Université de Montréal

Identification of Genomic Variants Associated with Adolescent Idiopathic Scoliosis (AIS) in French-Canadian Population

par:
Qi Lin TANG

Département de Biochimie et Médecine Moléculaire
Faculté de Médecine

Mémoire présenté à la Faculté des études supérieures et postdoctorales
en vue de l'obtention du grade de Maîtrise en Sciences (M.Sc)
en Biochimie
option Génomique Humaine

December 2014

Université de Montréal
Faculté des études supérieures et postdoctorales

Ce mémoire intitulé :

Identification of Genomic Variants Associated with Adolescent Idiopathic Scoliosis (AIS) in French-Canadian Population

présenté par :
Qi Lin TANG

a été évalué par un jury composé des personnes suivantes :

Muriel Aubry, président-rapporteur
René St-Arnaud, membre du jury
Alain Moreau, directeur de recherche

Résumé

La scoliose idiopathique est une déformation tridimensionnelle de la colonne vertébrale dont la pathogenèse reste obscure. Cette maladie affecte 2-4% des adolescents de 10-18 ans parmi les garçons et les filles. Il est à noter que les filles sont plus sévèrement affectées et ce en plus grand nombre que les garçons. Les études de jumeaux ont montré que les facteurs génétiques jouent un rôle important dans la scoliose idiopathique de l'adolescent (SIA).

Depuis 2010, les études d'association pan génomiques ont été multipliées dans les recherches, visant à trouver des gènes candidats impliqués dans la SIA à travers des examens des polymorphismes nucléotidiques (SNPs). Un test génétique nommé "ScoliScore" a été publié pour essayer de prédire la progression de courbure dans la population caucasienne. Cependant, l'association n'a pas été reproduite dans une grande étude japonaise, soulignant l'importance d'une étude de réplication dans une population caucasienne indépendante.

Dans ce contexte, mon projet de maîtrise a permis de génotyper plus de 1,4 millions de SNPs dans une cohorte canadienne-française dans le but: **1)** de valider l'association de ScoliScore^{TM}; et **2**) d'identifier les variants génomiques associées à la SIA dans la population québécoise.

Notre étude a montré qu'aucun des variants constituant le test ScoliScore^{TM} n'était associé à la SIA. Ceci suggère que l'absence d'association dans une cohorte japonaise n'est pas due à l'appartenance ethnique. Aussi, nous avons identifié des variants génomiques associés significativement à l'initiation et/ou la progression de SIA dans la population québécoise, suggérant des gènes candidats impliqués dans la pathogenèse de SIA.

Mots-Clés

scoliose idiopathique de l'adolescent, polymorphisme d'un seul nucléotide, variant génomique, étude d'association pan génomique, ScoliScore^{TM}, progression de la courbe de colonne vertébrale, population caucasienne, canadienne-française, analyse de l'association, génotypage

Abstract

Idiopathic scoliosis is a common spinal deformation occurring without clear reason. This disease affects 2-4% adolescents aging from 10-18 years old in both genders. Of note, girls are more affected in number and severity than boys. Twin studies demonstrated that genetic factors play an important role in adolescent idiopathic scoliosis (AIS).

Since 2010, Genome-wide association studies (GWAS) have been multiplied in AIS researches, aiming to find out candidate genes involved in the disease by an examination of single nucleotide polymorphisms (SNPs) throughout the entire genome. A genetic test named "ScoliScore" was released for the prediction of curvature progression in Caucasian AIS population using 53 SNPs. However, such association was not replicated in a larger Japanese-population study. Such a discrepancy could be explained by ethnicity, raising the importance of a replication study in an independent Caucasian population of European descent.

In that context, we genotyped over 1.4 million SNPs in a French-Canadian cohort: **1)** to validate the association in ScoliScore^{TM} test; and **2)** to identify genomic variants associated with AIS in the population of Quebec.

As a result, the association of ScoliScore^{TM} genomic markers could not be reproduced in French-Canadian AIS patients, suggesting that the lack of association of these SNPs in a Japanese cohort is not due to ethnicity. Meanwhile, we identified genome-wide significant variants associated with spinal curve initiation and/or progression in French-Canadian population, suggesting candidate genes involved in AIS pathogenesis.

Keywords

adolescent idiopathic scoliosis, single nucleotide polymorphism, genomic variant, genome-wide association study, ScoliScore^{TM}, spinal curve progression, Caucasian, French-Canadian, association analysis, genotype

Table of Contents

Résumé i
Abstract ii
Table of Contents iii
List of Figures v
List of Tables vi
List of Abbreviations vii
Acknowledgements x
CHAPTER 1. REVIEW OF LITERATURE 1
1.1 Introduction of idiopathic scoliosis 2
1.1.1 Sub-groups of IS by age at disease onset 2
1.1.2 Prevalence of IS in adolescents 3
1.2 Scoliosis detection and screening 4
1.3 Scoliosis management 6
1.3.1 Observation 6
1.3.2 Bracing 6
1.3.3 Surgery 7
1.4 Etiopathogenesis of scoliosis 9
1.4.1 Genetic theory 9
1.4.1.1 Linkage studies in pedigrees 10
1.4.1.2 Genome-wide association studies 11
1.4.1.3 Whole exome sequencing 13
1.4.2 Neurological theory 14
1.4.3 Muscular theory 14
1.4.4 Connective tissue theory 15
1.4.5 Bone growth mismatch theory 15
1.4.6 Endocrine abnormality theory 16
1.5 Hypothesis and objectives 18
1.5.1 Hypothesis 18
1.5.2 Objectives 18
CHAPTER 2. ARTICLE I 19
A Replication Study for Association of 53 Single Nucleotide Polymorphisms in ScoliScore™ Test with Adolescent Idiopathic Scoliosis in French-Canadian Population 20
Authors' contribution 20
Structured abstract 23
Mini abstract 24
Introduction 25
Materials and methods 26
Results 29

Discussion 30
Acknowledgements 33
Tables and figures 34
References 44
CHAPTER 3. ARTICLE II 46
A Genome-Wide Association Study of Adolescent Idiopathic Scoliosis in French-Canadian Population 47
Authors' contribution 47
Structured abstract 50
Mini abstract 51
Introduction 52
Materials and methods 53
Results 55
Discussion 56
Acknowledgements 59
Tables and figures 60
References 63
CHAPTER 4. DISCUSSION 65
4.1 Future work in GWAS approach 68
4.2 Other genetic hypotheses and relative approaches in AIS study 69
4.2.1 Common disease-rare variant hypothesis and whole exome sequencing 69
4.2.2 Gene expression studies on epigenetic modifications 71
4.2.3 Functional group classification among AIS patients 72
CHAPTER 5. CONCLUSION 73
CHAPTER 6. REFERENCES 75

List of Figures

Figure 1. The Adam's forward bending test by scoliometer. 5
Figure 2. The Cobb method to quantify spinal curve severity. 5
Figure 3. A corset worn in brace treatment for scoliosis. 8
Figure 4. Spinal fusion surgery for severe scoliosis case. 8
Figure 5. Manhattan plot showing the P values from genome-wide association study. 60

List of Tables

Table I. Demographic and clinical characteristics of severe patients and non-severe patients with AIS. 34
Table II. Twenty-five SNPs included both in ScoliScore™ and in Illumina genotyping microarray. 35
Table III. Twenty-seven SNPs in ScoliScore™ and their proxy SNPs in Illumina genotyping microarray. 36
Table IV. Association of 25 ScoliScore™ SNPs with AIS in French-Canadian population. .. 37
Table V. Association of 27 Proxy SNPs with AIS in French-Canadian population..... 38
Table VI. Association of 25 ScoliScore™ SNPs with AIS progression in French-Canadian population. 39
Table VII. Association of 27 Proxy SNPs with AIS progression in French-Canadian population. 40
Table VIII. Association of 25 ScoliScoreTM SNPs with AIS progression in French-Canadian population. 41
Table IX. Association of 27 Proxy SNPs with AIS progression in French-Canadian population. 42
Table X. Statistical power calculations for each association study in R software. 43
Table XI. Demographic and clinical characteristics of severe patients and non-severe patients with AIS at the last visit. 61
Table XII. Statistical power calculations for each association study in R software. 61
Table XIII. Significant SNPs and candidate genes identified by GWAS approach in French-Canadian population. 62

List of Abbreviations

AIS: Adolescent Idiopathic Scoliosis

ALGGEN: Algorithmics and Genetics Group

CaM: Calmodulin

CDCV: Common Disease-Common Variant

CDRV: Common Disease-Rare Variant

CELF2: CUGBP- and Elav-like family member 2

ChIP-seq: chromatin immunoprecipitation sequencing

CHL1: cell adhesion molecule with homology to L1CAM

CI: Confidence Interval

CNTNAP2: contactin associated protein-like 2

DSCAM: Down syndrome cell adhesion molecule

ENCODE: Encyclopedia of DNA Elements

FBN1: fibrillin 1

FBN2: fibrillin 2

FDA: U.S Food and Drug Administration

GCFC2: GC-rich sequence DNA-binding factor 2

GPR126: G protein-coupled receptor 126

GWAS: Genome-Wide Association Study

IBD: Identity-By-Descent

IS: Idiopathic Scoliosis

KCNJ2: potassium inwardly-rectifying channel, subfamily J, member 2

KLC4: kinesin light chain 4

LBX1: ladybird homeobox 1

LD: Linkage Disequilibrium

LRRTM4: leucine rich repeat transmembrane neuronal 4

MAF: Minor Allele Frequency

miRNA: micro ribonucleic acid

MRI: Magnetic Resonance Imaging

mRNA: messenger ribonucleic acid

MT2: melatonin receptor 1B or MTNR1B

OMIM: Online Mendelian Inheritance in Man

OR: Odds Ratio

PBMC: Peripheral Blood Mononuclear Cell

QC: Quality Control

qPCR: quantitative Polymerase Chain Reaction

SNP: Single Nucleotide Polymorphism

SOX9: sex determining region Y-box 9

SRS: Scoliosis Research Society

TGF-β: transforming growth factor beta

YY1: Yin and yang 1

To the ones who travelled with me on this journey

Acknowledgements

Firstly, I would like to express my sincere gratitude towards Dr. Alain Moreau, my research director, for giving me this opportunity to join in his wonderful lab, for his patience and encouragement during my master study. He is a wonderful role model and has provided me with many exciting opportunities over the last two years. I am deeply obliged.

Secondly, I would like to thank Dr. Cedric Julien, the research supervisor, for his guidance and support on my project.

I also wish to appreciate Dr. Kristen Fay Gorman for her ideas and advices which are very valuable for me to improve the project.

Special thanks to all of my colleagues for their help and encouragement during my study at university of Montreal: Anita Franco, DaShen Wang, Ginette Lacroix, Saadallah Bouhanik, Jean-François Lavoie, Niaz Olizadeh, Lakshmi Suvranan, Dina Nada, Nancy Karam, Mohamed Elbakry.

Finally, I cannot express enough my gratitude towards my parents and husband for their constant encouragement and support.

CHAPTER 1. REVIEW OF LITERATURE

1.1 Introduction of idiopathic scoliosis

Scoliosis was first documented in 400 B.C. by Hippocrates in Greece, characterized by a lateral spinal curvature, usually accompanied by vertebral rotation. It is a three dimensional spinal deformation in the frontal (lateral curvature), sagittal (thoracic lordosis) and transversal plane (vertebral rotation). The Scoliosis Research Society (SRS) has defined scoliosis as a lateral curvature of the spine exceeding 10 degrees as measured using the Cobb method on a standing radiograph *[Kane 1977]*.

There are four categories of scoliosis: 1) **idiopathic scoliosis** (IS) is the most common type of scoliosis. It occurs in 80% of scoliosis patients without clear reason; 2) **congenital scoliosis** is a rare type of scoliosis. It is often due to abnormal formation of the bones of the spine; 3) **neuromuscular scoliosis** is a lateral curvature of the spine due to loss of control of the nerves or muscles that support the spine; 4) **degenerative scoliosis** occurs in adults and is due to degeneration of the spine that occurs with aging.

1.1.1 Sub-groups of IS by age at disease onset

Idiopathic scoliosis (IS, OMIM# 181800) can be observed at any age. Traditionally, it is categorized by patient's age when the scoliosis is first identified. **Infantile idiopathic scoliosis** is defined by the age at disease onset as younger than 3 years and accounts for fewer than 1% of all IS cases in the United States. **Juvenile idiopathic scoliosis** is defined as scoliosis detected between ages 3 and 10. **Adolescent idiopathic scoliosis** is detected between the age of 10 years and skeletal maturity. Idiopathic scoliosis is more common in juveniles and adolescents when children are growing rapidly. Juvenile represents 12-21% of patients with IS, whereas adolescent makes up approximately 80% of all IS cases *[Dobbs and Weinstein 1999; James 1954; Riseborough and Wynne-Davies 1973]* .

In my project, we focus on scoliosis presenting in adolescents without clear undergoing cause, termed as adolescent idiopathic scoliosis (AIS), because AIS constitutes the majority of the IS cases.

1.1.2 Prevalence of IS in adolescents

The scoliosis affects 2% to 4% of adolescents in the world with unknown reason. Of adolescents diagnosed with scoliosis, only 10% have curve progression requiring medical intervention. The ratio of girls to boys with small curves around 10 degrees is equal. But the ratio increases, among cases with curves greater than 30 degrees, to an impressive 10 to 1. Scoliosis in girls tends to progress more frequently. Therefore, girls need treatment more commonly than boys do. Patient gender is one of the main factors to be taken under consideration in the estimation of curve progression risk by clinicians. Besides this, patient's age and the curve magnitude at the time of diagnosis need to be taken into account while estimating the risk. Younger patients having greater growth potential are at high risk of curve progression. The larger the initial curve, the greater the likelihood of curve progression *[Miller 1999; Roach 1999]*.

Several studies supported that AIS clusters in families. There is a higher incidence of AIS within the families of affected patients than in the general population. First-degree relatives of the affected individuals are at the highest risk and third-degree relatives are at the lowest risk *[Riseborough and Wynne-Davies 1973; Ward, Ogilvie, Argyle et al. 2010; Wynne-Davies 1968]*.

1.2 Scoliosis detection and screening

Currently, there is no diagnostic tool to predict the occurrence of idiopathic scoliosis among asymptomatic adolescents. Patients' family members are usually the first to notice the physical symptoms indicating scoliosis, such as one shoulder higher than the other or uneven leg lengths.

School-based scoliosis screening is recommended as a valuable tool to identify suspected cases which are sent for diagnostic confirmation. This screening allows the identification of scoliosis at an earlier stage. Given the statements of the SRS International Task Force on Scoliosis screening, supported by the SRS Board of Directors, females should be screened twice, at age 10 and 12, and boys once, at age 13 or 14.

At present, the scoliometer is a good tool in terms of reliability and validity to identify suspect individuals with spinal deformity in scoliosis screening. It is small and non-invasive that is placed over the spine while the person being measured is in a forward bending position (Adam's forward bend test, **Figure 1**). The scoliometer is a good indicator for trunk asymmetry, but should not be used as a diagnostic tool. The scolimeter measurement may underestimate the actual curve. An adolescent with positive screening results may be referred for a spinal x-ray. If so, the Cobb angle of the spinal curve(s) would be reported *[Cote, Kreitz, Cassidy et al. 1998; Kotwicki, Chowanska, Kinel et al. 2013]*.

The Cobb angle was first described in 1948 by Dr. John Robert Cobb (1903-1967), where he outlined how to measure the angle of the spinal curve. The Cobb angle measurement is used as the standard measurement to quantify and track the progression of scoliosis (**Figure 2**). Today, it is the "gold standard" of scoliosis evaluation endorsed by the Scoliosis Research Society (SRS). The Cobb angle degree is also an important parameter in our study for quantification of the severity of scoliosis deformation.

Figure 1. The Adam's forward bending test by scoliometer.
Clinicians identify the suspected adolescents with scoliosis by this screening, in which the individual bends from the waist as if touching the toes.
Figure adapted from http://www.posturetek.com/en/scoliometer.html

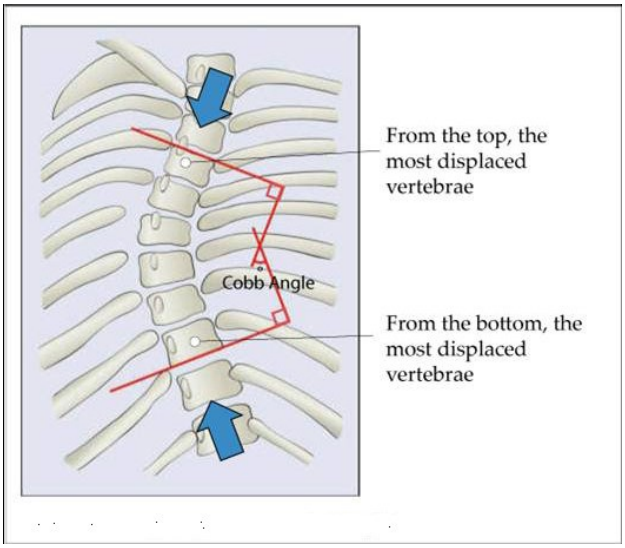

Figure 2. The Cobb method to quantify spinal curve severity.
Step 1. Identify the upper and lower end vertebrae
Step 2. Draw lines extending along the vertebral borders
Step 3. Measure the Cobb angle directly or geometrically
Figure adapted from e-radiography.net and core concepts

1.3 Scoliosis management

In clinics, scoliosis is defined when Cobb angle is greater than 10 degrees. At this time, there is no cure program for scoliosis but treatment options for scoliosis patients based on their severity of the curves, including observation, bracing, and surgical treatment *[Kotwicki, Chowanska, Kinel et al. 2013]*.

1.3.1 Observation

Patients with a spinal curve less than 25 degrees take routine x-ray testing periodically to observe the tendency of curve progression. In x-ray exams, two radiologic pictures are usually taken in a standing position, one from the back (postero-anterior or PA view) and one from the side (lateral view). The scoliosis patients might be asked to repeat the radiologic testing at regular intervals, sometimes every 3-12 months, to monitor the curve progression. If the curve remains below 25 degrees, no treatment is needed.

Although the amount of radiation used in an x-ray testing is small to minimize radiation hazards, adolescents in the growth stage are more vulnerable to radioactive harm. Thus, greater care is recommended in deciding which adolescents need further x-ray tests in their future.

1.3.2 Bracing

If the curve is between 25 and 45 degrees and the patients are still growing, adolescents need to wear a corset until their growth finish (**Figure 3**). More recently, Weinstein *et al*. reported a significant improvement on treatment success rate after bracing (72%) compared to the rate after observation (48%) among high-risk patients given references for bracing treatment. They revealed a positive association of average hours of daily brace wear with the treatment's success rate *[Weinstein, Dolan, Wright et al. 2013]*.

It is important to note that bracing does not correct scoliotic curvature, but may help slow or halt the spinal curve from getting worse until skeletal maturity. Patients reaching

skeletal maturity are unlikely to benefit from the use of a brace *[Weinstein, Dolan, Wright et al. 2013]*.

1.3.3 Surgery

Once the curve is greater than 45 degrees, it will probably continue getting worse for the rest of patients' life. It leads very much likely to lung or heart problems. As the last resort, a spinal fusion surgery is called, in which bone grafts combined with metal screws and rods are used to prevent further curvature in specific parts of the spine (**Figure 4**). In most cases, there is no need to remove the metal screws and rods from the spine. The goal of fusion surgery is to correct and stabilize the spinal curve.

The treatment cost for scoliosis varies by region in the world. Typically, it costs **$1,000** or more per year for observation, including periodic x-rays and doctor visits, about **$2,000-$6,000** for initial bracing, and about **$100,000-$150,000** or more for surgery. For example, according to a study of hospital charges to more than 76,000 patients, the average cost to the patient for scoliosis surgery was about **$113,000** *[Daffner, Beimesch and Wang 2010]*. Such expensive costs of scoliosis treatment raise the importance of developing a genetic test in the prediction of curve progression. Effective diagnostic/prognostic tool would help the AIS patients to be treated as soon as possible, notably with new fusionless devices and minimally invasive surgical approaches.

