC, Jr. Influences on skeletal mineralization in children and adolescents: evidence for varying effects of sexual maturation and physical activity. J Pediatr. 1994;125(2):201-7.

36. Riggs BL, Khosla S, Melton LJ, 3rd. Sex steroids and the construction and conservation of the adult skeleton. Endocr Rev. 2002;23(3):279-302.

37. Lombardi G, Akoume MY, Colombini A, Moreau A, Banfi G. Biochemistry of adolescent idiopathic scoliosis. Adv Clin Chem. 2011;54:165-82.

38. Eastell R. Role of oestrogen in the regulation of bone turnover at the menarche. Journal of Endocrinology. 2005;185(2):223-34.

39. Manolagas SC, Jilka RL. Cytokines, hematopoiesis, osteoclastogenesis, and estrogens. Calcif Tissue Int. 1992;50(3):199-202.

40. Burner WL, 3rd, Badger VM, Sherman FC. Osteoporosis and acquired back deformities. J Pediatr Orthop. 1982;2(4):383-5.

41. Thomas KA, Cook SD, Skalley TC, Renshaw SV, Makuch RS, Gross M, et al. Lumbar spine and femoral neck bone mineral density in idiopathic scoliosis: a follow-up study. J Pediatr Orthop. 1992;12(2):235-40.

42. Hung VW, Qin L, Cheung CS, Lam TP, Ng BK, Tse YK, et al. Osteopenia: a new prognostic factor of curve progression in adolescent idiopathic scoliosis. J Bone Joint Surg Am. 2005;87(12):2709-16.

43. Cheng JC, Guo X, Sher AH. Persistent osteopenia in adolescent idiopathic scoliosis. A longitudinal follow up study. Spine (Phila Pa 1976). 1999;24(12):1218-22.

44. Lee WT, Cheung CS, Tse YK, Guo X, Qin L, Lam TP, et al. Association of osteopenia with curve severity in adolescent idiopathic scoliosis: a study of 919 girls. Osteoporos Int. 2005;16(12):1924-32.

45. Moreau A, Wang DS, Forget S, Azeddine B, Angeloni D, Fraschini F, et al. Melatonin signaling dysfunction in adolescent idiopathic scoliosis. Spine (Phila Pa 1976). 2004;29(16):1772-81.

46. Papaioannou S, Tumber AM, Meikle MC, McDonald F. G-protein signalling pathways and oestrogen: a role of balanced maintenance in osteoblasts. Biochim Biophys Acta. 1999;1449(3):284-92.

47. Demirkiran G, Dede O, Yalcin N, Akel I, Marcucio R, Acaroglu E. Selective estrogen receptor modulation prevents scoliotic curve progression: radiologic and histomorphometric study on a bipedal C57Bl6 mice model. Eur Spine J. 2014;23(2):455-62.

48. Akel I, Demirkıran G, Alanay A, Karahan S, Marcucio R, Acaroglu E. The effect of calmodulin antagonists on scoliosis: bipedal C57BL/6 mice model. European Spine Journal. 2009;18(4):499-505.

49. Akel I, Kocak O, Bozkurt G, Alanay A, Marcucio R, Acaroglu E. The effect of calmodulin antagonists on experimental scoliosis: a pinealectomized chicken model. Spine (Phila Pa 1976). 2009;34(6):533-8.

50. Kindsfater K, Lowe T, Lawellin D, Weinstein D, Akmakjian J. Levels of platelet calmodulin for the prediction of progression and severity of adolescent idiopathic scoliosis. J Bone Joint Surg Am. 1994;76(8):1186-92.

51. Lowe T, Lawellin D, Smith D, Price C, Haher T, Merola A, et al. Platelet calmodulin levels in adolescent idiopathic scoliosis: do the levels correlate with curve progression and severity? Spine. 2002;27(7):768-75.

52. Cutler GB, Jr. The role of estrogen in bone growth and maturation during childhood and adolescence. J Steroid Biochem Mol Biol. 1997;61(3-6):141-4.

53. Lee PA, Witchel SF. The influence of estrogen on growth. Current opinion in pediatrics. 1997;9(4):431-6.

54. Nilsson O, Chrysis D, Pajulo O, Boman A, Holst M, Rubinstein J, et al. Localization of estrogen receptors-alpha and-beta and androgen receptor in the human growth plate at different pubertal stages. Journal of Endocrinology. 2003;177(2):319-26.