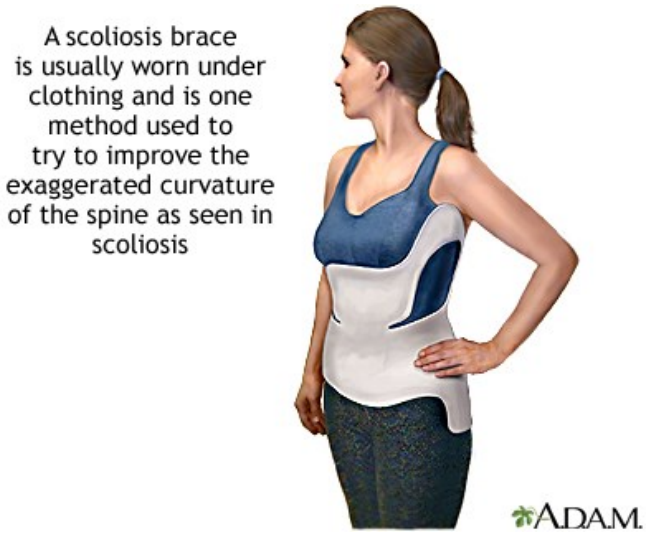

Figure 3. A corset worn in brace treatment for scoliosis.
The goal of brace treatment is to prevent the spinal curve from getting worse.
Figure adapted from http://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0002221/

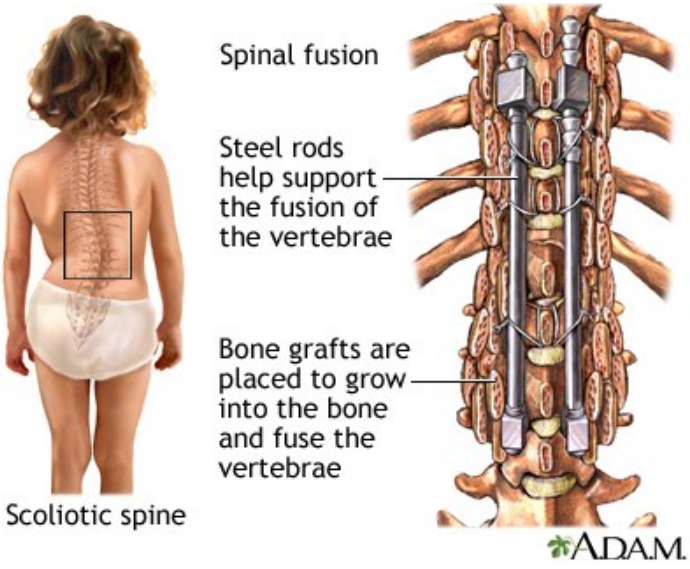

Figure 4. Spinal fusion surgery for severe scoliosis case.
The goal of fusion surgery is to correct and stabilize the spinal curve.
Figure adapted from http://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0002221/

1.4 Etiopathogenesis of scoliosis

Despite considerable advances made in the scoliosis management in the past decades, the etiopathogenesis of AIS has not been clarified. Etiologic hypotheses and concepts of AIS etiopathogenesis have been proposed, most including genetic theory, neurological theory, muscular theory, connective tissue theory, bone growth mismatch theory and endocrine abnormality theory *[Burwell and Dangerfield 2012; Burwell, Dangerfield, Moulton et al. 2011; Dayer, Haumont, Belaieff et al. 2013; Kouwenhoven and Castelein 2008; Wang, Yeung, Chu et al. 2011; Yagi, Machida and Asazuma 2014]*.

1.4.1 Genetic theory

AIS is sometimes abounded in certain families with multiple members affected, suggesting that AIS is inherited within families and that relatives of AIS patients have a greater risk than general populations *[Wynne-Davies 1968]*.

More evidence of a genetic contribution to AIS was revealed by studies in twins. Monozygotic twins have identical genetic information while dizygotic twins share half of their genetic information. A concordance rate is defined as the proportion of a certain condition's occurrence in both twins among total twin pairs that at least one of the twins has the condition. If the genetic contribution exists, this rate in monozygotic twins will be significantly different to that in dizygotic twins. By a meta-analysis of studies in twins, Kesling and Reinker reported a concordance rate of AIS at 73% in 37 pairs of monozygotic twins and at 36% in 31 pairs of dizygotic twins *[Kesling and Reinker 1997]*.

Recently, using the Danish Twin Registry, one of the most comprehensive registers of twins in the world, Andersen *et al*. reported concordance rates for AIS in 110 sets of twins, in which one or both of the twins were considered to have AIS. In their findings, 6 out of total 44 monozygotic pairs were affected by AIS in both twins. They did not find one pair that was both affected among 91 dizygotic twins. The concordance rates were 13% and zero for monozygotic and dizygotic twins, respectively *[Andersen, Thomsen and Kyvik 2007]*.

Both twin studies showed statistically significant concordance rates in monozygotic twins and in dizygotic twins, supporting the evidence of genetic contribution to AIS.

Nevertheless, within families of 207 AIS patients, Riseborough and Wynne-Davies reported the disease risk at 11%, 2.4% and 1.4% in first-, second- and third-degree relatives, respectively, suggesting a multifactorial mode of AIS inheritance which is distinct from single gene disease *[Riseborough and Wynne-Davies 1973]*. In addition, a heritability study of 69 extended Utah families with a history of AIS indicated that this disease is a polygenic and multifactorial condition, demonstrating the genetic and phenotypic complexity for AIS *[Ward, Ogilvie, Argyle et al. 2010]*.

Focusing on different hypotheses of genetic contribution to human complex diseases like AIS, researchers performed various approaches in genetic studies, including family-based linkage studies, population-based association studies and whole exome sequencing studies *[Gorman, Julien, Oliazadeh et al. 2014]*.

1.4.1.1 Linkage studies in pedigrees

Distinct observation of AIS aggregation within families suggested heritability of the disease, leading linkage studies in multiplex families. Linkage study is a statistical approach in a hypothesis-driven fashion, in which polymorphic markers are tested for linkage with disease. This approach has been successful in the discovery of Mendelian disease genes. But the majority has failed to identify causative genes for complex disease, such as AIS. The failure could possibly come from genetic and phenotypic heterogeneity *[Dawn Teare and Barrett 2005]*.

In AIS research field, candidate genes from clinical observation were examined in early linkage studies. The findings were limited by the studied sample size in pedigrees and uncertain gene functions at that time. Since 2000, through non-biased whole-genome linkage studies, several loci have been reported significant under different modes of inheritance: 3q12.1, 5q13.3, 9q31.2-34.2, 12p, 17p11, 19p13.3, Xq22.3-27.2, 6q15-q21, 10q23-q25.3 and 19p13.3, supporting that AIS is genetically heterogeneous and multifactorial disease *[Gorman, Julien and Moreau 2012]*.

1.4.1.2 Genome-wide association studies

Genome-wide association study (GWAS) is an effective and non-hypothesis based approach to discover risk variants associated with a trait through a large-scale genomic screening. It is an approach designed to identify common genetic variants with minor effect. Currently, only two genome-wide association studies have been documented in AIS field, one in the Caucasian population, the other in the Japanese population.

In 2011, the first GWAS study was conducted in Caucasian population. AIS-associated variants were first identified in 419 trio-families in Utah. Two most significant variants in the same gene were replicated in other three independent cohorts. Their findings demonstrated the most significant SNP (rs10510181, p-value=8.22×10^{-7}, odds ratio: OR=1.37, 95% confidence interval: CI=1.20-1.58) in the gene CHL1 (cell adhesion molecule with homology to L1CAM), suggesting the involvement of the axon guidance pathway in AIS susceptibility in the Caucasian population *[Sharma, Gao, Londono et al. 2011]*. Furthermore, they suggested another two genes, DSCAM (Down syndrome cell adhesion molecule) and CNTNAP2 (contactin associated protein-like 2), as candidate genes in AIS pathogenesis, which are involved in the axon guidance pathway. However, there was no statistical association between the polymorphisms and AIS susceptibility in Chinese populations *[Qiu, Lv, Zhu et al. 2014; Zhou, Zhu, Qiu et al. 2012]*.

The other GWAS study was conducted in a Japanese female population composed of 1033 AIS-affected patients and 1473 healthy individuals. Based on the genotype data from their GWAS and then combined with a replication study in a total of 11000 Japanese female cohort, they reported that three risk variants located near the gene LBX1 (ladybird homeobox 1) were significantly associated with AIS susceptibility *[Takahashi, Kou, Takahashi et al. 2011]*. The most significant association in Japanese population (rs11190870, combined p-value=1.24×10^{-19}, OR=1.56, 95% CI=1.41-1.71) was successfully replicated in three independent Chinese populations, suggesting that the abnormal somatosensory function was implicated in the etiology of spinal deformity in East Asia population *[Fan, Song, Chan et al. 2012; Gao, Peng, Liang et al. 2013; Jiang, Qiu, Dai et al. 2013; Liang, Xing, Li et al. 2014]*.

Likewise, another significant genetic association with AIS was identified through the above GWAS, and then followed by three replication studies using Japanese, Chinese and

Europe-ancestry populations (rs6570507, combined p-value=1.27×10^{-14}, OR=1.27, 95% CI=1.20-1.35). This time, the variant locates in the intron region of the GPR126 gene (G protein-coupled receptor 126), which is involved in the growth and ossification of developing spine and in neurological development. The variant reached sufficient significance level in East Asia and Europe-ancestry populations, suggesting the involvement of the GPR126 gene in AIS occurrence *[Kou, Takahashi, Johnson et al. 2013]*.

With a definition of severe curvature if the Cobb angle was above 40°, genotype data from the above GWAS in Japanese females was used to find risk variants associated with severe curves compared with control subjects. The association of rs12946942, located between two genes (SOX9 and KCNJ2), was identified with severe curves in females and followed by replication studies in Japanese and Chinese populations (combined p value=6.43×10^{-12}, OR=2.21, 95% CI=1.76-2.77). Although the variant rs12946942 was located in a region without clear effect yet, their findings suggested closest genes SOX9 (sex determining region Y-box 9) and KCNJ2 (potassium inwardly-rectifying channel, subfamily J, member 2) as promising candidate genes that played a role in AIS onset and/or progression in Japanese and Chinese female patients *[Miyake, Kou, Takahashi et al. 2013]*.

There were other independent GWA studies in AIS field presented in seminars and conferences, suggesting chromosome 3p25.3, 9p21.1, 10q24.3 and 12q12 as AIS susceptibility loci *[Dormans, Grant, Sampson et al. 2011; Nelson, Chettier, Ogilvie et al. 2011]*.

From another unpublished GWAS, 53 SNPs have been reported associated with AIS curve progression among Caucasian female patients in the United States. These genotyping data gave birth to an AIS progression prognostic tool *[Ward, Ogilvie, Singleton et al. 2010]*. Incorporating patients' initial Cobb angle measured between 9 and 13 years, this tool was built to quantify risk of spinal curve progression for Caucasian patients with a Cobb angle <25º, which was then commercialized under the name of ScoliScore™. Although not yet approved by the FDA (the U.S Food and Drug Administration), ScoliScore™ is the only DNA-based test developed to identify patients with mild AIS in Caucasian population who have a low risk of spinal curve progression. However, for some academic and/or commercial reasons, the authors did not describe enough details in their study design, leading to hesitation and consideration about the scientific foundation of ScoliScore™ *[Dobbs and Gurnett 2011; Grant and Dormans 2011]*. In addition, a recent study in an independent 85 Caucasian AIS

patients failed to replicate any genetic association between the 53 SNPs of ScoliScore™ and spinal curve progression *[Roye, Wright, Williams et al. 2012]*. Another study in a Japanese population did not yield any result supporting the presumed genetic associations by ScoliScore™ *[Ogura, Takahashi, Kou et al. 2013]*.

1.4.1.3 Whole exome sequencing

Most of the associated variants found in GWAS were located in non-protein-coding region with unexplained biological function.

Lately, the first study of rare variants was published in AIS field. Buchan *et al.* reported rare variants (defined as absent from the dbSNP database build 137) in the genes FBN1 (fibrillin 1) and FBN2 (fibrillin 2) that were concentrated in AIS patients with severe curve. Identified in an exome sequencing screen among 91 severe AIS cases (Cobb angle ≥40° or surgically treated) and 337 controls, frequency of rare variants in FBN1 among severe cases was significantly different from that among controls (p-value=3.17×10^{-4}, OR=10.4, 95% CI=2.7-39.5). Meanwhile the related gene FBN2 demonstrated a weak association to severe AIS (p-value=0.04). Verified in a larger cohort of European ancestry (323 severe cases versus 493 controls), the frequency of FBN1 and FBN2 rare variants in severe AIS was over 3 times the frequencies in two independent control cohorts (7.6% versus 2.4% and 2.3%, respectively). Moreover, FBN1 and FBN2 rare variants were not significantly associated with non-severe AIS cases compared to control cohorts (p-value=0.47 and 0.42, respectively). Similar results were observed in a replication study using 370 Chinese AIS patients (p-value=0.048) *[Buchan, Alvarado, Haller et al. 2014]*.

Of course, one of the limitations of this study is the fact that it remains to be proven that these variants have a pathological contribution by measuring changes in the expression of genes located in the vicinity of these variants. Furthermore, functional analysis in animal models will be required to further understand their contribution. We expect that in the next few years there will be more studies of rare variants (unknown or/and with low frequency) that can shed light on our understanding of AIS pathogenesis.

1.4.2 Neurological theory

The nervous system has been studied to explore potential factors playing a role in the etiopathogenesis of AIS. Children with AIS demonstrated abnormalities in electroencephalographic activity, postural balance, vestibular, somatosensory function equilibrium *[Beaulieu, Toulotte, Gatto et al. 2009; Cheng, Guo, Sher et al. 1999; Guo, Chau, Hui-Chan et al. 2006; Petersen, Sahlstrand and Sellden 1979; Sahlstrand and Petruson 1979; Sahlstrand, Petruson and Ortengren 1979; Simoneau, Richer, Mercier et al. 2006]*. Regional brain volume differences, examined via magnetic resonance imaging (MRI), were revealed among children with AIS when compared with age-matched healthy control individuals *[Liu, Chu, Young et al. 2008]*. Evidence in other MRI studies also revealed an uncoupled growth between the skeleton and the neural system in AIS cases. Mismatch of bone growth and spinal cord growth could induce stretching-tethering forces on the spine which result in spinal deformation with the continuing growth of the vertebral bodies *[Chu, Lam, Chan et al. 2006; Chu, Man, Lam et al. 2008; Porter 2001a; 2000; 2001b]*, proposing the asynchronous spinal neuro-osseous growth theory for AIS etiopathogenesis.

1.4.3 Muscular theory

The paraspinal muscles have been suggested as a possible causative factor in AIS etiology. Several electromyographic studies showed an increased activity of the paraspinal muscles on the convex side of the spine *[Alexander and Season 1978; Cheung, Halbertsma, Veldhuizen et al. 2005; Zetterberg, Bjork, Ortengren et al. 1984]*. However, interpretations of the electromyographic findings are quite different. It remains an argument whether the increased muscle activity is a causative factor to initiate the spinal curve initiation or a secondary consequence due to the curvature of spine.

1.4.4 Connective tissue theory

Because scoliosis is sometimes associated with connective tissue diseases, such as osteogenesis imperfecta and Marfan's syndrome *[Sponseller, Hobbs, Riley et al. 1995]*, connective tissues could also have implicated in the AIS pathogenesis. Hadley-Miller *et al*. reported that a high proportion (82%) of AIS patients exhibited disarrangement of elastic fibers in the ligamentum flavum. Moreover, 23% of AIS patients showed a marked decrease in fiber density. Seventeen percent of patients demonstrated a defect of fibrillin in the metabolism of its incorporation into the extracellular, suggesting the potential role of the elastic fiber system as a component in the pathogenesis of some AIS patients *[Hadley-Miller, Mims and Milewicz 1994]*. However, this could be secondary to the physical changes associated with the spinal deformity.

Most recently, through an exome sequencing study, a burden of rare variants in fibrillin genes, FBN1 (fibrillin 1) and FBN2 (fibrillin 2), was found in severely affected AIS cases *[Buchan, Alvarado, Haller et al. 2014]*. Previous studies have demonstrated that mutations in FBN1 are associated with Marfan's syndrome *[Dietz, Loeys, Carta et al. 2005; Kainulainen, Karttunen, Puhakka et al. 1994]*. Mutations in FBN2 are associated with Beals syndrome, a rare congenital connective tissue disorder *[Gupta, Putnam, Carmical et al. 2002; Putnam, Zhang, Ramirez et al. 1995]*. Although further studies are needed to prove the pathological contribution of these variants, this study suggests the role of fibrillin-related genes involved in AIS etiopathogenesis.

1.4.5 Bone growth mismatch theory

Idiopathic scoliosis occurs more often in adolescents when their skeletons are growing rapidly, proposing abnormal spinal growth as a contributing factor in the etiology of idiopathic scoliosis. A simple model of the spine shaped a scoliosis as a result of overgrowth of the anterior spine relative to the posterior spinal growth. The greater the overgrowth, the more pronounced the deformity *[Murray and Bulstrode 1996]*. However, the cause of this imbalance of the anterior and posterior structures of the spine has not been reported yet.

Factors inducing skeletal growth mismatch could play a role in the initiation and progression of a scoliosis.

1.4.6 Endocrine abnormality theory

Several endocrine abnormalities, such as calmodulin and melatonin, have been described associated with AIS disease.

Calmodulin (CaM), a calcium receptor protein modulating intracellular calcium activity, regulates the contractile properties of skeletal muscle and platelets through its interaction with actin-myosin system. Increased CaM levels over time in platelet have been shown in association with the curve progression of AIS patients. But these levels usually decreased in patients undergoing curve stabilization by bracing or spinal fusion *[Kindsfater, Lowe, Lawellin et al. 1994; Lowe, Lawellin, Smith et al. 2002]*. However, there was no establishment of the normal range for platelet CaM because of a large inexplicable discrepancy between baseline levels of different patients, necessitating the use of the AIS subjects as their own controls. Dr. Lowe considered the platelet as a "mini" skeletal muscle with a similar actin-myosin contractile system, suggesting the muscle hypothesis in the AIS etiology *[Lowe, Burwell and Dangerfield 2004]*. Furthermore, elevated CaM is a feature of activated platelets, which release growth factors as well, such as transforming growth factor beta (TGF-β), suggesting a skeletal hypothesis *[Geoffrey Burwell and Dangerfield 2003]*.

Idiopathic scoliosis-like changes were induced by experimental pinealectomy in chickens and bipedal rats, but not in quadrupedal rats, suggesting the importance of melatonin in the bipedal animal models *[Machida, Murai, Miyashita et al. 1999; Thillard 1959]*. Melatonin, also known as N-acetyl-5-methoxytryptamine, is a hormone secreted from the pineal gland. There were lower blood melatonin concentrations in pinealectomized chickens with scoliosis. Furthermore, melatonin administration may prevent the progression of scoliosis in the pinealectomized chickens model and in AIS patients *[Machida, Dubousset, Imamura et al. 1995; Machida, Dubousset, Yamada et al. 2009]*. However, there are no significant differences in circulating melatonin levels among AIS patients and healthy controls,

suggesting the role of other components in the melatonin signaling pathway *[Girardo, Bettini, Dema et al. 2011]*.

Dr. Moreau demonstrated several years ago the occurrence of a melatonin signaling impairment in AIS patients using their osteoblasts and PBMCs (peripheral blood mononuclear cells) *[Akoume, Azeddine, Turgeon et al. 2010; Moreau, Wang, Forget et al. 2004]*. Recently, a significantly lower expression of MT2 (or MTNR1B, melatonin receptor 1B) was found in AIS patients and was also correlated with abnormal systemic skeletal growth *[Yim, Yeung, Sun et al. 2013]*. Although the mechanism of melatonin signaling pathway in skeletal bone growth is not completely understood, the findings mentioned above suggest the important role of melatonin and its receptors and signaling pathway to the etiopathogenesis of AIS.

1.5 Hypothesis and objectives

1.5.1 Hypothesis

The contribution of genetic factors to the pathogenesis of Adolescent Idiopathic Scoliosis (AIS) has been revealed by twin studies. The identification of genetic variants associated with the susceptibility or severity of spinal curvature would facilitate the development of diagnostic/prognostic tools. The population in Quebec is unique because it is more isolated than the rest of North America and the incidence of AIS is higher than average here, leading to a valuable founder population with low genetic variability to medical genetic research. Thus, there is strong potential to identify variants aggregated in this population due to founder effects.

We assumed that AIS is a consequence of a moderate to large number of common genetic variants, each of which contributes to several percent of the risk for curvature and/or progression. For complex common diseases with an apparent polygenic inheritance, the common disease-common variant hypothesis (CDCV hypothesis) has motivated the pursuit of genome-wide association studies (GWAS). The goal of GWAS is to identify the causative variants that are underlying genomic markers associated with a disease, and then to characterize their functional effects.

1.5.2 Objectives

There have been a number of loci identified through genome-wide association studies in many populations. Here with a French-Canadian cohort, we performed a GWAS approach to: **1)** verify the AIS-associated genetic loci previously identified by ScoliScore^{TM} research team through GWAS; **2)** identify and validate the genetic variants associated with the development or/and the progression of adolescent idiopathic scoliosis, in order to determine their values in clinical practice and in further etiopathogenesis research.

CHAPTER 2. ARTICLE I

A Replication Study for Association of 53 Single Nucleotide Polymorphisms in ScoliScore^{TM} Test with Adolescent Idiopathic Scoliosis in French-Canadian Population

Authors' contribution

Qi Lin Tang performed association analyses and drafted the initial manuscript.

Alain Moreau is the research director of the study. He conceptualized and supervised study design. He revised the manuscript and approved the final manuscript.

Cedric Julien supervised and designed the study. He provided support and guidance in association analyses.

Robert Eveleigh and **Guillaume Bourque** participated in quality control measurement of genomic data.

Kristen Fay Gorman critically revised the manuscript.