55. Wang K, Shi D, Sun L, Jiang X, Lü Y, Dai J, et al. Association of estrogen receptor alpha gene polymorphisms with bone mineral density: a meta-analysis. Chinese medical journal. 2012;125(14):2589-97.

56. Grumbach MM, Auchus RJ. Estrogen: consequences and implications of human mutations in synthesis and action. J Clin Endocrinol Metab. 1999;84(12):4677-94.

57. Bilezikian JP, Morishima A, Bell J, Grumbach MM. Increased bone mass as a result of estrogen therapy in a man with aromatase deficiency. New England Journal of Medicine. 1998;339(9):599-603.

58. Fendri K, Patten S, Zaouter C, Parent S, Kaufman G, Labelle H, et al. Recent advances in the study of candidate genes for adolescent idiopathic scoliosis. Stud Health Technol Inform. 2010;158:3-7.

59. Inoue M, Minami S, Nakata Y, Kitahara H, Otsuka Y, Isobe K, et al. Association between estrogen receptor gene polymorphisms and curve severity of idiopathic scoliosis. Spine. 2002;27(21):2357-62.

60. Wu J, Qiu Y, Zhang L, Sun Q, Qiu X, He Y. Association of estrogen receptor gene polymorphisms with susceptibility to adolescent idiopathic scoliosis. Spine (Phila Pa 1976). 2006;31(10):1131-6.

61. Tang NL-S, Yeung H-Y, Lee K-M, Hung VW-Y, Cheung CS-K, Ng BK-W, et al. A relook into the association of the estrogen receptor α gene (PvuII, XbaI) and adolescent idiopathic scoliosis: a study of 540 Chinese cases. Spine. 2006;31(21):2463-8.

62. Zhang HQ, Lu SJ, Tang MX, Chen LQ, Liu SH, Guo CF, et al. Association of estrogen receptor beta gene polymorphisms with susceptibility to adolescent idiopathic scoliosis. Spine (Phila Pa 1976). 2009;34(8):760-4.

63. Zhao D, Qiu Gx, Wang Yp, Zhang Jg, Shen Jx, Wu Zh. Association between adolescent idiopathic scoliosis with double curve and polymorphisms of calmodulin1 gene/estrogen receptor-α gene. Orthopaedic surgery. 2009;1(3):222-30.

64. Takahashi Y, Matsumoto M, Karasugi T, Watanabe K, Chiba K, Kawakami N, et al. Replication study of the association between adolescent idiopathic scoliosis and two estrogen receptor genes. Journal of Orthopaedic Research. 2011;29(6):834-7.

65. Xu L, Qiu X, Sun X, Mao S, Liu Z, Qiao J, et al. Potential genetic markers predicting the outcome of brace treatment in patients with adolescent idiopathic scoliosis. Eur Spine J. 2011;20(10):1757-64.

66. Peng Y, Liang G, Pei Y, Ye W, Liang A, Su P. Genomic polymorphisms of G-protein estrogen receptor 1 are associated with severity of adolescent idiopathic scoliosis. Int Orthop. 2012;36(3):671-7.

67. Ogura Y, Takahashi Y, Kou I, Nakajima M, Kono K, Kawakami N, et al. A replication study for association of 5 single nucleotide polymorphisms with curve progression of adolescent idiopathic scoliosis in Japanese patients. Spine (Phila Pa 1976). 2013;38(7):571-5.

68. Janusz P, Kotwicki T, Andrusiewicz M, Kotwicka M. XbaI and PvuII polymorphisms of estrogen receptor 1 gene in females with idiopathic scoliosis: no association with occurrence or clinical form. PLoS One. 2013;8(10):e76806.

69. Yang M, Li C, Li M. The estrogen receptor alpha gene (XbaI, PvuII) polymorphisms and susceptibility to idiopathic scoliosis: a meta-analysis. J Orthop Sci. 2014;19(5):713-21.

70. Chen S, Zhao L, Roffey DM, Phan P, Wai EK. Association between the ESR1-351A> G single nucleotide polymorphism (rs9340799) and adolescent idiopathic scoliosis: a systematic review and meta-analysis. European Spine Journal. 2014;23(12):2586-93.

71. Kotwicki T, Janusz P, Andrusiewicz M, Chmielewska M, Kotwicka M. Estrogen receptor 2 gene polymorphism in idiopathic scoliosis. Spine (Phila Pa 1976). 2014;39(26):E1599-607.

72. Janusz P, Kotwicka M, Andrusiewicz M, Czaprowski D, Czubak J, Kotwicki T. Estrogen receptors genes polymorphisms and age at menarche in idiopathic scoliosis. BMC musculoskeletal disorders. 2014;15(1):383.

73. Liu Z, Tam EM, Sun GQ, Lam TP, Zhu ZZ, Sun X, et al. Abnormal leptin bioavailability in adolescent idiopathic scoliosis: an important new finding. Spine (Phila Pa 1976). 2012;37(7):599-604.

74. Tam E, Yu F, Hung V, Liu Z, Liu KL, Ng BK, et al. Are volumetric bone mineral density and bone micro-architecture associated with leptin and soluble leptin receptor levels in adolescent idiopathic scoliosis?--A case-control study. PloS one. 2014;9(2):e87939.

75. Qiu Y, Sun X, Qiu X, Li W, Zhu Z, Zhu F, et al. Decreased circulating leptin level and its association with body and bone mass in girls with adolescent idiopathic scoliosis. Spine (Phila Pa 1976). 2007;32(24):2703-10.

76. Elisa M, Lee W, Cheuk K-y, Lam T-P, Ng BK, Lee SK, et al. Morphological and bone strength indices in girls with adolescent idiopathic scoliosis and their correlations with leptin and soluble leptin receptor. Scoliosis. 2015;10(Suppl 1):O13.

77. Tam EM, Yeung K-H, Tang S, Lam T-P, Ng BK, Lee SK, et al. Abnormal functional responses of osteoblasts to leptin in adolescent idiopathic scoliosis. Scoliosis. 2015;10(Suppl 1):O6.

78. Liang G, Gao W, Liang A, Ye W, Peng Y, Zhang L, et al. Normal leptin expression, lower adipogenic ability, decreased leptin receptor and hyposensitivity to Leptin in Adolescent Idiopathic Scoliosis. PLoS One. 2012;7(5):e36648.

79. Burwell RG, Aujla RK, Grevitt MP, Dangerfield PH, Moulton A, Randell TL, et al. Pathogenesis of adolescent idiopathic scoliosis in girls-a double neuro-osseous theory involving disharmony between two nervous systems, somatic and autonomic expressed in the spine and trunk: possible dependency on sympathetic nervous system and hormones with implications for medical therapy. Scoliosis. 2009;4(1):24.

80. Canpolat S, Sandal S, Yilmaz B, Yasar A, Kutlu S, Baydas G, et al. Effects of pinealectomy and exogenous melatonin on serum leptin levels in male rat. Eur J Pharmacol. 2001;428(1):145-8.

81. Baltaci A, Mogulkoc R. Pinealectomy and melatonin administration in rats: their effects on plasma leptin levels and relationship with zinc. Acta Biologica Hungarica. 2007;58(4):335-43.

82. Alonso-Vale MI, Andreotti S, Peres SB, Anhe GF, das Neves Borges-Silva C, Neto JC, et al. Melatonin enhances leptin expression by rat adipocytes in the presence of insulin. Am J Physiol Endocrinol Metab. 2005;288(4):E805-12.

83. Azeddine B, Letellier K, Wang da S, Moldovan F, Moreau A. Molecular determinants of melatonin signaling dysfunction in adolescent idiopathic scoliosis. Clin Orthop Relat Res. 2007;462:45-52.

84. Akoume M-Y, Azeddine B, Turgeon I, Franco A, Labelle H, Poitras B, et al. Cell-based screening test for idiopathic scoliosis using cellular dielectric spectroscopy. Spine. 2010;35(13):E601-E8.

85. Sakata I, Sakai T. Ghrelin cells in the gastrointestinal tract. International journal of peptides. 2010;2010.

86. Maccarinelli G, Sibilia V, Torsello A, Raimondo F, Pitto M, Giustina A, et al. Ghrelin regulates proliferation and differentiation of osteoblastic cells. J Endocrinol. 2005;184(1):249-56.

87. Gennero I, Conte-Auriol F, Mus M, Molinas-Cazals C, Accadbled F, Tauber M, et al. Fasting total ghrelin levels are increased in patients with adolescent idiopathic scoliosis. Bone Abstracts 2013;2:122.