Anita Franco contributed to sample preparation for genotyping and revised the manuscript.

Hubert Labelle, Benoit Poitras, Guy Grimard, Stephan Parent, Jean Ouellet, and Jean-Marc Mac-Thiong participated in the study population screening and recorded patients' medical conditions.

All authors read and approved the final manuscript.

The manuscript has been submitted to Spine journal in July 2014 and is considered acceptable with minor revisions.

A Replication Study for Association of 53 Single Nucleotide Polymorphisms in ScoliScore™ Test with Adolescent Idiopathic Scoliosis in French-Canadian Population

Qi Lin Tang[1,2] BSc, Cedric Julien[1] PhD, Robert Eveleigh[3] BSc, Guillaume Bourque[3] PhD, Anita Franco[1] MSc, Hubert Labelle[4,5] MD, Benoit Poitras[5] MD, Guy Grimard[5] MD, Stephan Parent[5] MD, Jean Ouellet[6] MD, Jean-Marc Mac-Thiong[5] MD, Kristen F. Gorman[1,2] PhD, and Alain Moreau[1,2,7] PhD

[1]Viscogliosi Laboratory in Molecular Genetics of Musculoskeletal Diseases, Sainte-Justine University Hospital Research Center, Montréal, Qc, Canada;
[2]Department of Biochemistry, Faculty of Medicine, Université de Montréal, Montréal, Qc, Canada;
[3]Genome Quebec Innovation Center, McGill University, Montréal, Qc, Canada;
[4]LIS3D Laboratory, Sainte-Justine University Hospital Research Center, Montréal, Qc, Canada;
[5]Orthopedic Division, Sainte-Justine University Hospital and Department of Surgery, Faculty of Medicine, Université de Montréal, Montréal, Qc, Canada;
[6]Orthopedic Division, The Montreal Children's Hospital, Department of Surgery, McGill University, Montréal, Qc, Canada;
[7]Department of Stomatology, Faculty of Dentistry, Université de Montréal, Montréal, Qc, Canada.

*Corresponding author: Prof. Alain Moreau, Sainte-Justine University Hospital Research Center, Director of Viscogliosi Laboratory in Molecular Genetics of Musculoskeletal Diseases, 3175 Cote-Sainte-Catherine Road, Montreal, Quebec, H3T 1C5, Canada. Phone: 514-345-4931; Fax: 514-345-4801

This work was supported by grants from La Fondation Yves Cotrel de l'Institut de France, Paris, France (to Dr. Moreau), and Génome Québec (to Dr. Moreau).

The institutional review boards of The Sainte-Justine University Hospital, The Montreal Children's Hospital, The Shriners Hospital for Children in Montreal and McGill University approved the study.

Structured abstract

Study Design: A replication association study that used genomic data generated from French-Canadian case and control cohorts.

Objectives: To determine whether the 53 single nucleotide polymorphisms (SNPs) that were previously associated with spinal deformity progression in an American Caucasian cohort, are similarly associated in the French-Canadian population.

Summary of Background Data: It is widely accepted that genetic factors contribute to AIS. The identification of genetic variants associated with the predisposition or progression of curvature could facilitate diagnostic/prognostic tool development. Although 53 SNPs have been associated with spinal curve progression in Caucasian cohorts in the USA, these associations were not replicated in a large Japanese-population study, arguing that such a discrepancy could be explained by ethnicity, thus raising the importance of a replication study in an independent Caucasian population of European descent.

Methods: Genomic data were collected from the French-Canadian population, using the Illumina HumanOmni 2.5M BeadChip. Fifty-two SNPs, tested in ScoliScore^{TM} or in high linkage disequilibrium (LD) with SNPs in the test, were selected to assess their association with scoliosis generally, and with spinal curve progression. One SNP in ScoliScore^{TM}, rs16909285, could not be evaluated in our GWAS.

Results: None of the SNPs used in ScoliScore^{TM} were associated with AIS curve progression or curve occurrence in the French-Canadian population. We evaluated 52 SNPs in severe patients by comparing risk allele frequencies with those in non-severe patients and with those in control individuals. There was no significant difference between the severe group and the non-severe group or between the severe group and the control group.

Conclusions: Although the 52 SNPs studied here were previously associated with curve progression in an American population of European descent, we found no association in French-Canadian AIS patients. This second replication cohort suggests that the lack of association of these SNPs in a Japanese cohort is not due to ethnicity.

KEYWORDS: adolescent idiopathic scoliosis, single nucleotide polymorphism, ScoliScore^{TM}, spinal curve progression, French-Canadian, Caucasian, genetic test, genotype, association analysis, statistical power

Key Points

--Previously reported association of 53 SNPs with curve progression in white AIS patients was evaluated in a French-Canadian cohort.

--The association is not statistical significant in the first replication study in Caucasians.

--The lack of association of these SNPs in a previous Japanese cohort is not due to ethnicity.

Mini abstract

The association of 53 SNPs with scoliosis progression has been reported in a Caucasian population, generating a commercial product (ScoliScore^{TM}) that evaluates risk of curve progression. A previous study using a Japanese population failed to replicate the association. Our study indicates no genetic association between these SNPs and AIS among French-Canadian population.

Introduction

Adolescent idiopathic scoliosis (AIS) is the most common spinal deformity, affecting an average of about 4% of children globally, from 10 to 18 years old *[Kane 1977; Lonstein 1994; Weinstein, Dolan, Cheng et al. 2008]*. Among those affected, only 10% have curve progression so that medical intervention is required *[Miller 1999]*. It is observed that girls tend to develop progressive curves more often than boys, and the reason for this is unknown *[Roach 1999]*. In pedigrees, AIS tends to cluster so that the disease incidence in patients' relatives is much higher than in the general population, indicating a genetic basis *[Riseborough and Wynne-Davies 1973; Ward, Ogilvie, Argyle et al. 2010; Wynne-Davies 1968]*. Furthermore, there is strong evidence from twin studies showing that genetic factors contribute to AIS *[Andersen, Thomsen and Kyvik 2007; Kesling and Reinker 1997]*. Identification of genetic factors that are associated with AIS could facilitate screening for risk of curve onset and/or progression.

In 2010, from an unpublished Genome-Wide association study (GWAS), Ward *et al.* selected 53 single nucleotide polymorphisms (SNPs) associated with AIS curve progression among Caucasian female patients in the United States. Based on the genotype data for these 53 SNPs, as well as the patients' initial Cobb angle measured between 9 and 13 years, they built an algorithm to quantify risk of spinal curve progression for Caucasian patients with a Cobb angle <25º, which was then commercialized under the name of ScoliScore^{TM}. Although not yet approved by the FDA (the U.S Food and Drug Administration), ScoliScore^{TM} is the only DNA-based test developed to identify patients with mild AIS in Caucasian population who have low risk of spinal curve progression *[Ward, Ogilvie, Singleton et al. 2010]*.

However, in an independent study of Caucasian AIS patients who received ScoliScore^{TM} testing, no significant difference was found in risk scores between patients at low risk and at high risk, both being evaluated by traditional clinical estimates *[Roye, Wright, Williams et al. 2012]*. In addition, a recent study failed to replicate any genetic association between the 53 SNPs (of ScoliScore^{TM}) and spinal curve progression in a Japanese population *[Ogura, Takahashi, Kou et al. 2013]*. To determine whether this association is exclusive to the

Caucasian population, we conducted a replication study in a French-Canadian cohort using genomic data.

Materials and methods

Study population and data source

This study has the approval from the institutional review boards of The Sainte-Justine University Hospital, The Montreal Children's Hospital, The Shriners Hospital for Children in Montreal and McGill University, as well as the Affluent and Montreal English School Boards. We recruited 1056 individuals from schools. Additional genomic control data of 750 individuals were from the CARTaGENE project *[Awadalla, Boileau, Payette et al. 2013; Godard, Marshall and Laberge 2007]*. To rule out the presence of scoliosis among study populations, school screening was conducted by one of the orthopedic surgeons at Sainte-Justine Hospital in Montreal, Quebec, Canada, and the CARTaGENE adult phenotype records were checked.

Genomic DNA samples were extracted from the peripheral blood for the subjects at the hospital and schools and then genotyped by the Illumina HumanOmni 2.5-8 BeadChip. Control data from CARTaGENE was merged into the microarray outcome to generate files in the appropriate format to be analyzed.

Quality control for genomic data

Quality control (QC) measure was applied to genomic data following previously outlined standards *[Turner, Armstrong, Bradford et al. 2011; Weale 2010]*. PLINK *[Purcell, Neale, Todd-Brown et al. 2007]* and R *[Team 2012]* software packages were utilized to **1)** filter gender mismatches, **2)** filter missingness at both the sample- (< 2%) and SNP-level (< 2%), **3)** assessment of sample heterozygosity, **4)** filter SNPs with a minor allele frequency (MAF) less than 1%, and **5)** filter SNPs in Hardy-Weinberg disequilibrium *[Neale and Purcell 2008; Samani, Erdmann, Hall et al. 2007]*. Linkage disequilibrium (LD) thinning was performed on the filtered genomic data prior to ancestral and relatedness testing using EIGENSTRAT *[Price, Patterson, Plenge et al. 2006]* and PLINK *[Purcell, Neale, Todd-*

Brown et al. 2007] identity-by-descent (IBD), respectively. 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, 170 individuals from schools and 731 individuals from the CARTaGENE project).

Definition of severe and non-severe cases

Curve severity was defined by the Cobb angle that was recorded at the last clinic visit. As scoliosis curvatures vary from a single type to quadruple type, the worst curve or the major curve for each individual was used to determine AIS patients' severity. Severe patients were defined by a major curve Cobb angle ≥40º. Two patients were included even though their major curve Cobb angles were less than 40º at their last visits. One female with a major curve of 39º was considered as a severe case because she was less than 12 years old at the last recorded visit. The second exception was a patient with a curvature of 37º at her last visit, since her highest Cobb angle in prior records reached 41º.

Non-severe AIS patients were defined as their major curve Cobb angle between 10º and 39º by skeletal maturity. To simplify the concept "skeletal maturity", we fixed the cutoff age as 14 years for girls and 16 years for boys. Because generally AIS has an unclear curvature progression tendency, patients having major curve Cobb angle 10º-39º, younger than 14 for girls and younger than 16 for boys, were excluded from non-severe group, but were still kept in case group. The non-severe group consisted of skeletally mature patients only. Certain individuals, whose major curve Cobb angles were less than 10º at their last visit, were included in non-severe group if their prior spinal curve degrees had been ≥10º and were reduced because of bracing impact or unclear reason.

As a result of the criteria we define here, 148 patients were classified as severe and 302 patients were in a non-severe group. Clinical characteristics of two groups are shown in **Table I.**

Association study

Among the 53 SNPs previously associated with curve progression among Caucasians *[Ward, Ogilvie, Singleton et al. 2010]*, 25 were included in our genomic data (**Table II**). For the remaining 28 SNPs, we searched for proxy SNPs that were in high linkage disequilibrium (LD). Conceptually, LD is when an allele of one SNP is often observed with an allele of another SNP within a population. Thus, the allele of the one SNP is able to represent the allele of the other SNP. We queried the 28 SNPs using SNAP (www.broadinstiture.org/mpg/snap/), an online tool for SNP Annotation and Proxy, based on genotype data from the International HapMap Project and the 1000 Genomes Project *[Johnson, Handsaker, Pulit et al. 2008]*. We restricted our search to SNPs represented on the Illumina OmniChip 2.5M array, and used an $r^2>0.8$ as a cutoff for a proxy in European ancestry population. In the genetic analysis, LD is reported in terms of D' (D-prime) and r^2 (r-square). Both are statistical measures of linkage disequilibrium scaled from 0 to 1. The case D'=1 is referred to as complete LD, indicating no recombination between the two SNPs within the population. The case $r^2=1$ happens exclusively if 2 of the 4 possible haplotypes are present in the population and the two SNPs have the same allele frequencies, which is referred to as perfect LD. SNPs in perfect LD are necessarily in complete LD, but SNPs in complete LD may have low r^2 value if the alleles at two loci are not correlated. These values are represented for each SNP in **Table III**. We found 27 SNPs in our genomic dataset that are in high LD with their relative SNPs in ScoliScore™. However, no SNP matched our query criteria to represent the rs16909285 SNP in ScoliScore™.

Using PLINK software, we evaluated the association of 52 SNPs with AIS in French-Canadian population by chi-square test. Considering that the 53 SNPs in the original study were associated among Caucasian female AIS patients only, we conducted our association analyses in all French-Canadian samples as well as in females only. For all SNPs, we evaluated associations among totals and among female case versus female control for: **1)** presence of scoliosis versus controls; **2)** severe scoliosis versus controls; and **3)** severe scoliosis versus non-severe scoliosis.

For statistical significance, we used a conservative Bonferroni correction to adjust the p-value depicting probable association. We adjusted the probability of the false positive results from 0.05 to (0.05/k) where k is the number of SNPs tested in each independent association

test (k=52 in our study). Therefore, SNPs with a p-value $<1\times10^{-3}$ demonstrate significant association in statistics *[Bush and Moore 2012]*.

Statistical power calculation

The pwr package in R software (http://www.statmethods.net/stats/power.html) *[Champely and Champely 2007]* was used in statistical power calculations for each association study while effect size was defined as small, medium and large, respectively, as outlined by Cohen *[Cohen 1988]*.

Results

Disease-associated study

To evaluate whether the 52 SNPs are associated with the occurrence of spinal curvature, we compared the frequency of each SNP among all AIS patients to those of all controls. We applied the same comparison in female samples as well (545 cases vs. 476 controls). As shown in **Table IV-V**, none of the 52 SNPs were significantly associated (p-value$<1\times10^{-3}$) in either cohort.

Progression-associated study

To detect the genetic association with AIS progression, we conducted two association analyses independently: one between 148 severe AIS patients and 302 non-severe patients, the other between 148 severe patients and 901 healthy controls, in both genders as well as in females. All the association analysis results are shown in **Tables VI-IX**. There is no association between the SNPs and severe AIS, in total samples or in female samples in French-Canadian population.

Statistic power analysis

Power calculations for each association study are listed in **Table X**. We concluded that the statistical power for each association analysis is strong enough for medium genetic effect.

Discussion

In this study, we first attempted to replicate the association between AIS and the 53 SNPs from an unpublished GWAS, and then we attempted to replicate the association of these SNPs to severe curvature. Although the genome chip used in our study contained only 25 of the SNPs published by Ward *et al. [Ward, Ogilvie, Singleton et al. 2010]*, we identified 27 SNPs in high linkage disequilibrium with the remaining SNPs of the original study. One SNP was completely unavailable for us to study. We were unable to replicate any association between these 52 SNPs and AIS in our French-Canadian population.

This is the second study that has not replicated the association between AIS and the 53 SNPs used in the algorithm that the ScoliScore^{TM} test employs for its prediction of risk of curve progression. Recently, Y. Ogura *et al.* genotyped Japanese AIS patients, of which 600 individuals were divided into a progression group and 1114 individuals were divided into a non-progression group. With power greater than 80% in 24 out of 53 SNPs, no association with curve severity was found *[Ogura, Takahashi, Kou et al. 2013]*. However, it was possible that this lack of replication in the Japanese study population came from the ethnic admixture in Japanese and Caucasian cohorts between the two studies. Populations having distinct migration sources are likely to have a different disease penetrance due to varying degrees of genetic contributions, resulting in population stratification. Thus, our study sought to ascertain association of the SNPs in a Caucasian population of European descent, similar to that of the original study. Furthermore, an earlier study found no significant correlation in risk prediction of curve progression between ScoliScore^{TM} results and common clinical estimates in 83 Caucasians *[Roye, Wright, Williams et al. 2012]*, emphasizing the need for replication of the original genetic association in a Caucasian cohort.

That the original GWAS that produced the association between AIS and the 53 SNPs has not been published brought hesitation and consideration about the scientific foundation of ScoliScore^{TM} *[Dobbs and Gurnett 2011; Grant and Dormans 2011]*. To evaluate the statistical power of the initial study, we lack important details in the study design such as: control cohort definition, quality control criteria for SNPs and subjects, quantity of testing markers and adjusted significance level. In light of this missing information and assuming that the original GWAS was a classical case/control design, we first tested the association of the 53 SNPs to

AIS, and then tested for association to curve severity. With 100% power to detect a moderate to strong genetic effect and 75% power to detect a minor effect, we did not find an association between the 52 tested SNPs and AIS. Using two approaches, we did not find an association between curve severity and the 52 SNPs. By our calculations, we had sufficient power to detect a large and moderate effect, although we had reduced power to detect a minor genetic effect. However, with a study sample composed of 450 patients classified into a severe group and a non-severe group, the latter test was similar in size to that in the ScoliScore^{TM} validation study.

Possible reasons for the irreproducibility of genetic associations lie in various factors that affect the statistical power in association studies *[Hirschhorn, Lohmueller, Byrne et al. 2002; McClellan and King 2010; Sham and Purcell 2014]*. One important determinant of statistical power in association studies is variable LD between studied markers and the true causal variants *[Hirschhorn, Lohmueller, Byrne et al. 2002]*. For the proxy SNPs that we used in our study that are not in perfect LD with the query SNPs of the original study, it is possible that recombination events among individuals caused disassociation. Sixteen out of 27 proxy SNPs were in absolute LD (r^2=1, D'=1) with query SNPs. In 2 cases where D'<1, there was still chance in French-Canadian population that query SNPs have been separated from proxy SNPs by recombination events. Even no recombination happened in 9 cases (D'=1, r^2<1), proxy SNPs could not substitute completely query ones as the allele frequencies at two loci were not exactly the same. Still with $r^2 \geq 0.8$, we had a quite small number of subjects carrying mismatched alleles from SNPs in LD *[Wray 2005]*. Therefore, it is quite reasonable to expect that the negative result of genetic association of SNPs in ScoliScore^{TM} with spinal curve progression in French-Canadian population was not entirely due to our employment of proxy SNPs. Importantly, there was no evidence that the 53 identified SNPs were causal variants in AIS progression. They might have a correlation with the causal variants because of linkage disequilibrium in the initial study. Therefore, an increased sample size was required in the replication study to reach the same level of statistical power in the initial study *[Hirschhorn, Lohmueller, Byrne et al. 2002; Sham and Purcell 2014]*, which reduced the probability to reproduce the prior association results in our study.

Another important consideration for the discrepancy between our results and the original study is the criteria used to define the phenotype. Firstly, skeletal maturity was

defined slightly differently between our study and the initial study. Ward *et al.* defined mature patients by a Risser sign level or by chronological age. In our study, we used the age recorded at the last clinic visit. Secondly, the criteria to discriminate severe and non-severe curves were not exactly the same. Skeletally mature individuals having Cobb angle 40°-50° were allocated into the severe group in our study, while they were classified into non-severe group in the initial study *[Ward, Ogilvie, Singleton et al. 2010]*. Difference in disease severity definition influences the statistical power among independent studies, influencing in the reproducibility of the initial association study *[Sham and Purcell 2014]*.

To multigenetic disease such as AIS, weak genetic effects lead to false positive associations that cannot be replicated *[Hirschhorn, Lohmueller, Byrne et al. 2002]*. Obtained from initial GWAS *[Ward, Ogilvie, Singleton et al. 2010]*, the odds ratios, representing the effect size for the risk allele associated with severe scoliosis, varied from 0.26 to 1.94, suggesting that the effect size was small and might even be overestimated because of the phenomenon "Winner's curse" *[Xiao and Boehnke 2009]*. In auctions, the winners are likely to overestimate the true value of the item. In association studies, the first published positive report is equivalent to the winning bid. It is more likely that the underlying genetic effect size is upwardly biased in the original discovery study, causing the failure of replication study in small size.

Genetic heterogeneity is another factor that makes replication studies difficult regarding complex diseases. Although our cohorts and initial studied cohorts are both Caucasian, it is still possible that the same variants have distinct behaviors in clinical manifestation or phenotype, as well as that several causal variants lead to the same phenotype *[McClellan and King 2010]*. Ideally, a replication study for original GWAS would be conducted in another independent population, which is in perfect match with the initial one in the genetic and environmental background *[Hirschhorn, Lohmueller, Byrne et al. 2002]*. Although Ward *et al.* reported their results in validating AIS progression risk score, the discovery study from which they identified the genetic association of 53 SNPs was not mentioned, nor was any replication study by them in another matched population to strengthen their findings.

In summary, our study attempted to reproduce the association of SNPs originally associated with AIS and used to calculate risk of curve progression in the ScoliScore^{TM} test.

As demonstrated by an earlier study in a Japanese cohort, we did not find any significant association to AIS generally or to curve severity, in a Caucasian French-Canadian population. This study suggests that the lack of replication in the Japanese population is not due to ethnicity.

Acknowledgements

We are indebted to the subjects and families who participated in this study. We also thank all the orthopedic surgeons and nursing teams that have participated in this study at Sainte-Justine University Hospital, The Montreal's Children Hospital, and The Shriners Hospital for Children in Montreal.

This work was supported by grants from La Fondation Yves Cotrel de l'Institut de France, Paris, France, Paradigm Spine LLC/Fourth Dimension Spine LLC (NYC, USA) and Génome Québec. Dr. Gorman is the recipient of CHU Sainte-Justine Research Centre/CIHR MENTOR training program Postdoctoral Award for Excellence in Musculoskeletal Research. No other potential conflict of interest relevant to this article was reported.

Tables and figures

Table I. Demographic and clinical characteristics of severe patients and non-severe patients with AIS.

Subjects	Characteristics								
	All Subjects			Female			Male		
	N	Mean Age (Years)	Scoliosis Cobb Angles (°)	N	Mean Age (Years)	Scoliosis Cobb Angles (°)	N	Mean Age (Years)	Scoliosis Cobb Angles (°)
Severe AIS patients	148	15 ± 2 (10 – 25)	56 ± 12 (37 – 90)	129	15 ± 2 (10 – 25)	55 ± 12 (37 – 90)	19	16 ± 2 (12 – 19)	60 ± 11 (40 – 87)
Non-severe AIS patients	302	16 ± 1 (14 – 22)	21 ± 8 (3 – 39)	259	16 ± 1 (14 – 22)	21 ± 9 (3 – 39)	43	17 ± 1 (16 – 19)	19 ± 7 (7 – 35)

All values represent mean Cobb Angles ± standard deviation, and range values for respective groups.

Severe AIS patient was defined as Cobb angle ≥40º for major spinal curves.

Non-severe AIS patient was defined as the highest historical record of Cobb angle between 10º and 39º for spinal curves by skeletal maturity (girls ≥ 14 years and boys ≥ 16 years).

Table II. Twenty-five SNPs included both in ScoliScore™ and in Illumina genotyping microarray.

dbSNP	Chromosome	Associated Gene
rs6691909	chr1	AIM1L
rs10493083	chr1	RRAGC
rs17021437	chr1	AMY1C
rs2209158	chr1	KCNC4
rs12474952	chr2	ID2
rs1991127	chr2	APOB
rs17044552	chr2	KLHL29
rs6798946	chr3	ARPP21
rs6414345	chr3	CLSTN2
rs11747787	chr5	TNIP1
rs6420139	chr6	SCAF8
rs4724981	chr7	MICALL2
rs6952104	chr7	NPY
rs2976514	chr8	NRG1
rs2449539	chr8	LAPTM4B
rs10794280	chr11	MUC2
rs17210350	chr11	EED
rs1558729	chr12	NEDD1
rs1265566	chr12	CUX2
rs4765072	chr12	TMEM132B
rs17719756	chr14	EXOC5
rs1437480	chr15	FOXB1
rs9945359	chr18	SETBP1
rs17635546	chr19	NLRP11
rs132898	chr22	KIAA1671

Table III. Twenty-seven SNPs in ScoliScore™ and their proxy SNPs in Illumina genotyping microarray.

ScoliScore™		1000 GENOMES Pilot1, panel CEU		Illumina HumanOnmi2.5M	
dbSNP	Chromosome	r-square	D-prime	Proxy SNP	Associated Gene
rs4661748	chr1	1	1	rs4661747	SPATA21
rs6693477	chr1	1	1	rs7365544	EFNA3
rs10798036	chr1	1	1	rs6425017	HMCN1
rs16865244	chr2	0.831	1	rs16865273	CMPK2
rs12618119	chr2	0.926407	1	rs6711194	GYPC
rs10168146	chr2	1	1	rs6431278	ARL4C
rs7613792	chr3	1	1	rs13433861	ZBTB20
rs10004901	chr4	1	1	rs10011602	C4orf22
rs10000472	chr4	1	1	rs1384135	ARHGAP24
rs2045904	chr5	0.965	1	rs10512969	ITGA1
rs831653	chr5	1	1	rs831649	GPBP1
rs16902899	chr5	1	1	rs2178270	TMEM161B
rs239794	chr6	0.84	0.964	rs12192659	FAM83B
rs1349887	chr6	0.967	1	rs1902064	ARID1B
rs2700910	chr7	1	1	rs2726052	EEPD1
rs17165447	chr7	1	1	rs4729090	CALCR
rs7840870	chr8	0.934	0.966	rs17817357	RIMS2
rs10787096	chr10	0.962	1	rs10884639	SORCS1
rs16968878	chr16	1	1	rs2113177	CHD11
rs4782809	chr16	0.966	1	rs4782543	CHD13
rs16945692	chr17	1	1	rs12451910	INTS2
rs11083276	chr18	0.864	1	rs2311719	CDH2
rs8093693	chr18	0.93	1	rs2909638	SERPINB8
rs448013	chr20	1	1	rs447915	CBLN4
rs136187	chr22	0.894	1	rs767855	MYH9
rs6528028	chrX	1	1	rs952077	GPM6B
rs500243	chrX	1	1	rs485156	SLC16A2

Table IV. Association of 25 ScoliScore™ SNPs with AIS in French-Canadian population (667 cases vs. 901 controls; 545 female cases vs. 476 female controls).

ScoliScore™	AIS case vs. Control		female-only AIS case vs. Control	
dbSNP ID	P value	Odds Ratio (95% CI)	P value	Odds Ratio (95% CI)
rs6691909	0.9371	0.9943 (0.8627-1.146)	0.4078	0.9291 (0.7806-1.106)
rs10493083	0.353	0.7784 (0.4582-1.322)	0.4534	0.7825 (0.4114-1.488)
rs17021437	0.003928	0.6041 (0.4275-0.8536)	0.0318	0.6374 (0.4213-0.9645)
rs2209158	0.1936	0.9069 (0.7827-1.051)	0.05134	0.8357 (0.6976-1.001)
rs12474952	0.1656	0.9033 (0.7822-1.043)	0.1013	0.8624 (0.7225-1.03)
rs1991127	0.2083	1.36 (0.8409-2.199)	0.4647	1.247 (0.689-2.258)
rs17044552	0.4882	1.147 (0.7786-1.689)	0.2045	1.373 (0.8399-2.244)
rs6798946	0.4739	0.8906 (0.6485-1.223)	0.4126	0.8499 (0.5758-1.255)
rs6414345	0.3685	0.8303 (0.5534-1.246)	0.2847	0.7653 (0.4682-1.251)
rs11747787	0.4884	0.9509 (0.8246-1.096)	0.05752	0.8441 (0.7086-1.005)
rs6420139	0.3425	0.9261 (0.7902-1.085)	0.148	0.8672 (0.7149-1.052)
rs4724981	0.2588	1.086 (0.9413-1.252)	0.4653	1.067 (0.896-1.271)
rs6952104	0.7274	0.9751 (0.8461-1.124)	0.9847	0.9983 (0.8387-1.188)
rs2976514	0.8504	0.9863 (0.8542-1.139)	0.9077	1.011 (0.8467-1.206)
rs2449539	0.5843	0.8782 (0.5514-1.399)	0.5321	0.8348 (0.4735-1.472)
rs10794280	0.6155	1.039 (0.8956-1.205)	0.446	1.073 (0.8946-1.288)
rs17210350	0.2268	0.766 (0.4966-1.182)	0.4421	0.8121 (0.4772-1.382)
rs1558729	0.9912	1.001 (0.8576-1.168)	0.7451	0.9689 (0.8007-1.172)
rs1265566	0.6256	0.9629 (0.8274-1.121)	0.8049	0.977 (0.8122-1.175)
rs4765072	0.6903	1.029 (0.8931-1.186)	0.9763	0.9974 (0.8378-1.187)
rs17719756	0.2696	1.083 (0.9399-1.248)	0.2783	1.101 (0.9251-1.311)
rs1437480	0.06253	0.6451 (0.4053-1.027)	0.2936	0.7524 (0.4418-1.282)
rs9945359	0.288	1.12 (0.9087-1.381)	0.6897	1.054 (0.8127-1.368)
rs17635546	0.1921	0.8745 (0.7149-1.07)	0.4245	0.9048 (0.7077-1.157)
rs132898	0.3077	1.077 (0.9341-1.241)	0.5289	1.058 (0.8883-1.259)

CI: confidence interval
Significance level is p-value ≤ 0.001

Table V. Association of 27 Proxy SNPs with AIS in French-Canadian population (667 cases vs. 901 controls; 545 female cases vs. 476 female controls).

Proxy	AIS case vs. Control		female-only AIS case vs. Control	
dbSNP ID	P value	Odds Ratio (95% CI)	P value	Odds Ratio (95% CI)
rs4661748	0.2714	0.9187(0.7898-1.069)	0.2007	0.8854(0.7347-1.067)
rs6693477	0.9596	1.004(0.8712-1.156)	0.7137	1.033(0.8682-1.229)
rs10798036	0.07926	1.135(0.9853-1.308)	0.2867	1.099(0.9235-1.309)
rs16865244	0.9229	1.01(0.821-1.243)	0.8898	1.018(0.7903-1.311)
rs12618119	0.08106	1.143(0.9836-1.328)	0.3028	1.102(0.9161-1.325)
rs10168146	0.2655	1.094(0.9338-1.282)	0.2246	1.128(0.9287-1.37)
rs7613792	0.837	0.9482(0.5713-1.574)	0.1704	0.6578(0.3599-1.202)
rs10004901	0.8966	0.9847(0.781-1.242)	0.8375	0.9711(0.7338-1.285)
rs10000472	0.5558	1.049(0.8951-1.229)	0.5396	1.063(0.8753-1.29)
rs2045904	0.6947	0.972(0.8435-1.12)	0.6017	1.048(0.8799-1.247)
rs831653	0.9786	1.002(0.8685-1.156)	0.3455	0.9191(0.7712-1.095)
rs16902899	0.7969	0.9774(0.8213-1.163)	0.8609	1.019(0.8229-1.263)
rs239794	0.6133	0.9634(0.8337-1.113)	0.6778	1.039(0.8689-1.241)
rs1349887	0.5058	1.049(0.9107-1.209)	0.3844	1.08(0.9078-1.285)
rs2700910	0.6668	1.038(0.8773-1.227)	0.4831	1.076(0.8766-1.321)
rs17165447	0.9199	1.019(0.7125-1.456)	0.822	0.9526(0.624-1.454)
rs7840870	0.2563	1.086(0.9416-1.254)	0.2093	1.119(0.9387-1.335)
rs10787096	0.9363	1.006(0.8701-1.163)	0.9175	1.01(0.8439-1.208)
rs16968878	0.08869	0.8808(0.761-1.019)	0.03976	0.8291(0.6935-0.9913)
rs4782809	0.7197	1.026(0.8899-1.184)	0.7515	0.9721(0.8157-1.158)
rs16945692	0.2633	0.8919(0.7299-1.09)	0.2432	0.8666(0.6812-1.102)
rs11083276	0.8624	0.986(0.8402-1.157)	0.5932	1.055(0.8671-1.283)
rs8093693	0.7541	0.9772(0.8457-1.129)	0.5142	0.9426(0.7893-1.126)
rs448013	0.3316	0.7886(0.4877-1.275)	0.4427	0.7993(0.4506-1.418)
rs136187	0.6034	0.9607(0.8259-1.118)	0.9515	1.006(0.8344-1.213)
rs6528028	0.209	0.8955(0.7539-1.064)	0.4343	0.9252(0.7613-1.124)
rs500243	0.8811	0.9876(0.8384-1.163)	0.7203	1.034(0.8598-1.244)

CI: confidence interval
Significance level is p-value ≤ 0.001

Table VI. Association of 25 ScoliScore™ SNPs with AIS progression in French-Canadian population (148 severe cases vs. 901 controls; 129 severe female cases vs. 476 female controls).

ScoliScore™	Severe case vs. Control		female-only Severe case vs. Control	
dbSNP ID	P value	Odds Ratio (95% CI)	P value	Odds Ratio (95% CI)
rs6691909	0.04146	1.292 (1.01-1.653)	0.446	1.113 (0.8451-1.466)
rs10493083	0.3886	0.6359 (0.2253-1.795)	0.1575	0.3641 (0.08454-1.568)
rs17021437	0.04324	0.4968 (0.2488-0.992)	0.04771	0.4548 (0.2046-1.011)
rs2209158	0.1732	0.8351 (0.6442-1.083)	0.1467	0.8074 (0.6046-1.078)
rs12474952	0.08775	0.8023 (0.6229-1.033)	0.0979	0.7872 (0.5927-1.045)
rs1991127	0.1942	1.631 (0.774-3.436)	0.1572	1.775 (0.7933-3.971)
rs17044552	0.8856	1.051 (0.5312-2.081)	0.3895	1.381 (0.6598-2.892)
rs6798946	0.3935	0.7736 (0.4284-1.397)	0.5218	0.8112 (0.4272-1.54)
rs6414345	0.3138	0.6678 (0.3029-1.473)	0.4502	0.7291 (0.32-1.661)
rs11747787	0.8042	1.032 (0.8061-1.32)	0.7507	0.9562 (0.7256-1.26)
rs6420139	0.9903	0.9983 (0.7601-1.311)	0.6543	0.9328 (0.688-1.265)
rs4724981	0.6229	1.064 (0.8307-1.363)	0.6573	1.065 (0.8074-1.404)
rs6952104	0.5245	1.083 (0.847-1.385)	0.4714	1.106 (0.8402-1.457)
rs2976514	0.4468	1.101 (0.8592-1.411)	0.5118	1.098 (0.831-1.45)
rs2449539	0.5898	0.7898 (0.3343-1.866)	0.3163	0.5839 (0.2014-1.693)
rs10794280	0.9446	0.9909 (0.7651-1.283)	0.9647	1.007 (0.7536-1.344)
rs17210350	0.1523	0.516 (0.2052-1.297)	0.1902	0.5008 (0.1744-1.438)
rs1558729	0.972	1.005 (0.7685-1.314)	0.631	0.9281 (0.6846-1.258)
rs1265566	0.6028	0.932 (0.7149-1.215)	0.6779	0.9396 (0.7003-1.261)
rs4765072	0.97	1.005 (0.7853-1.285)	0.9005	1.018 (0.7726-1.341)
rs17719756	0.1599	0.8368 (0.6525-1.073)	0.3607	0.8787 (0.6658-1.16)
rs1437480	0.4986	0.7605 (0.3432-1.685)	0.7311	0.864 (0.375-1.99)
rs9945359	0.6268	1.094 (0.7621-1.57)	0.8254	0.9537 (0.6256-1.454)
rs17635546	0.3251	0.8346 (0.582-1.197)	0.4151	0.8463 (0.5664-1.265)
rs132898	0.6935	1.051 (0.8211-1.345)	0.8761	1.022 (0.7756-1.347)

CI: confidence interval
Significance level is p-value ≤ 0.001

Table VII. Association of 27 Proxy SNPs with AIS progression in French-Canadian population (148 severe cases vs. 901 controls; 129 severe female cases vs. 476 female controls).

Proxy	Severe case vs. Control		female-only Severe case vs. Control	
dbSNP ID	P value	Odds Ratio (95% CI)	P value	Odds Ratio (95% CI)
rs4661748	0.2831	0.8644 (0.6623-1.128)	0.4343	0.8883 (0.6599-1.196)
rs6693477	0.1778	0.8441 (0.6596-1.08)	0.4643	1.108 (0.8415-1.459)
rs10798036	0.1208	1.215 (0.9498-1.554)	0.3357	1.145 (0.8692-1.508)
rs16865244	0.591	1.101 (0.7753-1.563)	0.5188	1.136 (0.7708-1.675)
rs12618119	0.07728	1.259 (0.9748-1.626)	0.3013	1.164 (0.8724-1.554)
rs10168146	0.4919	1.101 (0.8373-1.447)	0.4241	1.132 (0.8348-1.536)
rs7613792	0.9764	0.987 (0.4129-2.359)	0.5284	0.7328 (0.2777-1.933)
rs10004901	0.7321	1.071 (0.7235-1.585)	0.8332	0.953 (0.6091-1.491)
rs10000472	0.9832	0.997 (0.756-1.315)	0.7597	0.9526 (0.6976-1.301)
rs2045904	0.4144	1.108 (0.8663-1.417)	0.2004	1.197 (0.9088-1.576)
rs831653	0.2515	1.155 (0.9026-1.478)	0.4656	1.108 (0.841-1.46)
rs16902899	0.07997	1.288 (0.9697-1.712)	0.03737	1.399 (1.019-1.921)
rs239794	0.6146	0.9373 (0.7286-1.206)	0.9584	0.9925 (0.748-1.317)
rs1349887	0.08914	1.238 (0.9676-1.584)	0.1576	1.22 (0.9258-1.607)
rs2700910	0.846	0.9712 (0.7231-1.304)	0.5899	0.9122 (0.6531-1.274)
rs17165447	0.9981	1.001 (0.5366-1.866)	0.6555	0.8524 (0.4223-1.72)
rs7840870	0.293	1.142 (0.8916-1.462)	0.5384	1.091 (0.8264-1.441)
rs10787096	0.005894	0.6907 (0.5303-0.8998)	0.03061	0.7225 (0.5378-0.9708)
rs16968878	0.1386	0.8239 (0.6374-1.065)	0.0635	0.7626 (0.5725-1.016)
rs4782809	0.7546	1.04 (0.8121-1.332)	0.7995	0.9647 (0.7307-1.273)
rs16945692	0.5617	0.9014 (0.6349-1.28)	0.4187	0.8521 (0.578-1.256)
rs11083276	0.4956	0.9063 (0.683-1.203)	0.7735	0.9549 (0.6972-1.308)
rs8093693	0.5068	0.9181 (0.7134-1.182)	0.3583	0.8758 (0.6598-1.162)
rs448013	0.2092	0.5229 (0.1868-1.464)	0.3163	0.5839 (0.2014-1.693)
rs136187	0.9478	0.9913 (0.7629-1.288)	0.9596	1.008 (0.7498-1.354)
rs6528028	0.9359	0.9884 (0.7435-1.314)	0.6799	1.066 (0.7874-1.442)
rs500243	0.3661	0.8792 (0.665-1.162)	0.6302	0.9299 (0.6918-1.25)

CI: confidence interval
Significance level is p-value ≤ 0.001

Table VIII. Association of 25 ScoliScoreTM SNPs with AIS progression in French-Canadian population (148 severe cases vs. 302 non-severe cases; 129 severe female cases vs. 259 non-severe female cases).

ScoliScore™	Severe case vs. Non-severe case		female-only Severe case vs. Non-severe case	
dbSNP ID	P value	Odds Ratio (95% CI)	P value	Odds Ratio (95% CI)
rs6691909	0.01073	1.438 (1.087-1.902)	0.047	1.354 (1.004-1.828)
rs10493083	0.498	0.6758 (0.2161-2.114)	0.2192	0.3969 (0.08632-1.825)
rs17021437	0.8289	0.9157 (0.4117-2.036)	0.5713	0.7747 (0.3194-1.879)
rs2209158	0.407	0.8835 (0.6591-1.184)	0.5835	0.9158 (0.6687-1.254)
rs12474952	0.6149	0.9291 (0.6977-1.237)	0.4899	0.8971 (0.6591-1.221)
rs1991127	0.5186	1.322 (0.5653-3.089)	0.3433	1.524 (0.6338-3.665)
rs17044552	0.7473	0.8832 (0.4148-1.881)	0.8065	0.9091 (0.4239-1.95)
rs6798946	0.3299	0.7248 (0.3784-1.388)	0.736	0.8871 (0.4418-1.781)
rs6414345	0.694	0.8364 (0.343-2.039)	0.9932	1.004 (0.4002-2.519)
rs11747787	0.384	1.132 (0.856-1.498)	0.4163	1.133 (0.8387-1.53)
rs6420139	0.1923	1.232 (0.9002-1.686)	0.4026	1.154 (0.8252-1.614)
rs4724981	0.661	1.065 (0.8044-1.409)	0.3872	1.142 (0.8452-1.543)
rs6952104	0.8548	1.026 (0.7771-1.355)	0.6123	1.08 (0.8013-1.456)
rs2976514	0.4419	1.117 (0.843-1.479)	0.5767	1.09 (0.8055-1.475)
rs2449539	0.4818	0.7144 (0.2787-1.831)	0.2533	0.5281 (0.1735-1.608)
rs10794280	0.8499	0.9722 (0.7258-1.302)	0.738	0.9479 (0.6929-1.297)
rs17210350	0.2433	0.5556 (0.2042-1.511)	0.1568	0.4603 (0.1533-1.382)
rs1558729	0.6963	1.063 (0.7836-1.441)	0.8848	1.025 (0.7352-1.429)
rs1265566	0.365	0.8713 (0.6467-1.174)	0.508	0.8982 (0.6537-1.234)
rs4765072	0.9104	0.9841 (0.7447-1.3)	0.5435	1.097 (0.8133-1.48)
rs17719756	0.005894	0.6746 (0.5094-0.8933)	0.03268	0.7208 (0.5335-0.9738)
rs1437480	0.5744	1.315 (0.5044-3.427)	0.5958	1.296 (0.4963-3.383)
rs9945359	0.9114	1.023 (0.6806-1.539)	0.6093	0.8887 (0.565-1.398)
rs17635546	0.6052	0.8983 (0.5981-1.349)	0.6595	0.9064 (0.5853-1.404)
rs132898	0.5285	0.9143 (0.6918-1.208)	0.4445	0.8898 (0.6596-1.2)

CI: confidence interval
Significance level is p-value ≤ 0.001

Table IX. Association of 27 Proxy SNPs with AIS progression in French-Canadian population (148 severe cases vs. 302 non-severe cases; 129 severe female cases vs. 259 non-severe female cases).