88. Brzezinski A. Melatonin in humans. N Engl J Med. 1997;336(3):186-95.

89. Grivas TB, Savvidou OD. Melatonin the" light of night" in human biology and adolescent idiopathic scoliosis. Scoliosis. 2007;2(6):1-14.

90. Morgan PJ, Barrett P, Howell HE, Helliwell R. Melatonin receptors: localization, molecular pharmacology and physiological significance. Neurochem Int. 1994;24(2):101-46.

91. Arendt J. Melatonin and the pineal gland: influence on mammalian seasonal and circadian physiology. Rev Reprod. 1998;3(1):13-22.

92. Reppert SM, Godson C, Mahle CD, Weaver DR, Slaugenhaupt SA, Gusella JF. Molecular characterization of a second melatonin receptor expressed in human retina

and brain: the Mel1b melatonin receptor. Proceedings of the National Academy of Sciences. 1995;92(19):8734-8.

93. Thillard MJ. Vertebral column deformities following epiphysectomy in the chick. C R Hebd Seances Acad Sci. 1959;248(8):1238-40.

94. Dubousset J, Queneau P, Thillard M. Experimental scoliosis induced by pineal and diencephalic lesions in young chickens: its relation with clinical findings. Orthop Trans. 1983;7(7):4.

95. Machida M, Dubousset J, Imamura Y, Iwaya T, Yamada T, Kimura J. An experimental study in chickens for the pathogenesis of idiopathic scoliosis. Spine. 1993;18(12):1609-15.

96. Machida M, Dubousset J, Imamura Y, Iwaya T, Yamada T, Kimura J. Role of melatonin deficiency in the development of scoliosis in pinealectomised chickens. J Bone Joint Surg Br. 1995;77(1):134-8.

97. Acaroglu E, Bobe R, Enouf J, Marcucio R, Moldovan F, Moreau A. The metabolic basis of adolescent idiopathic scoliosis: 2011 report of the "metabolic" workgroup of the Fondation Yves Cotrel. Eur Spine J. 2012;21(6):1033-42.

98. Machida M, Miyashita Y, Murai I, Dubousset J, Yamada T, Kimura J. Role of serotonin for scoliotic deformity in pinealectomized chicken. Spine. 1997;22(12):1297-301.

99. Machida M, Murai I, Miyashita Y, Dubousset J, Yamada T, Kimura J. Pathogenesis of idiopathic scoliosis. Experimental study in rats. Spine (Phila Pa 1976). 1999;24(19):1985-9.

100. Machida M, Dubousset J, Yamada T, Kimura J, Saito M, Shiraishi T, et al. Experimental scoliosis in melatonin-deficient C57BL/6J mice without pinealectomy. Journal of pineal research. 2006;41(1):1-7.

101. Machida M, Dubousset J, Yamada T, Kimura J. Serum melatonin levels in adolescent idiopathic scoliosis prediction and prevention for curve progression--a prospective study. J Pineal Res. 2009;46(3):344-8.

102. Bagnall K, Raso J, Moreau M, Mahood J, Wang X, Beuerlein M. The Development of Scoliosis Following Pinealectomy in Young Chickens is not the Result of an Artifact of the Surgical. Research Into Spinal Deformities 3. 2002;3:3.

103. Turgut M, Yenisey C, Uysal A, Bozkurt M, Yurtseven ME. The effects of pineal gland transplantation on the production of spinal deformity and serum melatonin level following pinealectomy in the chicken. Eur Spine J. 2003;12(5):487-94.

104. Turhan E, Acaroglu E, Bozkurt G, Alanay A, Yazici M, Surat A. Unilateral enucleation affects the laterality but not the incidence of scoliosis in pinealectomized chicken. Spine. 2006;31(2):133-8.

105. Bagnall K, Raso VJ, Moreau M, Mahood J, Wang X, Zhao J. The effects of melatonin therapy on the development of scoliosis after pinealectomy in the chicken. J Bone Joint Surg Am. 1999;81(2):191-9.

106. Bagnall KM, Beuerlein M, Johnson P, Wilson J, Raso VJ, Moreau M. Pineal transplantation after pinealectomy in young chickens has no effect on the development of scoliosis. Spine. 2001;26(9):1022-7.

107. Wang X, Moreau M, Raso VJ, Zhao J, Jiang H, Mahood J, et al. Changes in serum melatonin levels in response to pinealectomy in the chicken and its correlation with development of scoliosis. Spine (Phila Pa 1976). 1998;23(22):2377-81; discussion 82.

108. Dede O, Akel I, Demirkiran G, Yalcin N, Marcucio R, Acaroglu E. Is decreased bone mineral density associated with development of scoliosis? A bipedal osteopenic rat model. scoliosis. 2011;6(1):24.

109. Cheung KM, Wang T, Poon AM, Carl A, Tranmer B, Hu Y, et al. The effect of pinealectomy on scoliosis development in young nonhuman primates. Spine (Phila Pa 1976). 2005;30(18):2009-13.

110. Bagnall KM, Raso VJ, Hill DL, Moreau M, Mahood JK, Jiang H, et al. Melatonin levels in idiopathic scoliosis: diurnal and nocturnal serum melatonin levels in girls with adolescent idiopathic scoliosis. Spine. 1996;21(17):1974-8.

111. Goultidis TT, Papavasiliou KA, Petropoulos AS, Philippopoulos A, Kapetanos GA. Higher levels of melatonin in early stages of adolescent idiopathic scoliosis: toward a new scenario. J Pediatr Orthop. 2014;34(8):768-73.

112. Girardo M, Bettini N, Dema E, Cervellati S. The role of melatonin in the pathogenesis of adolescent idiopathic scoliosis (AIS). Eur Spine J. 2011;20 Suppl 1:S68-74.

113. Day GA, McPhee IB, Tuffley J, Tomlinson F, Chaseling R, Kellie S, et al. Idiopathic scoliosis and pineal lesions in Australian children. J Orthop Surg (Hong Kong). 2007;15(3):327-33.

114. Qiu Y, Wu L, Wang B, Yu Y, Zhu Z. Asymmetric expression of melatonin receptor mRNA in bilateral paravertebral muscles in adolescent idiopathic scoliosis. Spine. 2007;32(6):667-72.

115. Yim AP-y, Yeung H-y, Sun G, Lee K-m, Ng T-b, Lam T-p, et al. Abnormal skeletal growth in adolescent idiopathic scoliosis is associated with abnormal quantitative expression of melatonin receptor, mt2. International journal of molecular sciences. 2013;14(3):6345-58.

116. Wang H, Wu Z, Zhuang Q, Fei Q, Zhang J, Liu Y, et al. Association study of tryptophan hydroxylase 1 and arylalkylamine N-acetyltransferase polymorphisms with adolescent idiopathic scoliosis in Han Chinese. Spine. 2008;33(20):2199-203.

117. Qiu XS, Tang NL, Yeung HY, Lee K-M, Hung VW, Ng BK, et al. Melatonin receptor 1B (MTNR1B) gene polymorphism is associated with the occurrence of adolescent idiopathic scoliosis. Spine. 2007;32(16):1748-53.

118. Qiu X-S, Tang NL, Yeung H-Y, Cheng JC, Qiu Y. Lack of association between the promoter polymorphism of the MTNR1A gene and adolescent idiopathic scoliosis. Spine. 2008;33(20):2204-7.

119. Yang M, Wei X, Yang W, Li Y, Ni H, Zhao Y, et al. The polymorphisms of melatonin receptor 1B gene (MTNR1B) (rs4753426 and rs10830963) and susceptibility to adolescent idiopathic scoliosis: a meta-analysis. J Orthop Sci. 2015;20(4):593-600.

120. Wang WW, Man GC, Wong JH, Ng TB, Lee KM, Ng BK, et al. Abnormal response of the proliferation and differentiation of growth plate chondrocytes to melatonin in adolescent idiopathic scoliosis. Int J Mol Sci. 2014;15(9):17100-14.

121. Hamm HE, Gilchrist A. Heterotrimeric G proteins. Current opinion in cell biology. 1996;8(2):189-96.

122. Gilman AG. G proteins and dual control of adenylate cyclase. Cell. 1984;36(3):577-9.

123. Verdonk E, Johnson K, McGuinness R, Leung G, Chen YW, Tang HR, et al. Cellular dielectric spectroscopy: a label-free comprehensive platform for functional evaluation of endogenous receptors. Assay Drug Dev Technol. 2006;4(5):609-19.