Proxy	Severe case vs. Non-severe case		female-only Severe case vs. Non-severe case	
dbSNP ID	P value	Odds Ratio (95% CI)	P value	Odds Ratio (95% CI)
rs4661748	0.7112	0.9447 (0.699-1.277)	0.8529	0.9699 (0.7019-1.34)
rs6693477	0.04717	0.7541 (0.5705-0.9968)	0.243	1.195 (0.8861-1.611)
rs10798036	0.8081	1.035 (0.7837-1.367)	1	1 (0.7418-1.348)
rs16865244	0.5989	1.113 (0.7463-1.66)	0.7023	1.086 (0.7125-1.654)
rs12618119	0.3165	1.16 (0.8678-1.549)	0.3995	1.145 (0.836-1.567)
rs10168146	0.7736	0.9559 (0.7028-1.3)	0.6818	0.9338 (0.6731-1.296)
rs7613792	0.5544	1.368 (0.4823-3.879)	0.6874	1.26 (0.408-3.89)
rs10004901	0.9205	1.023 (0.6567-1.593)	0.6419	0.8922 (0.5516-1.443)
rs10000472	0.3353	0.8588 (0.63-1.171)	0.1917	0.8008 (0.5735-1.118)
rs2045904	0.3996	1.127 (0.8533-1.489)	0.3885	1.14 (0.8459-1.538)
rs831653	0.3103	1.156 (0.8739-1.528)	0.1259	1.264 (0.9361-1.707)
rs16902899	0.03038	1.438 (1.034-2)	0.008447	1.605 (1.127-2.286)
rs239794	0.9365	1.012 (0.7605-1.346)	0.9903	0.9981 (0.7342-1.357)
rs1349887	0.4985	1.101 (0.8332-1.455)	0.9567	1.008 (0.7474-1.36)
rs2700910	0.4652	0.8841 (0.6353-1.231)	0.2445	0.8086 (0.5653-1.157)
rs17165447	0.9519	0.9786 (0.4846-1.976)	0.905	0.9543 (0.4426-2.058)
rs7840870	0.498	1.102 (0.8325-1.458)	0.8729	0.9757 (0.722-1.319)
rs10787096	0.003236	0.6424 (0.478-0.8633)	0.0152	0.675 (0.491-0.9278)
rs16968878	0.7888	0.961 (0.7184-1.285)	0.6338	0.9269 (0.6782-1.267)
rs4782809	0.3369	1.148 (0.8665-1.52)	0.4139	1.134 (0.8381-1.536)
rs16945692	0.5621	0.8901 (0.6004-1.32)	0.5584	0.882 (0.5792-1.343)
rs11083276	0.4595	0.8869 (0.6452-1.219)	0.477	0.8843 (0.6299-1.241)
rs8093693	0.6206	0.9305 (0.6997-1.237)	0.6678	0.9349 (0.6874-1.272)
rs448013	0.2147	0.5034 (0.1668-1.519)	0.3151	0.5669 (0.1847-1.74)
rs136187	0.8838	1.022 (0.7599-1.375)	0.929	1.015 (0.7361-1.399)
rs6528028	0.4119	1.144 (0.8298-1.576)	0.2891	1.197 (0.8585-1.668)
rs500243	0.3386	0.8598 (0.6309-1.172)	0.3516	0.8593 (0.6245-1.182)

CI: confidence interval
Significance level is p-value ≤ 0.001

Table X. Statistical power calculations for each association study in R software.

	Power (%)					
	Case vs. Control		Severe case vs. Control		Severe case vs. Non-severe case	
	ALL	FEMALE	ALL	FEMALE	ALL	FEMALE
Effect size[a]						
small	74.8	46.2	47.9	20.3	12.1	9.3
medium	100	100	100	100	99.9	99.6
large	100	100	100	100	100	100

[a] Cohen proposed rules of thumb for interpreting effect sizes: a "small" effect size is 20%, a "medium" effect size is 50%, and a "large" effect size is 80%.

References

1. Kane WJ. Scoliosis prevalence: a call for a statement of terms. *Clinical orthopaedics and related research* 1977:43-6.
2. Weinstein SL, Dolan LA, Cheng JC, et al. Adolescent idiopathic scoliosis. *Lancet* 2008;371:1527-37.
3. Lonstein JE. Adolescent idiopathic scoliosis. *Lancet* 1994;344:1407-12.
4. Miller NH. Cause and natural history of adolescent idiopathic scoliosis. *The Orthopedic clinics of North America* 1999;30:343-52, vii.
5. Roach JW. Adolescent idiopathic scoliosis. *The Orthopedic clinics of North America* 1999;30:353-65, vii-viii.
6. Riseborough EJ, Wynne-Davies R. A genetic survey of idiopathic scoliosis in Boston, Massachusetts. *The Journal of bone and joint surgery. American volume* 1973;55:974-82.
7. Wynne-Davies R. Familial (idiopathic) scoliosis. A family survey. *The Journal of bone and joint surgery. British volume* 1968;50:24-30.
8. 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;152A:1178-88.
9. Andersen MO, Thomsen K, Kyvik KO. Adolescent idiopathic scoliosis in twins: a population-based survey. *Spine* 2007;32:927-30.
10. Kesling KL, Reinker KA. Scoliosis in twins. A meta-analysis of the literature and report of six cases. *Spine* 1997;22:2009-14; discussion 15.
11. Ward K, Ogilvie JW, Singleton MV, et al. Validation of DNA-based prognostic testing to predict spinal curve progression in adolescent idiopathic scoliosis. *Spine* 2010;35:E1455-64.
12. Roye BD, Wright ML, Williams BA, et al. Does ScoliScore provide more information than traditional clinical estimates of curve progression? *Spine* 2012;37:2099-103.
13. Ogura Y, Takahashi Y, Kou I, et al. A replication study for association of 53 single nucleotide polymorphisms in a scoliosis prognostic test with progression of adolescent idiopathic scoliosis in Japanese. *Spine* 2013;38:1375-9.
14. Godard B, Marshall J, Laberge C. Community engagement in genetic research: results of the first public consultation for the Quebec CARTaGENE project. *Community genetics* 2007;10:147-58.
15. 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* 2013;42:1285-99.
16. Turner S, Armstrong LL, Bradford Y, et al. Quality control procedures for genome-wide association studies. *Current protocols in human genetics / editorial board, Jonathan L. Haines ... [et al.]* 2011;Chapter 1:Unit1 19.
17. Weale ME. Quality control for genome-wide association studies. *Methods in molecular biology* 2010;628:341-72.
18. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* 2007;81:559-75.
19. Team RC. R: A language and environment for statistical computing. 2012.

20. Samani NJ, Erdmann J, Hall AS, et al. Genomewide association analysis of coronary artery disease. *The New England journal of medicine* 2007;357:443-53.
21. Neale BM, Purcell S. The positives, protocols, and perils of genome-wide association. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* 2008;147B:1288-94.
22. 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.
23. Johnson AD, Handsaker RE, Pulit SL, et al. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 2008;24:2938-9.
24. Bush WS, Moore JH. Chapter 11: Genome-wide association studies. *PLoS computational biology* 2012;8:e1002822.
25. Champely S, Champely MS. The pwr Package. 2007.
26. Cohen J. Statistical power analysis for the behavioral sciences: New York: Academic Press. y= 0.2981 ln (x), 1988.
27. Grant SF, Dormans JP. Re: Ward K, Ogilvie JW, Singleton, et al. Validation of DNA-based prognostic testing to predict spinal curve progression in adolescent idiopathic scoliosis. Spine 2010;35:E1455-64. *Spine* 2011;36:1258; discussion -9.
28. Dobbs MB, Gurnett CA. Re: Ward K, Ogilvie JW, Singleton MV, et al. Validation of DNA-based prognostic testing to predict spinal curve progression in adolescent idiopathic scoliosis. Spine 2010;35:E1455-64. *Spine* 2011;36:1257; author reply
29. Hirschhorn JN, Lohmueller K, Byrne E, et al. A comprehensive review of genetic association studies. *Genetics in medicine : official journal of the American College of Medical Genetics* 2002;4:45-61.
30. Sham PC, Purcell SM. Statistical power and significance testing in large-scale genetic studies. *Nature reviews. Genetics* 2014;15:335-46.
31. McClellan J, King MC. Genetic heterogeneity in human disease. *Cell* 2010;141:210-7.
32. Wray NR. Allele frequencies and the r2 measure of linkage disequilibrium: impact on design and interpretation of association studies. *Twin research and human genetics : the official journal of the International Society for Twin Studies* 2005;8:87-94.
33. Xiao R, Boehnke M. Quantifying and correcting for the winner's curse in genetic association studies. *Genetic epidemiology* 2009;33:453-62.

CHAPTER 3. ARTICLE II

A Genome-Wide Association Study of Adolescent Idiopathic Scoliosis in French-Canadian Population

Authors' contribution

Qi Lin Tang performed association analyses and drafted the initial manuscript.

Alain Moreau is the research director of the study. He conceptualized and supervised study design. He revised the manuscript and approved the final manuscript.

Cedric Julien supervised and designed the study. He provided support and guidance in association analyses.

Robert Eveleigh and **Guillaume Bourque** participated in quality control measurement of genomic data.

Kristen Fay Gorman critically revised the manuscript.

Anita Franco contributed to sample preparation for genotyping and revised the manuscript.

Hubert Labelle, Benoit Poitras, Guy Grimard, Stephan Parent, Jean Ouellet, and Jean-Marc Mac-Thiong participated in the study population screening and recorded patients' medical conditions.

The manuscript is currently in preparation.

A Genome-Wide Association Study of Adolescent Idiopathic Scoliosis in French-Canadian Population

Qi Lin Tang[1,2] BSc, Cedric Julien[1] PhD, Robert Eveleigh[3] BSc, Guillaume Bourque[3] PhD, Anita Franco[1] MSc, Hubert Labelle[4,5] MD, Benoit Poitras[5] MD, Guy Grimard[5] MD, Stephan Parent[5] MD, Jean Ouellet[6] MD, Jean-Marc Mac-Thiong[5] MD, Kristen F. Gorman[1,2] PhD, and Alain Moreau[1,2,7] PhD

[1]Viscogliosi Laboratory in Molecular Genetics of Musculoskeletal Diseases, Sainte-Justine University Hospital Research Center, Montréal, Qc, Canada;
[2]Department of Biochemistry, Faculty of Medicine, Université de Montréal, Montréal, Qc, Canada;
[3]Genome Quebec Innovation Center, McGill University, Montréal, Qc, Canada;
[4]LIS3D Laboratory, Sainte-Justine University Hospital Research Center, Montréal, Qc, Canada;
[5]Orthopedic Division, Sainte-Justine University Hospital and Department of Surgery, Faculty of Medicine, Université de Montréal, Montréal, Qc, Canada;
[6]Orthopedic Division, The Montreal Children's Hospital, Department of Surgery, McGill University, Montréal, Qc, Canada;
[7]Department of Stomatology, Faculty of Dentistry, Université de Montréal, Montréal, Qc, Canada.

*Corresponding author: Prof. Alain Moreau, Sainte-Justine University Hospital Research Center, Director of Viscogliosi Laboratory in Molecular Genetics of Musculoskeletal Diseases, 3175 Cote-Sainte-Catherine Road, Montreal, Quebec, H3T 1C5, Canada. Phone: 514-345-4931; Fax: 514-345-4801

This work was supported by grants from La Fondation Yves Cotrel de l'Institut de France, Paris, France (to Dr. Moreau), and Génome Québec (to Dr. Moreau).

The institutional review boards of The Sainte-Justine University Hospital, The Montreal Children's Hospital, The Shriners Hospital for Children in Montreal and McGill University approved the study.

Structured abstract

Study Design: An association study that used genomic data generated from French-Canadian case and control cohorts.

Objectives: To identify Single Nucleotide Polymorphisms (SNPs) associated with Adolescent Idiopathic Scoliosis (AIS) through a genome-wide association study (GWAS) in a French-Canadian cohort, which represents a Caucasian population of European descent.

Summary of Background Data: The contribution of genetic factors to the pathogenesis of AIS has been widely recognized. The identification of genetic variants associated with the susceptibility or severity of spinal curvature would facilitate the development of diagnostic/prognostic tools. There have been a number of loci identified through GWAS in other populations. The population in Quebec is unique because the incidence of idiopathic scoliosis is generally higher in Quebec (average 4.5%), and because 51% of our cohort reported a familial incidence of scoliosis, we expected a strong genetic effect in our population.

Methods: We recruited 667 AIS patients and 901 healthy control individuals from the French-Canadian population. Genomic DNA was extracted from blood and was genotyped using the Illumina HumanOmni 2.5M BeadChip, a commercial genotyping platform with a high density of SNPs. Genotyping data quality control was ensured by previously outlined standards.

Results: We evaluated the association of 1.4 million SNPs through allelic association analysis. Three variants were identified significantly associated with spinal curve predisposition and/or progress, suggesting several novel candidate genes involved in the disease etiopathogenesis.

Conclusions: A genome-wide association study was performed to find genomic variants linked to Adolescent Idiopathic Scoliosis in French-Canadian population using a genotyping microarray with the highest density of SNPs used in the AIS research field so far. We identified several genetic variants linked to disease susceptibility and/or severity and suggested novel candidate genes in etiopathogenesis. Associated loci already reported were not significant in the French-Canadian cohort. The observation of non-association may derive from population stratification and genetic heterogeneity of AIS. Further replication with larger samples is required to validate our findings.

KEYWORDS: adolescent idiopathic scoliosis, single nucleotide polymorphism, genetic variant, genome-wide association study, French-Canadian population, Caucasian, spinal curvature severity, genotype, association analysis

Key Points

- A genome-wide association study demonstrated a number of genetic variants linked to AIS susceptibility and/or severity, suggesting novel candidate genes that play a role in disease etiopathogenesis.
- Using a microarray with a very high density of SNPs, we have performed the most comprehensive genomic survey done yet.
- The observation of non-association of previously reported variants may derive from population stratification and genetic heterogeneity of AIS disease.
- Requirement of further replication study with larger samples and in other ethnic populations is important to validate our findings in GWAS.

Mini abstract

Genetic association of a number of loci has been identified with adolescent idiopathic scoliosis (AIS) through genome-wide association studies. Here with an association study in a French-Canadian population, we identified significant variants linked to AIS, suggesting novel candidate genes that may play a role in spinal curve predisposition and/or progress.

Introduction

Idiopathic scoliosis (IS, OMIM #181800) is the most common spinal deformity, characterized by a lateral curvature of the spine greater than 10 degrees, although it usually manifests in three-dimensions, without a clear cause. Adolescents make up 80% of all IS cases *[Riseborough and Wynne-Davies 1973]*, which is termed as adolescent idiopathic scoliosis (AIS). AIS affects about 2.5% children in the world *[Asher and Burton 2006]*. Females have a tenfold the risk of curve progression requiring medical intervention than males *[Miller 1999]*.

Although that the pathogenesis of the disease is still unknown, AIS is observed aggregated within families of patients *[Riseborough and Wynne-Davies 1973; Wynne-Davies 1968]*, suggesting heritability. The scoliosis curves in monozygous twins are more likely to develop and progress together than in dizygous twins, demonstrating strong evidence for a genetic etiology in AIS *[Andersen, Thomsen and Kyvik 2007; Kesling and Reinker 1997]*. Based on families or populations, genetic studies have identified several candidate genes or loci in AIS etiology. However, these studies have a poor success rate in replication studies *[Gorman, Julien and Moreau 2012]*.

Since 2010, genome-wide association studies (GWAS) have been applied to AIS. This type of genomic survey presents a non-hypothesis based approach, by genotyping population-defined single nucleotide polymorphisms (SNPs). To date, only two case-control discovery studies have been published, one in Utah using a Caucasian population *[Sharma, Gao, Londono et al. 2011]*, the other in Japan with Japanese population *[Takahashi, Kou, Takahashi et al. 2011]*. Based on the genotyping outcome and subsequent replication studies, three loci: 3p26.3 *[Sharma, Gao, Londono et al. 2011]*, 10q24.31 *[Takahashi, Kou, Takahashi et al. 2011]* and 6q24.1 *[Kou, Takahashi, Johnson et al. 2013]*, have been associated with AIS predisposition. However, these significant variants have a minor effect size (odds ratio, OR<2), indicating that they are not major contributors to the disease etiology. Using the same genomic data in the Japanese population, one locus, 17q24.3, was significantly associated with AIS severity of medium effect (rs12946942, combined OR=2.2 in East-Asia population) *[Miyake, Kou, Takahashi et al. 2013]*. But their findings have not been validated in other ethnic groups.

Here with a French-Canadian cohort, we conducted a GWAS to identify genetic variants in AIS patients. The French-Canadian population was founded on migrants who moved from Europe in the 17th and 18th centuries. Compared to the rest of North America, it is relatively isolated, leading the Quebec Founder Population which is valuable to medical genetic research with low genetic variability *[De Braekeleer and Dao 1994]*. In addition, the incidence of AIS is higher in the population in Quebec than average *[Rogala, Drummond and Gurr 1978]*. A survey of our clinical database shows that among 920 patients, 467 (51%) reported a familial incidence of scoliosis, suggesting a strong genetic effect in our population.

Materials and methods

Study population and data source

This study has the approval from the institutional review boards of The Sainte-Justine University Hospital, The Montreal Children's Hospital, The Shriners Hospital for Children in Montreal and McGill University, as well as the Affluent and Montreal English School Boards. We recruited 1056 individuals from schools. Additional genomic control data of 750 individuals were from the CARTaGENE project *[Awadalla, Boileau, Payette et al. 2013; Godard, Marshall and Laberge 2007]*. To rule out the presence of scoliosis from the controls of the studied population, school screening was conducted by one of the orthopedic surgeons at Sainte-Justine Hospital in Montreal, Quebec, Canada, and the CARTaGENE adult phenotype records were checked.

Genome wide association study (GWAS)

Genomic DNA samples were derived from the peripheral blood of the subjects at the hospital and schools and then genotyped by the Illumina HumanOmni 2.5M BeadChip, which genotyped 2.5 million SNPs per sample. Control data from CARTaGENE was merged into the microarray outcome to generate files in the appropriate format to be analyzed.

Quality control (QC) measures were applied to genomic data following standards previously outlined *[Turner, Armstrong, Bradford et al. 2011; Weale 2010]*. We used PLINK *[Purcell, Neale, Todd-Brown et al. 2007]* and R *[Team 2012]* software packages to: 1) filter

gender mismatches, 2) filter missingness at both the sample-level (< 2%) and SNP-level (< 2%), 3) assess sample heterozygosity, 4) filter SNPs with a minor allele frequency (MAF) less than 1% and 5) filter SNPs in Hardy-Weinberg disequilibrium *[Neale and Purcell 2008; Samani, Erdmann, Hall et al. 2007]*.

Linkage disequilibrium (LD) thinning was performed on the filtered genomic data prior to ancestral and relatedness testing by applying respectively EIGENSTRAT *[Price, Patterson, Plenge et al. 2006]* and PLINK *[Purcell, Neale, Todd-Brown et al. 2007]* identity-by-descent (IBD). Ancestral outliers and related samples (pi_hat >0.1875) were thus 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, 170 individuals from schools and 731 individuals from the CARTaGENE project).

Definition of severe and non-severe cases

To identify genetic variants associated with the spinal curve severity, we classified AIS-affected patients into severe and non-severe groups, defined by the Cobb angles of major curve records. Severe patients were defined by a major curve Cobb angle ≥40º. Non-severe AIS patients were defined as their major curve Cobb angle between 10º and 39º by skeletal maturity. To simplify the concept "skeletal maturity", we fixed the cutoff age as 14 years for girls and 16 years for boys. The non-severe group consisted of skeletally mature patients only.

As a result of the criteria we define here, 148 patients were classified as severe and 302 patients were in a non-severe group. Clinical characteristics of two groups at the last clinic visit are shown in **Table XI.**

Statistical analyses

Using PLINK software, we evaluated the association of 1.4 million SNPs with AIS predisposition and severity in French-Canadian population by Chi-square test for allele model. To achieve statistical significance, we used a conservative Bonferroni correction to adjust the p-value depicting probable association *[Bush and Moore 2012]*. We adjusted the probability for the false positive results from 0.05 to (0.05/k) where k is the number of SNPs tested in each independent association test ($k=1.4\times10^6$ in our genome-wide test). Therefore, SNPs with

a p-value $<10^{-8}$ demonstrate a statistically significant association with AIS. SNPs with a p-value $<10^{-5}$ are treated as suggestively significant variants.

Statistical power calculation

The pwr package in R software (http://www.statmethods.net/stats/power.html) *[Champely and Champely 2007]* was used in statistical power calculations for each association study while effect size was defined as small, medium or large, as outlined by Cohen *[Cohen 1988]*.