124. Akoume M-Y, Franco A, Moreau A. Cell-based Assay Protocol for the Prognostic Prediction of Idiopathic Scoliosis Using Cellular Dielectric Spectroscopy. Journal of visualized experiments: JoVE. 2013(80).

125. Julien C, Gorman KF, Akoume M-Y, Moreau A. Towards a comprehensive diagnostic assay for scoliosis. Personalized Medicine. 2013;10(1):97-103.

126. Denhardt DT, Noda M, O'Regan AW, Pavlin D, Berman JS. Osteopontin as a means to cope with environmental insults: regulation of inflammation, tissue remodeling, and cell survival. Journal of Clinical Investigation. 2001;107(9):1055.

127. Weintraub AS, Schnapp LM, Lin X, Taubman MB. Osteopontin deficiency in rat vascular smooth muscle cells is associated with an inability to adhere to collagen and increased apoptosis. Laboratory investigation. 2000;80(11):1603-15.

128. Mazzali M, Kipari T, Ophascharoensuk V, Wesson J, Johnson R, Hughes J. Osteopontin—a molecule for all seasons. QJM: An International Journal of Medicine. 2002;95(1):3-13.

129. Standal T, Borset M, Sundan A. Role of osteopontin in adhesion, migration, cell survival and bone remodeling. Exp Oncol. 2004;26(3):179-84.

130. Terai K, Takano-Yamamoto T, Ohba Y, Hiura K, Sugimoto M, Sato M, et al. Role of osteopontin in bone remodeling caused by mechanical stress. Journal of Bone and Mineral Research. 1999;14(6):839-49.

131. Moreau A, Franco A, Azeddine B, Rompré PH, Gagnon M-HR, Bagnall KM, et al. High Circulating Levels of Osteopontin Are Associated with Idiopathic Scoliosis Onset and Spinal Deformity Progression 7. In: Aubin C-E, Stokes IAF, Labelle H, Moreau A, editors. Research into spinal deformities. Amsterdam, Netherlands: IOS Press BV; 2010.

132. Oyama J, Murai I, Kanazawa K, Machida M. Bipedal ambulation induces experimental scoliosis in C57BL/6J mice with reduced plasma and pineal melatonin levels. Journal of pineal research. 2006;40(3):219-24.

133. Yadav MC, Huesa C, Narisawa S, Hoylaerts MF, Moreau A, Farquharson C, et al. Ablation of osteopontin improves the skeletal phenotype of phospho1−/− mice. Journal of Bone and Mineral Research. 2014;29(11):2369-81.

134. Xie N, Li M, Wu T, Liu J, Wang B, Tang F. Does elevated osteopontin level play an important role in the development of scoliosis in bipedal mice? The Spine Journal. 2015.

135. Stevens FC. Calmodulin: an introduction. Canadian journal of biochemistry and cell biology. 1983;61(8):906-10.

136. Chin D, Means AR. Calmodulin: a prototypical calcium sensor. Trends in cell biology. 2000;10(8):322-8.

137. Yarom R, Muhlrad A, Hodges S, Robin G. Platelet pathology in patients with idiopathic scoliosis: Ultrastructural morphometry, agrregations, x-ray spectrometry, and biochemical analysis. Laboratory investigation; a journal of technical methods and pathology. 1980;43(3):208-16.

138. Peleg I, Eldor A, Kahane I, Muhlrad A, Liebergall M, Floman Y. Altered structural and functional properties of myosins, from platelets of idiopathic scoliosis patients. Journal of Orthopaedic Research. 1989;7(2):260-5.

139. Muhlrad A, Yarom R. Contractile protein studies on platelets from patients with idiopathic scoliosis. Pathophysiology of Haemostasis and Thrombosis. 1982;11(3):154-60.

140. Sabato S, Rotman A, Robin GC, Floman Y. Platelet aggregation abnormalities in idiopathic scoliosis. Journal of Pediatric Orthopaedics. 1985;5(5):558-63.

141. Bredoux R, Corvazier E, Dally S, Chaabane C, Bobe R, Raies A, et al. Human platelet Ca2+-ATPases: new markers of cell differentiation as illustrated in idiopathic scoliosis. Platelets. 2006;17(6):421-33.

142. KAHMANN RD, DONOHUE JM, BRADFORD DS, WHITE JG, RAO GH. Platelet function in adolescent idiopathic scoliosis. Spine. 1992;17(2):145-8.