Results

Power calculation for association analysis is listed in **Table XII**. We concluded that the statistical power for each association analysis is strong enough to detect a medium genetic effect.

Two SNPs reached genome-wide significance level through the GWAS association analysis (**Figure 5**). The most significant was SNP rs114646323 (p-value=1.34×10^{-9}) in the intron region of the KLC4 gene on chromosome 6. This variant was found in about 2% AIS-affected patients (14 out of 667 cases), but not in healthy control subjects. Neither was it in severe AIS cases. However, the association of this variant was significant to the non-severe cases (p-value=2.04×10^{-11}). About 2.6% patients in non-severe group had this variant (8 out of 302). This was the only variant significantly associated with non-severe cases when compared to control subjects.

Another significant SNP in GWAS, rs1607639, is located on chromosome 2 between genes GCFC2 and LRRTM4 (p-value=8.68×10^{-9}, odds ratio, OR=1.52, 95% Confidence Interval, CI=1.32-1.75). The association of this variant with severe cases and with non-severe cases reached a suggestive significance level (p-value=3.42×10^{-5} and 2.55×10^{-6}, respectively).

Through the association analysis between AIS severe cases and healthy control subjects, SNP rs201793089, located in the intron of gene CELF2 on chromosome 10, attained the genome-wide significance level (p-value=1.28×10^{-9}). The associations of this variant with non-severe cases and total AIS cases reached our genome-wide suggestive significance threshold (p-value=7.18×10^{-7} in case/control study and 4.05×10^{-8} in non-severe/control study).

None of the control individuals had this variant. This variant was detected in about 1.4% of total affected cases (9 out of 667), 2% of severe cases (3 out of 148) and 1.6% of non-severe cases (5 out of 302).

In summary, **Table XIII** listed the three significant SNPs and the candidate genes.

Discussion

We performed a genome-wide association study to find genomic variants linked to adolescent idiopathic scoliosis in French-Canadian population using the Illumina HumanOmni 2.5M BeadChip, a genotyping microarray with the highest density of SNPs used in the AIS research field so far. A genome-wide association study (GWAS) is an examination tool to detect genetic association with a trait in different individuals. By genotyping millions of single nucleotide polymorphisms (SNPs) through the entire genome and comparing allele frequencies between two different groups, researchers are capable of detecting genetic polymorphisms that have an association with one group. These genetic polymorphisms may be in linkage disequilibrium (LD) with causal genes that are located within or around the polymorphism loci. In general, the association study is a statistical approach. SNPs detected in GWAS are usually common in the population (minor allele frequency, MAF>5%) *[Bush and Moore 2012; Spencer, Su, Donnelly et al. 2009]*. Here with the highest density of the genotyping chip in AIS study, we were capable of detecting the variants more rare (MAF>1%). In addition, with a French-Canadian population, a valuable founder population in genetic research, we identified three variants statistically correlated to AIS disease.

Two SNPs, rs114646323 and rs1607639, were identified significantly associated with AIS disease in GWAS case/control analysis (P-value $<10^{-8}$) while SNP rs201793089 was found significantly associated with AIS-affected severe cases (P-value $<10^{-8}$). It is noteworthy that the variant rs114646323 demonstrated significant association in case/control study but was not observed in severe cases. A t-test (http://www.socscistatistics.com/tests/ztest/Default2.aspx) suggests this significant distinction between severe and non-severe groups (P-value=0.0455). Our hypothesis is that since the allele is associated with curve predisposition, it may have protective effects against curve

progression. However, the small cohort size being taken under consideration, the distinction may represent a possible statistical bias. Therefore, extensive genotyping among severe cases is necessary to investigate this trend further.

The variants identified in the association study suggest novel candidate genes that play a role in AIS susceptibility and/or severity pathogenesis in French-Canadian population. The KLC4 gene encodes the protein kinesin light chain 4 in humans. Kinesins, composed of two heavy chains and two light chains, are microtubule-based molecular motors that transport various intracellular cargos, including neurons and ciliated cells. Intracellular transport is important in the regulation of several physiological processes in mammals, including the development of the body axis, brain wiring and development, higher brain function, left–right body determination and tumor suppression *[Hirokawa, Noda, Tanaka et al. 2009]*. The gene LRRTM4 (leucine rich repeat transmembrane neuronal 4, OMIM# 610870) positively regulates excitatory synapse development in cultured neurons and in vivo, playing a role in regulation of synapse development and function *[de Wit, O'Sullivan, Savas et al. 2013]*. Variants in the GCFC2 gene (GC-rich sequence DNA-binding factor 2, OMIM# 189901) were found to have a genome-wide significant association with Alzheimer disease-related quantitative measures of hippocampal volume *[Melville, Buros, Parrado et al. 2012]*. A haplotype for this GC-rich sequence DNA-binding factor gene has been associated with dyslexia in a set of Finnish families *[Anthoni, Zucchelli, Matsson et al. 2007]*. The protein encoded by the gene CELF2 (CUGBP- and Elav-like family member 2, OMIM# 602538) is a member of CELF/BRUNOL protein family, which was implicated in the regulation of several post-transcriptional events. It mediates exon inclusion and/or exclusion in tissue-specific pre-mRNA alternative splicing, including cardiac and skeletal muscle, smooth muscle and neuronal cells *[Barreau, Paillard, Mereau et al. 2006]*.

The fact that these significantly associated variants are located in non-protein-coding regions, creates barrier to a forthcoming explanation of their biological functions. However, as assayed by ChIP-seq (Chromatin Immunoprecipitation sequencing) from the ENCODE (Encyclopedia of DNA Elements) project, the most significant variant rs114646323 is located in a putative binding site for transcription factor YY1 (Yin and yang 1, OMIM# 600013). Inquiring the sequence surrounding the variant rs114646323 (5'-CTGCC[A/G]TCTC-3'

where A is reference allele and G is alternative allele) on the ALGGEN (Algorithmics and Genetics Group) website (http://alggen.lsi.upc.es/), we found that the reference sequence showed the most similarity with the YY1 binding site while the alternative sequence did not match with any binding site of transcription factor in human beings *[Farre, Roset, Huerta et al. 2003; Messeguer, Escudero, Farre et al. 2002]*. Binding to the consensus sequence 5'-CCGCCATNTT-3', YY1 exhibits multiple controls (initiate, activate, or repress transcription) on a large number of genes by binding to sites overlapping the transcription start site. There is evidence that this transcription factor plays an important role in embryogenesis, differentiation, and cellular replication, proliferation, senescence, and response to genotoxic stimuli *[Gordon, Akopyan, Garban et al. 2006]*. Therefore, this variant, albeit its location in a non-coding region, is worthy of further investigation because of its putative functions in transcriptional regulation of other genes.

We surveyed the significant SNPs that were previously reported through the GWAS approach in AIS disease by other research groups *[Kou, Takahashi, Johnson et al. 2013; Miyake, Kou, Takahashi et al. 2013; Sharma, Gao, Londono et al. 2011; Takahashi, Kou, Takahashi et al. 2011]*. However, they did not attain the genome-wide suggestive significance level as we found in French-Canadian population (p-value>10^{-5}). The irreproducibility of the association could mainly come from population stratification. Population in Quebec shows relatively low genetic variation due to the founder effect.

Genetic heterogeneity is another possible barrier contributing to AIS genetic and phenotypic complexity. On one hand, individuals carrying a same variant may demonstrate various clinical manifestations. On the other hand, the same disorder may be triggered by variants in different genes due to the involvement in the same or related biological pathways *[McClellan and King 2010]*. Genetic heterogeneity could be a reason that significant variants were associated with a small fraction of a particular population.

From this GWAS in French-Canadian population, we demonstrated three genetic variants significantly associated with spinal curve susceptibility and/or progress. It should keep in mind that our findings might be overestimated as impacted by the phenomenon of "winner's curse" *[Xiao and Boehnke 2009]*. To prevent this upward bias in our GWAS discovery, it is ideal to validate our findings, for the next step, in a larger cohort which has an exact match with the original French-Canadian cohort both in genetics and environmental

background. Again, an independent replication study is also required to verify our significant variants in other ethnic population with a large sample size to confirm the GWAS result.

Acknowledgements

We are indebted to the subjects and families who participated in this study. We also thank all the orthopedic surgeons and nursing teams who have participated in this study at Sainte-Justine University Hospital, The Montreal's Children Hospital, and The Shriners Hospital for Children in Montreal.

This work was supported by grants from La Fondation Yves Cotrel de l'Institut de France, Paris, France and Génome Québec. No other potential conflict of interest relevant to this article was reported.

Tables and figures

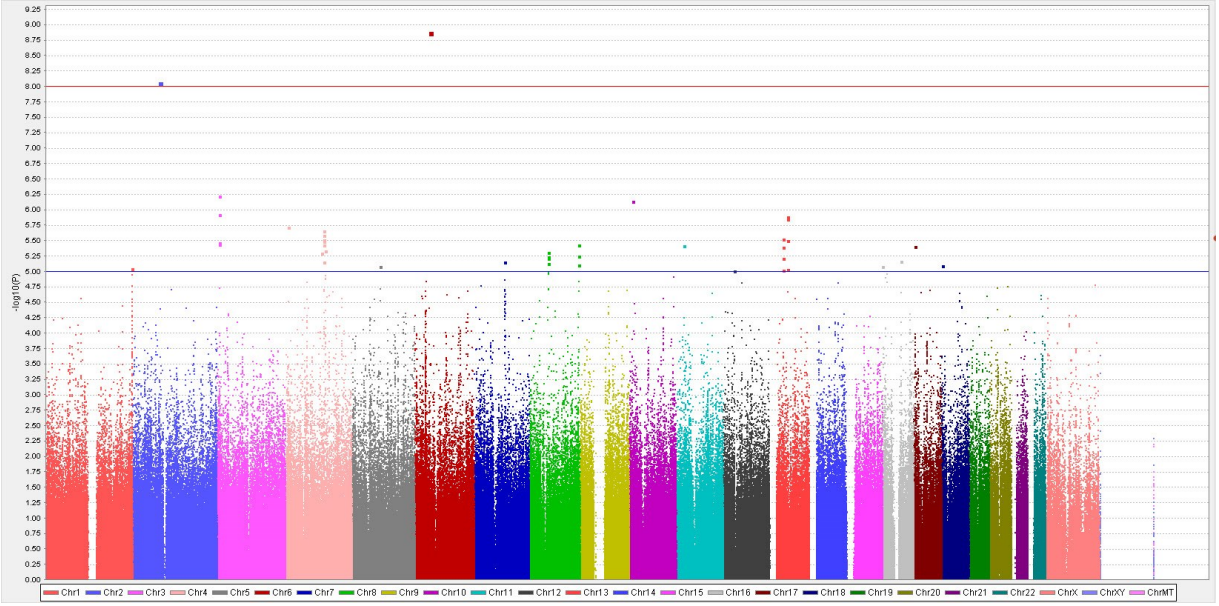

Figure 5. Manhattan plot showing the P values from genome-wide association study. The horizontal lines represent the genome-wide significant threshold (P-value $=10^{-8}$) and suggestive significant threshold (P-value $=10^{-5}$).

Table XI. Demographic and clinical characteristics of severe patients and non-severe patients with AIS at the last visit.

Subjects	Characteristics								
	All Subjects			Female			Male		
	N	Mean Age (Years)	Scoliosis Cobb Angles (°)	N	Mean Age (Years)	Scoliosis Cobb Angles (°)	N	Mean Age (Years)	Scoliosis Cobb Angles (°)
Severe AIS patients	148	15 ± 2 (10 – 25)	56 ± 12 (37 – 90)	129	15 ± 2 (10 – 25)	55 ± 12 (37 – 90)	19	16 ± 2 (12 – 19)	60 ± 11 (40 – 87)
Non-severe AIS patients	302	16 ± 1 (14 – 22)	21 ± 8 (3 – 39)	259	16 ± 1 (14 – 22)	21 ± 9 (3 – 39)	43	17 ± 1 (16 – 19)	19 ± 7 (7 – 35)

All values represent mean Cobb Angles ± standard deviation, and range values for respective groups.

Severe AIS patient was defined as Cobb angle ≥40º for major spinal curves.

Non-severe AIS patient was defined as the highest historical record of Cobb angle between 10º and 39º for spinal curves by skeletal maturity (girls ≥ 14 years and boys ≥ 16 years).

Table XII. Statistical power calculations for each association study in R software.

	Power (%)		
	Case vs. Control	Severe case vs. Control	Non-severe case vs. Control
	667 vs. 901	148 vs. 901	302 vs. 901
Effect size [a]			
small	74.8	47.9	12.1
medium	100	100	99.9
large	100	100	100

[a] Cohen proposed rules of thumb for interpreting effect sizes: a "small" effect size is 20%, a "medium" effect size is 50%, and a "large" effect size is 80%.

Table XIII. Significant SNPs and candidate genes identified by GWAS approach in French-Canadian population.

dbSNP	CHR	P value [a]			candidate gene
		case vs. control	severe vs. control	non-severe vs. control	
rs114646323	6	**1.34×10^{-9}**	NaN	**2.04×10^{-11}**	KLC4
rs1607639	2	**8.68×10^{-9}**	3.42×10^{-5}	2.55×10^{-6}	GCFC2, LRRTM4
rs201793089	10	7.18×10^{-7}	**1.28×10^{-9}**	4.05×10^{-8}	CELF2

[a] calculated by Chi-square test.
P values below the genome-wide significance level (P-value $<10^{-8}$) are in bold.

References

1. Riseborough EJ, Wynne-Davies R. A genetic survey of idiopathic scoliosis in Boston, Massachusetts. *The Journal of bone and joint surgery. American volume* 1973;55:974-82.
2. Asher MA, Burton DC. Adolescent idiopathic scoliosis: natural history and long term treatment effects. *Scoliosis* 2006;1:2.
3. Miller NH. Cause and natural history of adolescent idiopathic scoliosis. *The Orthopedic clinics of North America* 1999;30:343-52, vii.
4. Wynne-Davies R. Familial (idiopathic) scoliosis. A family survey. *The Journal of bone and joint surgery. British volume* 1968;50:24-30.
5. Andersen MO, Thomsen K, Kyvik KO. Adolescent idiopathic scoliosis in twins: a population-based survey. *Spine* 2007;32:927-30.
6. Kesling KL, Reinker KA. Scoliosis in twins. A meta-analysis of the literature and report of six cases. *Spine* 1997;22:2009-14; discussion 15.
7. Gorman KF, Julien C, Moreau A. The genetic epidemiology of idiopathic scoliosis. *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society* 2012;21:1905-19.
8. 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.
9. 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.
10. Kou I, Takahashi Y, Johnson TA, et al. Genetic variants in GPR126 are associated with adolescent idiopathic scoliosis. *Nature genetics* 2013;45:676-9.
11. 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.
12. De Braekeleer M, Dao TN. Hereditary disorders in the French Canadian population of Quebec. I. In search of founders. *Human biology* 1994;66:205-23.
13. Rogala EJ, Drummond DS, Gurr J. Scoliosis: incidence and natural history. A prospective epidemiological study. *The Journal of bone and joint surgery. American volume* 1978;60:173-6.
14. Godard B, Marshall J, Laberge C. Community engagement in genetic research: results of the first public consultation for the Quebec CARTaGENE project. *Community genetics* 2007;10:147-58.
15. 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* 2013;42:1285-99.
16. Turner S, Armstrong LL, Bradford Y, et al. Quality control procedures for genome-wide association studies. *Current protocols in human genetics / editorial board, Jonathan L. Haines ... [et al.]* 2011;Chapter 1:Unit1 19.
17. Weale ME. Quality control for genome-wide association studies. *Methods in molecular biology* 2010;628:341-72.

18. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* 2007;81:559-75.
19. Team RC. R: A language and environment for statistical computing. 2012.
20. Samani NJ, Erdmann J, Hall AS, et al. Genomewide association analysis of coronary artery disease. *The New England journal of medicine* 2007;357:443-53.
21. Neale BM, Purcell S. The positives, protocols, and perils of genome-wide association. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* 2008;147B:1288-94.
22. 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.
23. Bush WS, Moore JH. Chapter 11: Genome-wide association studies. *PLoS computational biology* 2012;8:e1002822.
24. Champely S, Champely MS. The pwr Package. 2007.
25. Cohen J. Statistical power analysis for the behavioral sciences: New York: Academic Press. y= 0.2981 ln (x), 1988.
26. Spencer CC, Su Z, Donnelly P, et al. Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. *PLoS genetics* 2009;5:e1000477.
27. Hirokawa N, Noda Y, Tanaka Y, et al. Kinesin superfamily motor proteins and intracellular transport. *Nature reviews. Molecular cell biology* 2009;10:682-96.
28. de Wit J, O'Sullivan ML, Savas JN, et al. Unbiased discovery of glypican as a receptor for LRRTM4 in regulating excitatory synapse development. *Neuron* 2013;79:696-711.
29. Melville SA, Buros J, Parrado AR, et al. Multiple loci influencing hippocampal degeneration identified by genome scan. *Annals of neurology* 2012;72:65-75.
30. Anthoni H, Zucchelli M, Matsson H, et al. A locus on 2p12 containing the co-regulated MRPL19 and C2ORF3 genes is associated to dyslexia. *Human molecular genetics* 2007;16:667-77.
31. Barreau C, Paillard L, Mereau A, et al. Mammalian CELF/Bruno-like RNA-binding proteins: molecular characteristics and biological functions. *Biochimie* 2006;88:515-25.
32. Messeguer X, Escudero R, Farre D, et al. PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics* 2002;18:333-4.
33. Farre D, Roset R, Huerta M, et al. Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. *Nucleic acids research* 2003;31:3651-3.
34. Gordon S, Akopyan G, Garban H, et al. Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. *Oncogene* 2006;25:1125-42.
35. McClellan J, King MC. Genetic heterogeneity in human disease. *Cell* 2010;141:210-7.
36. Xiao R, Boehnke M. Quantifying and correcting for the winner's curse in genetic association studies. *Genetic epidemiology* 2009;33:453-62.

CHAPTER 4. DISCUSSION

Adolescent idiopathic scoliosis (AIS) is the most common form of pediatric spinal malformation with unknown cause. There is strong evidence of genetic factors' contribution to the pathogenesis of AIS. It is a complex common disease with polygenic inheritance *[Kesling and Reinker 1997; Ward, Ogilvie, Argyle et al. 2010]*. The hypothesis suggests that AIS is a consequence of a moderate to large number of common genetic variants, each of which contributes to several percent of the risk for curvature and/or progression. Genome-wide association study (GWAS) is designed to test this so-called common disease-common variant hypothesis (CDCV hypothesis) *[Bush and Moore 2012; McCarthy, Abecasis, Cardon et al. 2008]*.

With the GWAS approach, we test common SNPs across the entire genome in thousands of individuals without the necessity of any biological knowledge. By comparing SNPs' frequencies between patients and controls, we may find out associated SNPs and identify genomic regions of interest thanks to the linkage disequilibrium (LD) method. However, common SNPs are often located in introns and intergenic regions with unclear functions, making it difficult to comprehend these SNPs' roles in human diseases' pathogenesis. When evaluating the effectiveness of a GWAS, it is important to take under consideration multiple factors, such as the sample sizes, the odds ratios, the allele frequencies, the threshold of significance, and the performance of the commercial microarrays in a population *[Hong and Park 2012; Jorgenson and Witte 2006; Korte and Farlow 2013; Magi, Pfeufer, Nelis et al. 2007; McCarthy, Abecasis, Cardon et al. 2008; Riancho 2012; Stranger, Stahl and Raj 2011]*. Even with such limitation, GWAS has become more and more prevalent in genetic researches of human disease *[Wellcome Trust Case Control 2007]*. There were a total of 689 GWA studies in 2012, and 860 studies in 2013 (http://hugenavigator.net). This is likely due to the decreasing cost and improved power for the technologies. The recent generation of commercially available chips also has improved the genomic coverage and the representation of alleles that occur at a minor frequency in the population. For example, variants with minor allele frequencies (MAFs) of greater than 5% are used to be defined as common and be tested in a GWAS. The microarray, we used in our GWAS approach, covers genomic variants with MAF as low as 1%.

To identify genomic variants associated with AIS disease, we performed a genome-wide association study (GWAS) in French-Canadian population, by genotyping over 1.4

million SNPs among 667 AIS patients and 901 healthy controls. Using a microarray with a very high density of single nucleotide polymorphisms (SNPs), our study has the capacity to detect more genomic variants with more rare frequency (minor allele frequency, MAF>1% by quality control of genotyping data), representing the most comprehensive genomic survey done yet in AIS research field.

In the first article, we evaluated 53 SNPs that were previously associated with spinal curve progression in an American population of European descent to determine whether there is a similar association in a Caucasian French-Canadian population. As demonstrated by an earlier study in a Japanese cohort, we did not find any significant association to AIS initiation or to curve severity, in a French-Canadian cohort, suggesting that the lack of replication in the Japanese population is not due to ethnicity.