143. Suk S, Kim I, Lee C, Koh Y, Yeom J. A study on platelet function in idiopathic scoliosis. Orthopedics. 1991;14(10):1079-83.

144. Zhao Y, Qiu GX. Expression of calmodulin and nNOS in the paraspinal muscles in idiopathic scoliosis. Zhonghua Yi Xue Za Zhi. 2004;84(16):1358-61.

145. Acaroglu E, Akel I, Alanay A, Yazici M, Marcucio R. Comparison of the melatonin and calmodulin in paravertebral muscle and platelets of patients with or without adolescent idiopathic scoliosis. Spine. 2009;34(18):E659-E63.

146. Zhao D, Qiu GX, Wang YP, Zhang JG, Shen JX, Wu ZH, et al. Association of calmodulin1 gene polymorphisms with susceptibility to adolescent idiopathic scoliosis. Orthop Surg. 2009;1(1):58-65.

147. Zhang Y, Gu Z, Qiu G. The association study of calmodulin 1 gene polymorphisms with susceptibility to adolescent idiopathic scoliosis. Biomed Res Int. 2014;2014:168106.

148. Xia Z, Storm DR. Calmodulin-regulated adenylyl cyclases and neuromodulation. Curr Opin Neurobiol. 1997;7(3):391-6.

149. Hardeland R. Melatonin, hormone of darkness and more–occurrence, control mechanisms, actions and bioactive metabolites. Cellular and Molecular Life Sciences. 2008;65(13):2001-18.

150. Benítez-King G. Melatonin as a cytoskeletal modulator: implications for cell physiology and disease. Journal of pineal research. 2006;40(1):1-9.

151. Porter RW. The pathogenesis of idiopathic scoliosis: uncoupled neuro-osseous growth? European Spine Journal. 2001;10(6):473-81.

152. Boyce BF, Xing L. Biology of RANK, RANKL, and osteoprotegerin. Arthritis Research and Therapy. 2007;9(1):S1.

153. Suh KT, Lee S-S, Hwang SH, Kim S-J, Lee JS. Elevated soluble receptor activator of nuclear factor-κB ligand and reduced bone mineral density in patients with adolescent idiopathic scoliosis. European Spine Journal. 2007;16(10):1563-9.

154. Eun I-S, Park WW, Suh KT, Kim JI, Lee JS. Association between osteoprotegerin gene polymorphism and bone mineral density in patients with adolescent idiopathic scoliosis. European Spine Journal. 2009;18(12):1936-40.

155. Zhou S, Wang W, Zhu Z, Sun X, Zhu F, Yu Y, et al. Increased expression of receptor activator of nuclear factor-κB ligand in osteoblasts from adolescent idiopathic scoliosis patients with low bone mineral density. Journal of Huazhong University of Science and Technology [Medical Sciences]. 2012;32:686-90.

156. Kulis A, Gozdzialska A, Drag J, Jaskiewicz J, Knapik-Czajka M, Lipik E, et al. Participation of sex hormones in multifactorial pathogenesis of adolescent idiopathic scoliosis. Int Orthop. 2015;39(6):1227-36.

157. Chiru M. Adolescent idiopathic scoliosis and osteopenia. Maedica (Buchar). 2011;6(1):17-22.

158. Lu L, Dai Z, Lv G, Kang Y, Jiang Y. A novel therapeutic strategy for adolescent idiopathic scoliosis based on osteoporotic concept. Med Hypotheses. 2013;80(6):773-5.

159. Ram VS, Parthiban, Sudhakar U, Mithradas N, Prabhakar R. Bonebiomarkers in periodontal disease: a review article. J Clin Diagn Res. 2015;9(1):ZE07-10.

160. Wagener R, Ehlen HW, Ko YP, Kobbe B, Mann HH, Sengle G, et al. The matrilins--adaptor proteins in the extracellular matrix. FEBS Lett. 2005;579(15):3323-9.

161. Lowe TG, Edgar M, Margulies JY, Miller NH, Raso VJ, Reinker KA, et al. Etiology of idiopathic scoliosis: Current trends in research*. The Journal of Bone & Joint Surgery. 2000;82(8):1157-.

162. Montanaro L, Parisini P, Greggi T, Di Silvestre M, Campoccia D, Rizzi S, et al. Evidence of a linkage between matrilin-1 gene (MATN1) and idiopathic scoliosis. Scoliosis. 2006;1(1):21.