The irreproducibility may come from ethnic differences (also termed as population stratification), over-estimate of the original GWAS findings because of the "winner's curse", or uncertain statistical power that was affected by various determinants. Commercially available genotyping chips show diverse performances in a same ethnic group, as well as among distinct ethnic groups. The disagreement of important parameters among original association studies, such as the criteria of quality control in genotyping data and the ascertainment of phenotypes to survey, sets obstacle to evaluate the statistical power of original studies and to reproduce the original signals of association in replication studies.

In the second article, we identified genome-wide significant SNPs linked to spinal curve predisposition and/or severity through a GWAS in French-Canadian population. So far, the fact that these significantly associated variants are located in non-protein-coding regions, creates barrier to a forthcoming explanation of their biological functions. However, thanks to the linkage disequilibrium (LD) between the associated variants and the causal genes, the significant variants suggested novel candidate genes involved in the incidence and/or progression of the spinal curvature.

The KLC4 gene (kinesin light chain 4) encodes a composition of kinesin which is a molecular motor of neurons and ciliated cells in intracellular transport for the development of body axis and brain wiring and development. LRRTM4 (leucine rich repeat transmembrane

neuronal 4, OMIM#610870) has a function in synapse development. A GC-rich sequence DNA-binding factor gene (GCFC2, OMIM#189901) has an association with dyslexia and Alzheimer disease. The protein encoded by the gene CELF2 (CUGBP- and Elav-like family member 2, OMIM# 602538) mediates exon inclusion/exclusion in cardiac and skeletal muscle, smooth muscle and neuronal cells. Functional study of these candidate genes may highlight the pathogenesis of AIS disease.

In addition, it is noteworthy that the most significant variant rs114646323 is located in a putative binding site for transcription factor YY1 (Yin and yang 1, OMIM# 600013), which has multiple functions in transcriptional regulation of a large number of genes involved in basic cellular functions. Thus, this variant calls for further investigation, albeit its location in a non-coding region.

4.1 Future work in GWAS approach

Population-based association study through whole genome is a statistical approach. Sampling is an important determinant to generate a true association signal. The small cohort recruited in the discovery phase of association study may not have enough power to detect variants with small to medium effect sizes. Meanwhile, current commercially available genotyping technologies have the possibility to introduce a range of errors and biases in GWAS analysis. Therefore, the next step for our GWAS approach is a replication study, using a second genotyping platform to genotype significant variants in another independent French-Canadian cohort, e.g., 1000 AIS cases vs. 1000 healthy controls. It allows early validation of false positive association signals coming from technical errors and validates our original findings in larger samples. Functional studies of the candidate genes in animal models are also important to validate gene functions in the pathogenesis of AIS disease for further clinical applications and drug innovations *[Manolio 2013]*.

GWAS is an effective approach on the basis of the hypothesis “common disease-common variant (CDCV)”. Since most of the common variants are located in non-coding region, we missed most of the variants in protein-coding region, where exome sequencing is effective.

4.2 Other genetic hypotheses and relative approaches in AIS study

Despite the notable success of GWAS in revealing numerous new disease-associated genes and loci, all the identified SNPs collectively account for a small proportion of the heritability for common complex diseases. This has led to the important consideration on the reasons for "missing heritability" *[Chaufan and Joseph 2013; Eichler, Flint, Gibson et al. 2010; Manolio, Collins, Cox et al. 2009]*. Rare variants and genetic interactions are likely to play an important role in complex diseases *[Zuk, Hechter, Sunyaev et al. 2012; Zuk, Schaffner, Samocha et al. 2014]*, but neither has yet been well examined in GWAS. Studies of gene-gene and gene-environment interactions are still a challenge for researchers *[Ackert-Bicknell and Karasik 2013; Jiao, Hsu, Berndt et al. 2012; Okser, Pahikkala and Aittokallio 2013]*. Rare variants, assumed in "common disease-rare variant" (CDRV) hypothesis, is likely to be the major contributors to genetic susceptibility to complex disease, each with relative major effect *[Gibson 2011; Schork, Murray, Frazer et al. 2009]*. Besides CDRV hypothesis, changes in gene expression because of epigenetic modification may modulate the phenotype in complex diseases *[Feinberg 2010; Schumacher and Petronis 2006]*. In addition, Moreau *et al.* have demonstrated the existence of functional sub-groups among AIS patients, suggesting a genetic heterogeneity in AIS *[Akoume, Azeddine, Turgeon et al. 2010; McClellan and King 2010]*.

4.2.1 Common disease-rare variant hypothesis and whole exome sequencing

Rare variants, the allele frequency of which is typically <1%, are thought to exist as recently derived highly penetrant alleles that account for high disease susceptibility *[Gibson 2011]*. The disease could occur from an accumulation of these rare variants in a functional class or network. Currently, tools such as whole exome sequencing is available for the investigation of rare variants *[Bamshad, Ng, Bigham et al. 2011]*. Earlier investigations have suggested promising results associated with complex disease like AIS *[Buchan, Alvarado, Haller et al. 2014; Christodoulou, Wiskin, Gibson et al. 2013; de Ligt, Veltman and Vissers 2013]*.

Rare variants can occur as point mutations, or as gene deletions/duplications. To enrich for harmful alleles in each gene, whole exome sequencing typically focuses on non-synonymous variants in protein-coding region (missense, nonsense, gain/loss of start/stop codon, splice site/frameshift change) with a low population frequency (typically <1%). Generally, comparing sequences of diseased individuals with a healthy control cohort and/or with existing databases like 1000 Genomes Project (www.1000genomes.org) allows to identify genes in which there is an elevated aggregation of rare variants *[Li and Leal 2008; Morris and Zeggini 2010]*.

When applied to a large pedigree, exome sequencing has the potential to identify family specific causative genes that might explain some of the cases in the population *[Peng, Fan, Palculict et al. 2013]*. When applied to a population, exome sequencing can suggest important genes and pathways that are accumulated with rare variants in cases versus controls *[Moens, De Rijk, Reumers et al. 2011]*. It is possible to validate or predict whether a variant/gene is likely to have damaging effects via in vitro biochemical experiments or computational programs *[Romeo, Yin, Kozlitina et al. 2009; Sunyaev 2012]*. Studying gene sets that are aggregated in genetic loci identified by GWAS is a potentially powerful strategy, because genes linked to a trait are likely to harbor both common and rare variants *[Rivas, Beaudoin, Gardet et al. 2011; Teslovich, Musunuru, Smith et al. 2010]*.

The limitation with exome sequencing is that these rare variants are studied in the protein-coding region, which accounts for only 1% of the genome. Large sample collections are required for both common and rare variants studies. CDCV and CDRV hypotheses are complementary to each other.

4.2.2 Gene expression studies on epigenetic modifications

Gene expression can be influenced by genome modifications other than variations in the DNA sequence itself. Unlike DNA sequence changes, genome functional changes, such as DNA methylation and histone modification, can regulate how genes are expressed without altering the underlying DNA sequence. It is thought that such epigenetic modifications influence the observed phenotypic variability for complex diseases *[Feinberg 2010; Schumacher and Petronis 2006]*. With a concordance rate less than 100% among monozygotic twins, who have nearly identical genetic information, it is feasible to presume that epigenetic modification may modulate the phenotype with or without genetic variations.

A real-time polymerase chain reaction (also known as quantitative PCR, qPCR) *[Bustin, Benes, Garson et al. 2009]* has been employed in quantification of gene expression. Quantitative PCR technology is more rapid, cost-effective, easier to use, and capable of higher throughput in molecular biology applications *[VanGuilder, Vrana and Freeman 2008]*. However, for those diseases such as AIS, multiple tissues are affected. The choice of tissue to be explored and the timing of sample collection could have an important impact on the outcome of gene expression inquiries.

Micro ribonucleic acids (miRNAs) can regulate gene expression in a tissue-specific way. Micro RNA is a class of short (22 nucleotides) noncoding RNA, which targets messenger RNAs (mRNAs) in a sequence-specific manner. The human genome may encode over 1000 miRNAs, which may regulate about 60% of human protein coding genes *[Friedman, Farh, Burge et al. 2009]*. Most recently, the Affymetrix Company introduced a commercially available high-density miRNA Target Site Genotyping Arrays. Announced by the Encyclopedia of DNA Elements (ENCODE) Consortium, miRNAs contain important regulatory elements with functional importance. We expect that studies using miRNA arrays could elucidate important biological pathways involved in AIS pathogenesis *[Xiao, Diao, Yang et al. 2013]*.

4.2.3 Functional group classification among AIS patients

Recent work on the biological basis of AIS by Moreau and colleagues has demonstrated the existence of functional groups among patients *[Akoume, Azeddine, Turgeon et al. 2010; Letellier, Azeddine, Blain et al. 2007; Moreau, Wang, Forget et al. 2004]*. Based on their experimental data, melatonin signaling dysfunction was found only in AIS patients and not in healthy controls. Using osteoblasts and peripheral blood mononuclear cells (PBMCs), they have validated the signaling impairment by functional in vitro assays. Moreover, depending on the cellular response to melatonin, they suggested a classification of AIS patients in three different functional groups. There is a hypothesis that diverse variants/genes and their related signaling pathways are implicated in each group. The application of functional group classification of AIS studies in large cohorts of patients and controls could reduce genetic heterogeneity and increase the chances of detecting more genomic variants associated with each subgroup.

CHAPTER 5. CONCLUSION

The main goal of this project was to identify genomic variants significantly associated with idiopathic scoliosis among adolescents in French-Canadian population. Firstly, using genomic data from a genome-wide association study, we verified that previously reported association in ScoliScore^{TM} was not found in a French-Canadian cohort. This second replication cohort suggested that the lack of association of these variants in a Japanese cohort was not due to ethnicity. Secondly, through the typical case/control GWA study, we identified several genomic variants which were significantly associated with spinal curve initiation and/or progression in French-Canadian population. Our results suggested novel candidate genes that may play a role in AIS pathogenesis. The observed non-association of previously reported genomic loci suggests possible bias from the population stratification and/or the genetic heterogeneity involved in AIS etiopathogenesis.

With comprehensive genetic studies in complex human diseases such as AIS, we expect that in the next few years there will be more breakthroughs on genomic variants associated with diseases. Our findings are worthy of further investigation and could make contribution to clinical applications. Disease-associated genomic variants like those we singled out are valuable in early detection of high-risk individuals, patients' classification for improved medical care and drug innovation for disease prevention.

CHAPTER 6. REFERENCES

Ackert-Bicknell, C. L., and D. Karasik. "*Impact of the Environment on the Skeleton: Is It Modulated by Genetic Factors?*" *Curr Osteoporos Rep* 11.3 (2013): 219-28.

Akoume, M. Y., B. Azeddine, I. Turgeon*, et al.* "*Cell-Based Screening Test for Idiopathic Scoliosis Using Cellular Dielectric Spectroscopy*." *Spine (Phila Pa 1976)* 35.13 (2010): E601-8.

Alexander, M. A., and E. H. Season. "*Idiopathic Scoliosis: An Electromyographic Study*." *Arch Phys Med Rehabil* 59.7 (1978): 314-5.

Andersen, M. O., K. Thomsen, and K. O. Kyvik. "*Adolescent Idiopathic Scoliosis in Twins: A Population-Based Survey*." *Spine (Phila Pa 1976)* 32.8 (2007): 927-30.

Anthoni, H., M. Zucchelli, H. Matsson*, et al.* "*A Locus on 2p12 Containing the Co-Regulated Mrpl19 and C2orf3 Genes Is Associated to Dyslexia*." *Hum Mol Genet* 16.6 (2007): 667-77.

Asher, M. A., and D. C. Burton. "*Adolescent Idiopathic Scoliosis: Natural History and Long Term Treatment Effects*." *Scoliosis* 1.1 (2006): 2.

Awadalla, P., C. Boileau, Y. Payette*, et al.* "*Cohort Profile of the Cartagene Study: Quebec's Population-Based Biobank for Public Health and Personalized Genomics*." *Int J Epidemiol* 42.5 (2013): 1285-99.

Bamshad, M. J., S. B. Ng, A. W. Bigham*, et al.* "*Exome Sequencing as a Tool for Mendelian Disease Gene Discovery*." *Nat Rev Genet* 12.11 (2011): 745-55.

Barreau, C., L. Paillard, A. Mereau*, et al.* "*Mammalian Celf/Bruno-Like Rna-Binding Proteins: Molecular Characteristics and Biological Functions*." *Biochimie* 88.5 (2006): 515-25.

Beaulieu, M., C. Toulotte, L. Gatto*, et al.* "*Postural Imbalance in Non-Treated Adolescent Idiopathic Scoliosis at Different Periods of Progression*." *Eur Spine J* 18.1 (2009): 38-44.

Buchan, J. G., D. M. Alvarado, G. E. Haller*, et al.* "*Rare Variants in Fbn1 and Fbn2 Are Associated with Severe Adolescent Idiopathic Scoliosis*." *Hum Mol Genet* 23.19 (2014): 5271-82.

Burwell, R. G., and P. H. Dangerfield. "*Whither the Etiopathogenesis (and Scoliogeny) of Adolescent Idiopathic Scoliosis?*" *Stud Health Technol Inform* 176 (2012): 3-19.

Burwell, R. G., P. H. Dangerfield, A. Moulton*, et al.* "*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* 6.1 (2011): 26.

Bush, W. S., and J. H. Moore. "*Chapter 11: Genome-Wide Association Studies*." *PLoS Comput Biol* 8.12 (2012): e1002822.

Bustin, S. A., V. Benes, J. A. Garson*, et al.* "*The Miqe Guidelines: Minimum Information for Publication of Quantitative Real-Time Pcr Experiments*." *Clin Chem* 55.4 (2009): 611-22.

Champely, Stéphane, and Maintainer Stéphane Champely. "*The Pwr Package*." (2007).

Chaufan, C., and J. Joseph. "*The 'Missing Heritability' of Common Disorders: Should Health Researchers Care?*" *Int J Health Serv* 43.2 (2013): 281-303.

Cheng, J. C., X. Guo, A. H. Sher*, et al.* "*Correlation between Curve Severity, Somatosensory Evoked Potentials, and Magnetic Resonance Imaging in Adolescent Idiopathic Scoliosis*." *Spine (Phila Pa 1976)* 24.16 (1999): 1679-84.

Cheung, J., J. P. Halbertsma, A. G. Veldhuizen*, et al.* "*A Preliminary Study on Electromyographic Analysis of the Paraspinal Musculature in Idiopathic Scoliosis*." *Eur Spine J* 14.2 (2005): 130-7.

Christodoulou, K., A. E. Wiskin, J. Gibson*, et al.* "*Next Generation Exome Sequencing of Paediatric Inflammatory Bowel Disease Patients Identifies Rare and Novel Variants in Candidate Genes*." *Gut* 62.7 (2013): 977-84.

Chu, W. C., W. W. Lam, Y. L. Chan*, 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 (Phila Pa 1976)* 31.1 (2006): E19-25.

Chu, W. C., G. C. Man, W. W. Lam*, et al.* "*Morphological and Functional Electrophysiological Evidence of Relative Spinal Cord Tethering in Adolescent Idiopathic Scoliosis*." *Spine (Phila Pa 1976)* 33.6 (2008): 673-80.

Cohen, J. "*Statistical Power Analysis for the Behavioral Sciences*." New York: Academic Press. y= 0.2981 ln (x), 1988.

Cote, P., B. G. Kreitz, J. D. Cassidy*, et al.* "*A Study of the Diagnostic Accuracy and Reliability of the Scoliometer and Adam's Forward Bend Test*." *Spine (Phila Pa 1976)* 23.7 (1998): 796-802; discussion 03.

Daffner, S. D., C. F. Beimesch, and J. C. Wang. "*Geographic and Demographic Variability of Cost and Surgical Treatment of Idiopathic Scoliosis*." *Spine (Phila Pa 1976)* 35.11 (2010): 1165-9.

Dawn Teare, M., and J. H. Barrett. "*Genetic Linkage Studies*." *Lancet* 366.9490 (2005): 1036-44.

Dayer, R., T. Haumont, W. Belaieff*, et al.* "*Idiopathic Scoliosis: Etiological Concepts and Hypotheses*." *J Child Orthop* 7.1 (2013): 11-6.

De Braekeleer, M., and T. N. Dao. "*Hereditary Disorders in the French Canadian Population of Quebec. I. In Search of Founders*." *Hum Biol* 66.2 (1994): 205-23.

de Ligt, J., J. A. Veltman, and L. E. Vissers. "*Point Mutations as a Source of De Novo Genetic Disease*." *Curr Opin Genet Dev* 23.3 (2013): 257-63.

de Wit, J., M. L. O'Sullivan, J. N. Savas*, et al.* "*Unbiased Discovery of Glypican as a Receptor for Lrrtm4 in Regulating Excitatory Synapse Development*." *Neuron* 79.4 (2013): 696-711.

Dietz, H. C., B. Loeys, L. Carta*, et al.* "*Recent Progress Towards a Molecular Understanding of Marfan Syndrome*." *Am J Med Genet C Semin Med Genet* 139C.1 (2005): 4-9.

Dobbs, M. B., and C. A. Gurnett. "*Re: Ward K, Ogilvie Jw, Singleton Mv, Et Al. Validation of DNA-Based Prognostic Testing to Predict Spinal Curve Progression in Adolescent Idiopathic Scoliosis. Spine 2010;35:E1455-64*." *Spine (Phila Pa 1976)* 36.15 (2011): 1257; author reply 57.

Dobbs, M. B., and S. L. Weinstein. "*Infantile and Juvenile Scoliosis*." *Orthop Clin North Am* 30.3 (1999): 331-41, vii.

Dormans, John P., Struan F. Grant, Norma Rendon Sampson*, et al.* "*A Genome Wide Association Study Identifies Il17rc as an Adolescent Idiopathic Scoliosis Locus: Paper #80†*." *Spine Journal Meeting Abstracts* (2011): 96.

Eichler, E. E., J. Flint, G. Gibson*, et al.* "*Missing Heritability and Strategies for Finding the Underlying Causes of Complex Disease*." *Nat Rev Genet* 11.6 (2010): 446-50.

Fan, Y. H., Y. Q. Song, D. Chan*, et al.* "*Snp Rs11190870 near Lbx1 Is Associated with Adolescent Idiopathic Scoliosis in Southern Chinese*." *J Hum Genet* 57.4 (2012): 244-6.

Farre, D., R. Roset, M. Huerta*, et al.* "*Identification of Patterns in Biological Sequences at the Alggen Server: Promo and Malgen*." *Nucleic Acids Res* 31.13 (2003): 3651-3.

Feinberg, A. P. "*Genome-Scale Approaches to the Epigenetics of Common Human Disease*." *Virchows Arch* 456.1 (2010): 13-21.

Friedman, R. C., K. K. Farh, C. B. Burge*, et al.* "*Most Mammalian Mrnas Are Conserved Targets of Microrna*s." *Genome Res* 19.1 (2009): 92-105.

Gao, W., Y. Peng, G. Liang*, et al.* "*Association between Common Variants near Lbx1 and Adolescent Idiopathic Scoliosis Replicated in the Chinese Han Population*." *PLoS One* 8.1 (2013): e53234.

Geoffrey Burwell, R., and P. H. Dangerfield. "*Platelet Calmodulin Levels in Adolescent Idiopathic Scoliosis. Do the Levels Correlate with Curve Progression and Severity?*" *Spine (Phila Pa 1976)* 28.17 (2003): 2036-7; author reply 37-8.

Gibson, G. "*Rare and Common Variants: Twenty Arguments*." *Nat Rev Genet* 13.2 (2011): 135-45.

Girardo, M., N. Bettini, E. Dema*, et al.* "*The Role of Melatonin in the Pathogenesis of Adolescent Idiopathic Scoliosis (Ais)*." *Eur Spine J* 20 Suppl 1 (2011): S68-74.

Godard, B., J. Marshall, and C. Laberge. "*Community Engagement in Genetic Research: Results of the First Public Consultation for the Quebec Cartagene Project*." *Community Genet* 10.3 (2007): 147-58.

Gordon, S., G. Akopyan, H. Garban*, et al.* "*Transcription Factor Yy1: Structure, Function, and Therapeutic Implications in Cancer Biology*." *Oncogene* 25.8 (2006): 1125-42.

Gorman, K. F., C. Julien, and A. Moreau. "*The Genetic Epidemiology of Idiopathic Scoliosis*." *Eur Spine J* 21.10 (2012): 1905-19.

Gorman, Kristen F, Cedric Julien, Niaz Oliazadeh*, et al.* "*Genetics of Idiopathic Scoliosis*." *eLS* (2014).

Grant, S. F., and J. P. Dormans. "*Re: Ward K, Ogilvie Jw, Singleton, Et Al. Validation of DNA-Based Prognostic Testing to Predict Spinal Curve Progression in Adolescent Idiopathic Scoliosis. Spine 2010;35:E1455-64*." *Spine (Phila Pa 1976)* 36.15 (2011): 1258; discussion 58-9.

Guo, X., W. W. Chau, C. W. Hui-Chan*, et al.* "*Balance Control in Adolescents with Idiopathic Scoliosis and Disturbed Somatosensory Function*." *Spine (Phila Pa 1976)* 31.14 (2006): E437-40.

Gupta, P. A., E. A. Putnam, S. G. Carmical*, et al.* "*Ten Novel Fbn2 Mutations in Congenital Contractural Arachnodactyly: Delineation of the Molecular Pathogenesis and Clinical Phenotype*." *Hum Mutat* 19.1 (2002): 39-48.

Hadley-Miller, N., B. Mims, and D. M. Milewicz. "*The Potential Role of the Elastic Fiber System in Adolescent Idiopathic Scoliosis*." *J Bone Joint Surg Am* 76.8 (1994): 1193-206.

Hirokawa, N., Y. Noda, Y. Tanaka*, et al.* "*Kinesin Superfamily Motor Proteins and Intracellular Transport*." *Nat Rev Mol Cell Biol* 10.10 (2009): 682-96.

Hirschhorn, J. N., K. Lohmueller, E. Byrne*, et al.* "*A Comprehensive Review of Genetic Association Studies*." *Genet Med* 4.2 (2002): 45-61.

Hong, E. P., and J. W. Park. "*Sample Size and Statistical Power Calculation in Genetic Association Studies*." *Genomics Inform* 10.2 (2012): 117-22.

James, J. I. "*Idiopathic Scoliosis; the Prognosis, Diagnosis, and Operative Indications Related to Curve Patterns and the Age at Onset*." *J Bone Joint Surg Br* 36-B.1 (1954): 36-49.

Jiang, H., X. Qiu, J. Dai*, et al.* "*Association of Rs11190870 near Lbx1 with Adolescent Idiopathic Scoliosis Susceptibility in a Han Chinese Population*." *Eur Spine J* 22.2 (2013): 282-6.

Jiao, S., L. Hsu, S. Berndt*, et al.* "*Genome-Wide Search for Gene-Gene Interactions in Colorectal Cancer*." *PLoS One* 7.12 (2012): e52535.

Johnson, A. D., R. E. Handsaker, S. L. Pulit*, et al.* "*Snap: A Web-Based Tool for Identification and Annotation of Proxy Snps Using Hapmap*." *Bioinformatics* 24.24 (2008): 2938-9.

Jorgenson, E., and J. S. Witte. "*Coverage and Power in Genomewide Association Studies*." *Am J Hum Genet* 78.5 (2006): 884-8.

Kainulainen, K., L. Karttunen, L. Puhakka*, et al.* "*Mutations in the Fibrillin Gene Responsible for Dominant Ectopia Lentis and Neonatal Marfan Syndrome*." *Nat Genet* 6.1 (1994): 64-9.

Kane, W. J. "*Scoliosis Prevalence: A Call for a Statement of Terms*." *Clin Orthop Relat Res*.126 (1977): 43-6.

Kesling, K. L., and K. A. Reinker. "*Scoliosis in Twins. A Meta-Analysis of the Literature and Report of Six Cases*." *Spine (Phila Pa 1976)* 22.17 (1997): 2009-14; discussion 15.

Kindsfater, K., T. Lowe, D. Lawellin*, et al.* "*Levels of Platelet Calmodulin for the Prediction of Progression and Severity of Adolescent Idiopathic Scoliosis*." *J Bone Joint Surg Am* 76.8 (1994): 1186-92.

Korte, A., and A. Farlow. "*The Advantages and Limitations of Trait Analysis with Gwas: A Review*." *Plant Methods* 9 (2013): 29.

Kotwicki, T., J. Chowanska, E. Kinel*, et al.* "*Optimal Management of Idiopathic Scoliosis in Adolescence*." *Adolesc Health Med Ther* 4 (2013): 59-73.

Kou, I., Y. Takahashi, T. A. Johnson*, et al.* "*Genetic Variants in Gpr126 Are Associated with Adolescent Idiopathic Scoliosis*." *Nat Genet* 45.6 (2013): 676-9.

Kouwenhoven, J. W., and R. M. Castelein. "*The Pathogenesis of Adolescent Idiopathic Scoliosis: Review of the Literature*." *Spine (Phila Pa 1976)* 33.26 (2008): 2898-908.

Letellier, K., B. Azeddine, S. Blain*, et al.* "*[Etiopathogenesis of Adolescent Idiopathic Scoliosis and New Molecular Concepts]*." *Med Sci (Paris)* 23.11 (2007): 910-6.

Li, B., and S. M. Leal. "*Methods for Detecting Associations with Rare Variants for Common Diseases: Application to Analysis of Sequence Data*." *Am J Hum Genet* 83.3 (2008): 311-21.

Liang, J., D. Xing, Z. Li*, et al.* "*Association between Rs11190870 Polymorphism near Lbx1 and Susceptibility to Adolescent Idiopathic Scoliosis in East Asian Population: A Genetic Meta-Analysis*." *Spine (Phila Pa 1976)* (2014).

Liu, T., W. C. Chu, G. Young*, et al.* "*Mr Analysis of Regional Brain Volume in Adolescent Idiopathic Scoliosis: Neurological Manifestation of a Systemic Disease*." *J Magn Reson Imaging* 27.4 (2008): 732-6.

Lonstein, J. E. "*Adolescent Idiopathic Scoliosis*." *Lancet* 344.8934 (1994): 1407-12.

Lowe, T. G., R. G. Burwell, and P. H. Dangerfield. "*Platelet Calmodulin Levels in Adolescent Idiopathic Scoliosis (Ais): Can They Predict Curve Progression and Severity? Summary of an Electronic Focus Group Debate of the Ibse*." *Eur Spine J* 13.3 (2004): 257-65.

Lowe, T., D. Lawellin, D. Smith*, et al.* "*Platelet Calmodulin Levels in Adolescent Idiopathic Scoliosis: Do the Levels Correlate with Curve Progression and Severity?*" *Spine (Phila Pa 1976)* 27.7 (2002): 768-75.

Machida, M., J. Dubousset, Y. Imamura*, et al.* "*Role of Melatonin Deficiency in the Development of Scoliosis in Pinealectomised Chickens*." *J Bone Joint Surg Br* 77.1 (1995): 134-8.

Machida, M., J. Dubousset, T. Yamada*, et al.* "*Serum Melatonin Levels in Adolescent Idiopathic Scoliosis Prediction and Prevention for Curve Progression--a Prospective Study*." *J Pineal Res* 46.3 (2009): 344-8.

Machida, M., I. Murai, Y. Miyashita*, et al.* "*Pathogenesis of Idiopathic Scoliosis. Experimental Study in Rats*." *Spine (Phila Pa 1976)* 24.19 (1999): 1985-9.

Magi, R., A. Pfeufer, M. Nelis*, et al.* "*Evaluating the Performance of Commercial Whole-Genome Marker Sets for Capturing Common Genetic Variation*." *BMC Genomics* 8 (2007): 159.

Manolio, T. A. "*Bringing Genome-Wide Association Findings into Clinical Use*." *Nat Rev Genet* 14.8 (2013): 549-58.

Manolio, T. A., F. S. Collins, N. J. Cox*, et al.* "*Finding the Missing Heritability of Complex Diseases*." *Nature* 461.7265 (2009): 747-53.

McCarthy, M. I., G. R. Abecasis, L. R. Cardon*, et al.* "*Genome-Wide Association Studies for Complex Traits: Consensus, Uncertainty and Challenges*." *Nat Rev Genet* 9.5 (2008): 356-69.

McClellan, J., and M. C. King. "*Genetic Heterogeneity in Human Disease*." *Cell* 141.2 (2010): 210-7.

Melville, S. A., J. Buros, A. R. Parrado*, et al.* "*Multiple Loci Influencing Hippocampal Degeneration Identified by Genome Scan*." *Ann Neurol* 72.1 (2012): 65-75.

Messeguer, X., R. Escudero, D. Farre*, et al.* "*Promo: Detection of Known Transcription Regulatory Elements Using Species-Tailored Searches*." *Bioinformatics* 18.2 (2002): 333-4.

Miller, N. H. "*Cause and Natural History of Adolescent Idiopathic Scoliosis*." *Orthop Clin North Am* 30.3 (1999): 343-52, vii.

Miyake, A., I. Kou, Y. Takahashi*, et al.* "*Identification of a Susceptibility Locus for Severe Adolescent Idiopathic Scoliosis on Chromosome 17q24.3*." *PLoS One* 8.9 (2013): e72802.

Moens, L. N., P. De Rijk, J. Reumers*, et al.* "*Sequencing of Disc1 Pathway Genes Reveals Increased Burden of Rare Missense Variants in Schizophrenia Patients from a Northern Swedish Population*." *PLoS One* 6.8 (2011): e23450.

Moreau, A., D. S. Wang, S. Forget*, et al.* "*Melatonin Signaling Dysfunction in Adolescent Idiopathic Scoliosis*." *Spine (Phila Pa 1976)* 29.16 (2004): 1772-81.

Morris, A. P., and E. Zeggini. "*An Evaluation of Statistical Approaches to Rare Variant Analysis in Genetic Association Studies*." *Genet Epidemiol* 34.2 (2010): 188-93.

Murray, D. W., and C. J. Bulstrode. "*The Development of Adolescent Idiopathic Scoliosis*." *Eur Spine J* 5.4 (1996): 251-7.

Neale, B. M., and S. Purcell. "*The Positives, Protocols, and Perils of Genome-Wide Association*." *Am J Med Genet B Neuropsychiatr Genet* 147B.7 (2008): 1288-94.

Nelson, Lesa M., Rakesh Chettier, James W. Ogilvie*, et al.* "*Candidate Genes for Susceptibility of Adolescent Idiopathic Scoliosis Identified through a Large Genome - Wide Association Study: Paper #81†*." *Spine Journal Meeting Abstracts* (2011): 96-97.

Ogura, Y., Y. Takahashi, I. Kou*, et al.* "*A Replication Study for Association of 53 Single Nucleotide Polymorphisms in a Scoliosis Prognostic Test with Progression of Adolescent Idiopathic Scoliosis in Japanese*." *Spine (Phila Pa 1976)* 38.16 (2013): 1375-9.

Okser, S., T. Pahikkala, and T. Aittokallio. "*Genetic Variants and Their Interactions in Disease Risk Prediction - Machine Learning and Network Perspectives*." *BioData Min* 6.1 (2013): 5.

Peng, G., Y. Fan, T. B. Palculict*, et al.* "*Rare Variant Detection Using Family-Based Sequencing Analysis*." *Proc Natl Acad Sci U S A* 110.10 (2013): 3985-90.

Petersen, I., T. Sahlstrand, and U. Sellden. "*Electroencephalographic Investigation of Patients with Adolescent Idiopathic Scoliosis*." *Acta Orthop Scand* 50.3 (1979): 283-93.

Porter, R. W. "*Can a Short Spinal Cord Produce Scoliosis?*" *Eur Spine J* 10.1 (2001a): 2-9.

Porter, R. W. "*Idiopathic Scoliosis: The Relation between the Vertebral Canal and the Vertebral Bodies*." *Spine (Phila Pa 1976)* 25.11 (2000): 1360-6.

Porter, R. W. "*The Pathogenesis of Idiopathic Scoliosis: Uncoupled Neuro-Osseous Growth?*" *Eur Spine J* 10.6 (2001b): 473-81.

Price, A. L., N. J. Patterson, R. M. Plenge*, et al.* "*Principal Components Analysis Corrects for Stratification in Genome-Wide Association Studies*." *Nat Genet* 38.8 (2006): 904-9.

Purcell, S., B. Neale, K. Todd-Brown*, et al.* "*Plink: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses*." *Am J Hum Genet* 81.3 (2007): 559-75.

Putnam, E. A., H. Zhang, F. Ramirez*, et al.* "*Fibrillin-2 (Fbn2) Mutations Result in the Marfan-Like Disorder, Congenital Contractural Arachnodactyly*." *Nat Genet* 11.4 (1995): 456-8.

Qiu, X. S., F. Lv, Z. Z. Zhu*, et al.* "*Lack of Association between the Chl1 Gene and Adolescent Idiopathic Scoliosis Susceptibility in Han Chinese: A Case-Control Study*." *BMC Musculoskelet Disord* 15 (2014): 38.

Riancho, J. A. "*Genome-Wide Association Studies (Gwas) in Complex Diseases: Advantages and Limitations*." *Reumatol Clin* 8.2 (2012): 56-7.

Riseborough, E. J., and R. Wynne-Davies. "*A Genetic Survey of Idiopathic Scoliosis in Boston, Massachusetts*." *J Bone Joint Surg Am* 55.5 (1973): 974-82.

Rivas, M. A., M. Beaudoin, A. Gardet*, et al.* "*Deep Resequencing of Gwas Loci Identifies Independent Rare Variants Associated with Inflammatory Bowel Disease*." *Nat Genet* 43.11 (2011): 1066-73.

Roach, J. W. "*Adolescent Idiopathic Scoliosis*." *Orthop Clin North Am* 30.3 (1999): 353-65, vii-viii.

Rogala, E. J., D. S. Drummond, and J. Gurr. "*Scoliosis: Incidence and Natural History. A Prospective Epidemiological Study*." *J Bone Joint Surg Am* 60.2 (1978): 173-6.

Romeo, S., W. Yin, J. Kozlitina*, et al.* "*Rare Loss-of-Function Mutations in Angptl Family Members Contribute to Plasma Triglyceride Levels in Humans*." *J Clin Invest* 119.1 (2009): 70-9.

Roye, B. D., M. L. Wright, B. A. Williams*, et al.* "*Does Scoliscore Provide More Information Than Traditional Clinical Estimates of Curve Progression?*" *Spine (Phila Pa 1976)* 37.25 (2012): 2099-103.

Sahlstrand, T., and B. Petruson. "*A Study of Labyrinthine Function in Patients with Adolescent Idiopathic Scoliosis. I. An Electro-Nystagmographic Study*." *Acta Orthop Scand* 50.6 Pt 2 (1979): 759-69.

Sahlstrand, T., B. Petruson, and R. Ortengren. "*Vestibulospinal Reflex Activity in Patients with Adolescent Idiopathic Scoliosis. Postural Effects During Caloric Labyrinthine Stimulation Recorded by Stabilometry*." *Acta Orthop Scand* 50.3 (1979): 275-81.

Samani, N. J., J. Erdmann, A. S. Hall*, et al.* "*Genomewide Association Analysis of Coronary Artery Disease*." *N Engl J Med* 357.5 (2007): 443-53.

Schork, N. J., S. S. Murray, K. A. Frazer*, et al.* "*Common Vs. Rare Allele Hypotheses for Complex Diseases*." *Curr Opin Genet Dev* 19.3 (2009): 212-9.

Schumacher, A., and A. Petronis. "*Epigenetics of Complex Diseases: From General Theory to Laboratory Experiments*." *Curr Top Microbiol Immunol* 310 (2006): 81-115.

Sham, P. C., and S. M. Purcell. "*Statistical Power and Significance Testing in Large-Scale Genetic Studies*." *Nat Rev Genet* 15.5 (2014): 335-46.

Sharma, S., X. Gao, D. Londono*, et al.* "*Genome-Wide Association Studies of Adolescent Idiopathic Scoliosis Suggest Candidate Susceptibility Genes*." *Hum Mol Genet* 20.7 (2011): 1456-66.

Simoneau, M., N. Richer, P. Mercier*, et al.* "*Sensory Deprivation and Balance Control in Idiopathic Scoliosis Adolescent*." *Exp Brain Res* 170.4 (2006): 576-82.

Spencer, C. C., Z. Su, P. Donnelly*, et al.* "*Designing Genome-Wide Association Studies: Sample Size, Power, Imputation, and the Choice of Genotyping Chip*." *PLoS Genet* 5.5 (2009): e1000477.

Sponseller, P. D., W. Hobbs, L. H. Riley, 3rd*, et al.* "*The Thoracolumbar Spine in Marfan Syndrome*." *J Bone Joint Surg Am* 77.6 (1995): 867-76.

Stranger, B. E., E. A. Stahl, and T. Raj. "*Progress and Promise of Genome-Wide Association Studies for Human Complex Trait Genetics*." *Genetics* 187.2 (2011): 367-83.

Sunyaev, S. R. "*Inferring Causality and Functional Significance of Human Coding DNA Variants*." *Hum Mol Genet* 21.R1 (2012): R10-7.

Takahashi, Y., I. Kou, A. Takahashi*, et al.* "*A Genome-Wide Association Study Identifies Common Variants near Lbx1 Associated with Adolescent Idiopathic Scoliosis*." *Nat Genet* 43.12 (2011): 1237-40.

Team, R Core. "*R: A Language and Environment for Statistical Computing*." (2012).

Teslovich, T. M., K. Musunuru, A. V. Smith*, et al.* "*Biological, Clinical and Population Relevance of 95 Loci for Blood Lipids*." *Nature* 466.7307 (2010): 707-13.

Thillard, M. J. "*[Vertebral Column Deformities Following Epiphysectomy in the Chick]*." *C R Hebd Seances Acad Sci* 248.8 (1959): 1238-40.

Turner, S., L. L. Armstrong, Y. Bradford*, et al.* "*Quality Control Procedures for Genome-Wide Association Studies*." *Curr Protoc Hum Genet* Chapter 1 (2011): Unit1 19.

VanGuilder, H. D., K. E. Vrana, and W. M. Freeman. "*Twenty-Five Years of Quantitative Pcr for Gene Expression Analysis*." *Biotechniques* 44.5 (2008): 619-26.

Wang, W. J., H. Y. Yeung, W. C. Chu*, et al.* "*Top Theories for the Etiopathogenesis of Adolescent Idiopathic Scoliosis*." *J Pediatr Orthop* 31.1 Suppl (2011): S14-27.

Ward, K., J. Ogilvie, V. Argyle*, et al.* "*Polygenic Inheritance of Adolescent Idiopathic Scoliosis: A Study of Extended Families in Utah*." *Am J Med Genet A* 152A.5 (2010): 1178-88.

Ward, K., J. W. Ogilvie, M. V. Singleton*, et al.* "*Validation of DNA-Based Prognostic Testing to Predict Spinal Curve Progression in Adolescent Idiopathic Scoliosis*." *Spine (Phila Pa 1976)* 35.25 (2010): E1455-64.

Weale, M. E. "*Quality Control for Genome-Wide Association Studies*." *Methods Mol Biol* 628 (2010): 341-72.

Weinstein, S. L., L. A. Dolan, J. C. Cheng*, et al.* "*Adolescent Idiopathic Scoliosis*." *Lancet* 371.9623 (2008): 1527-37.

Weinstein, S. L., L. A. Dolan, J. G. Wright*, et al.* "*Effects of Bracing in Adolescents with Idiopathic Scoliosis*." *N Engl J Med* 369.16 (2013): 1512-21.

Wellcome Trust Case Control, Consortium. "*Genome-Wide Association Study of 14,000 Cases of Seven Common Diseases and 3,000 Shared Controls*." *Nature* 447.7145 (2007): 661-78.

Wray, N. R. "*Allele Frequencies and the R2 Measure of Linkage Disequilibrium: Impact on Design and Interpretation of Association Studies*." *Twin Res Hum Genet* 8.2 (2005): 87-94.

Wynne-Davies, R. "*Familial (Idiopathic) Scoliosis. A Family Survey*." *J Bone Joint Surg Br* 50.1 (1968): 24-30.

Xiao, R., and M. Boehnke. "*Quantifying and Correcting for the Winner's Curse in Genetic Association Studies*." *Genet Epidemiol* 33.5 (2009): 453-62.

Xiao, Z. D., L. T. Diao, J. H. Yang*, et al.* "*Deciphering the Transcriptional Regulation of Microrna Genes in Humans with Actlocater*." *Nucleic Acids Res* 41.1 (2013): e5.

Yagi, Mitsuru, Masafumi Machida, and Takashi Asazuma. "*Pathogenesis of Adolescent Idiopathic Scoliosis*." *JBJS Reviews* 2.1 (2014).

Yim, A. P., H. Y. Yeung, G. Sun*, et al.* "*Abnormal Skeletal Growth in Adolescent Idiopathic Scoliosis Is Associated with Abnormal Quantitative Expression of Melatonin Receptor, Mt2*." *Int J Mol Sci* 14.3 (2013): 6345-58.

Zetterberg, C., R. Bjork, R. Ortengren*, et al.* "*Electromyography of the Paravertebral Muscles in Idiopathic Scoliosis. Measurements of Amplitude and Spectral Changes under Load*." *Acta Orthop Scand* 55.3 (1984): 304-9.

Zhou, S., Z. Zhu, X. Qiu*, et al.* "*Association Study of Il-17rc, Chl1, Dscam and Cntnap2 Genes Polymorphisms with Adolescent Idiopathic Scoliosis Susceptibility in a Chinese Han Population*." *Stud Health Technol Inform* 176 (2012): 47-51.

Zuk, O., E. Hechter, S. R. Sunyaev*, et al.* "*The Mystery of Missing Heritability: Genetic Interactions Create Phantom Heritability*." *Proc Natl Acad Sci U S A* 109.4 (2012): 1193-8.

Zuk, O., S. F. Schaffner, K. Samocha*, et al.* "*Searching for Missing Heritability: Designing Rare Variant Association Studies*." *Proc Natl Acad Sci U S A* 111.4 (2014): E455-64.