163. Chen Z, Tang NL, Cao X, Qiao D, Yi L, Cheng JC, et al. Promoter polymorphism of matrilin-1 gene predisposes to adolescent idiopathic scoliosis in a Chinese population. Eur J Hum Genet. 2009;17(4):525-32.

164. Bae JW, Cho C-H, Min W-K, Kim U-K. Associations between matrilin-1 gene polymorphisms and adolescent idiopathic scoliosis curve patterns in a Korean population. Molecular biology reports. 2012;39(5):5561-7.

165. Yilmaz H, Zateri C, Uludag A, Bakar C, Kosar S, Ozdemir O. Single-nucleotide polymorphism in Turkish patients with adolescent idiopathic scoliosis: Curve progression is not related with MATN-1, LCT C/T-13910, and VDR BsmI. Journal of Orthopaedic Research. 2012;30(9):1459-63.

166. Zhang H, Zhao S, Zhao Z, Tang L, Guo Q, Liu S, et al. The association of rs1149048 polymorphism in matrilin-1(MATN1) gene with adolescent idiopathic scoliosis susceptibility: a meta-analysis. Mol Biol Rep. 2014;41(4):2543-9.

167. Wang B, Chen ZJ, Qiu Y, Liu WJ. Decreased circulating matrilin-1 levels in adolescent idiopathic scoliosis. Zhonghua Wai Ke Za Zhi. 2009;47(21):1638-41.

168. Posey KL, Hecht JT. The role of cartilage oligomeric matrix protein (COMP) in skeletal disease. Curr Drug Targets. 2008;9(10):869-77.

169. Xu K, Zhang Y, Ilalov K, Carlson CS, Feng JQ, Di Cesare PE, et al. Cartilage oligomeric matrix protein associates with granulin-epithelin precursor (GEP) and potentiates GEP-stimulated chondrocyte proliferation. J Biol Chem. 2007;282(15):11347-55.

170. Ishii Y, Thomas AO, Guo XE, Hung CT, Chen FH. Localization and distribution of cartilage oligomeric matrix protein in the rat intervertebral disc. Spine (Phila Pa 1976). 2006;31(14):1539-46.

171. Di Cesare PE, Fang C, Leslie MP, Tulli H, Perris R, Carlson CS. Expression of cartilage oligomeric matrix protein (COMP) by embryonic and adult osteoblasts. J Orthop Res. 2000;18(5):713-20.

172. Kong L, Tian Q, Guo F, Mucignat MT, Perris R, Sercu S, et al. Interaction between cartilage oligomeric matrix protein and extracellular matrix protein 1 mediates endochondral bone growth. Matrix Biology. 2010;29(4):276-86.

173. Fendri K, Moldovan F. Potential role of COMP as a biomarker for adolescent idiopathic scoliosis. Medical hypotheses. 2011;76(5):762-3.

174. Hecht JT, Nelson LD, Crowder E, Wang Y, Elder FF, Harrison WR, et al. Mutations in exon 17B of cartilage oligomeric matrix protein (COMP) cause pseudoachondroplasia. Nat Genet. 1995;10(3):325-9.

175. Briggs M, Hoffman S, King L, Olsen A, Mohrenweiser H, Leroy J, et al. Pseudoachondroplasia and multiple epiphyseal dysplasia due to mutations in the cartilage oligomeric matrix protein gene. Nature genetics. 1995;10(3):330-6.

176. Hecht JT, Makitie O, Hayes E, Haynes R, Susic M, Montufar-Solis D, et al. Chondrocyte cell death and intracellular distribution of COMP and type IX collagen in the pseudoachondroplasia growth plate. J Orthop Res. 2004;22(4):759-67.

177. Bjornhart B, Juul A, Nielsen S, Zak M, Svenningsen P, Muller K. Cartilage oligomeric matrix protein in patients with juvenile idiopathic arthritis: relation to growth and disease activity. J Rheumatol. 2009;36(8):1749-54.

178. Gerdhem P, Topalis C, Grauers A, Stubendorff J, Ohlin A, Karlsson K. Serum level of cartilage oligomeric matrix protein is lower in children with idiopathic scoliosis than in non-scoliotic controls. European Spine Journal. 2015;24(2):256-61.