(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 December 2014 (31.12.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/207246 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/EP2014/063780

(22) International Filing Date:
27 June 2014 (27.06.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
13305897.4 27 June 2013 (27.06.2013) EP

(71) Applicants: **INSERM (INSTITUT NATIONAL DE LA SANTÉ ET DE LA RECHERCHE MÉDICALE)** [FR/FR]; 101, rue de Tolbiac, F-75013 Paris (FR). **UNIVERSITÉ CLAUDE BERNARD - LYON 1** [FR/FR]; 43 boulevard du 11 novembre 1918, F-69100 Villeurbanne (FR). **CHU SAINTE JUSTINE** [CA/CA]; 3175 Côte-Ste-Catherine, Montreal, Québec H3T 1C5 (CA). **CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS)** [FR/FR]; 3, rue Michel Ange, F-75016 Paris (FR). **HOSPICES CIVILS DE LYON** [FR/FR]; 3 Quai des Célestins, F-69002 Lyon (FR).

(72) Inventors: **EDERY, Patrick**; Centre de Recherche en Neurosciences de Lyon, INSERM1028 - CNRS UMR5292, 95 Boulevard Pinel, CH LE VINATIER/ Bat 452, F-69675 Bron Cedex (FR). **MOLDOVAN, Florina**; CHU Sainte-Justine, Centre de Recherche 3175, Chemin de la Côte Sainte-Catherine, Montréal, Québec H3T 1C5 (CA). **PATTEN, Shunmoogum**; CHU Sainte-Justine, Centre de Recherche 3175, Chemin de la Côte Sainte-Catherine, Montréal, Québec H3T 1C5 (CA).

(74) Agent: **HIRSCH, Denise**; Inserm Transfert, 7 rue Watt, F-75013 Paris (FR).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2014/207246 A1

(54) Title: NEW POLYMORPHISMS FOR THE DIAGNOSIS OF IDIOPATHIC SCOLIOSIS DISEASE

(57) Abstract: The present invention relates to a method of identifying a subject having or at risk of having or developing idiopathic scoliosis disease, comprising determining, in a sample obtained from said subject, the presence or absence of an allelic variant in a single nucleotide polymorphism (SNP) located at *POC5* gene.

NEW POLYMORPHISMS FOR THE DIAGNOSIS OF IDIOPATHIC SCOLIOSIS DISEASE

RELATED APPLICATION

The present application claims priority to European Patent Application No. EP13305897.4 which was filed on June 27, 2013. The European patent application is incorporated herein by reference in its entirety.

FIELD OF THE INVENTION:

The invention is in the field of scoliosis disease/condition diagnosis and therapy. In particular, the invention relates to specific single nucleotide polymorphisms (SNPs) in the human genome and their association with idiopathic scoliosis disease.

BACKGROUND OF THE INVENTION:

Idiopathic scoliosis (IS) is a common spinal deformity characterised by rotation of the vertebral bodies. It affects ~3% of the population at age 16. IS is defined[1] by a structurally fixed lateral curvature of the spine with a Cobb's angle of at least 10°. This disease occurs in otherwise healthy adolescents (~85%) or in younger children (~15%), and is more frequent in girls than in boys[2,3]. About 3 per 1000 children present severe forms, with Cobb's angles greater than 30°. These forms may require braces or surgical intervention and can have severe functional, aesthetic and social consequences[1-3]. Multiple aetiopathogenic hypotheses have been suggested, including neurological mechanisms[4-6], muscular[7,8], mechanical[9], extracellular matrix[10] or metabolic disorders[11,12], but the aetiology of IS remains unknown despite extensive studies. Familial aggregation of IS cases was first observed many decades ago[13-17], and familial occurrence of the disease is seen in ~40% of cases[15,18]. However, genome-wide linkage studies have provided very few sound results, favouring the view that idiopathic scoliosis is a genetically heterogeneous disorder[19-24]. Genome-wide association studies (GWAS) recently reported candidate loci for IS susceptibility[25,26]; however, efforts to find the causative genes have thus far been unsuccessful. Identifying IS-causing genes should improve our understanding of the pathogenesis of this common paediatric disorder, and trigger advances in diagnostic testing.

Identifying additional risk genes and understanding their mechanisms of action through experimental and functional studies represent current challenges and are critical for the use of this knowledge in disease prevention or therapy.

Nonetheless, up to this point, no single biomarker is sufficiently specific to provide adequate clinical utility for the diagnosis of IS in an individual patient.

Therefore, there is a need for identifying factors that provide a more accurate diagnosis/prognosis of IS.

Here, the inventors describe the first ever identification of an IS-causing gene, present in ~10% of IS families. Screening for this gene should improve the clinical outcome for young members of these families by allowing early detection and appropriate clinical management of IS from its earliest stages.

SUMMARY OF THE INVENTION:

A first object of the invention is a method of identifying a subject having or at risk of having or developing a idiopathic scoliosis disease, comprising determining, in a sample obtained from said subject, the presence or absence of an allelic variant in a single nucleotide polymorphism (SNP) located in *POC5* nucleic sequence.

In preferred embodiment the SNP is located in exon 10 of the *POC5* nucleic sequence.

In a particular embodiment, the SNP is selected from the group consisting of rs34678567, rs146984380; c.G1363C and wherein :

- the presence of the allele (A) of SNP rs34678567 indicates an increased risk of having or being at risk of having or developing an idiopathic scoliosis disease, and
- the presence of the allele (T) of SNP rs146984380 indicates a increased risk of having or being at risk of having or developing an idiopathic scoliosis disease;
- the presence of the allele (C) of SNP c.G1363C indicates a increased risk of having or being at risk of having or developing an idiopathic scoliosis disease.

A second object of the invention is a kit for identifying whether a subject has or is at risk of having or developing a idiopathic scoliosis, comprising :

- at least a means for detecting the SNP selected from the group consisting of rs34678567, rs146984380; c.G1363C and

- instructions for us.

A third object of the invention is a nuclease for use in treating Idiopathic Scoliosis and/or preventing progression of Idiopathic Scoliosis in a patient, wherein the presence of SNPs in exon 10 of the *POC5* nucleic sequence in a sample previously obtained from said patient, have been detected by a method of the invention previously described.

DETAILED DESCRIPTION OF THE INVENTION:

Definitions:

Throughout the specification, several terms are employed and are defined in the following paragraphs.

The term "Scoliosis" means a. is a medical condition in which a subject's spine is curved from side to side. Although it is a complex three-dimensional deformity, on an X-ray, viewed from the rear, the spine of an individual with scoliosis may look more like an "S" or a "C", rather than a straight line. IS is defined[1] by a structurally fixed lateral curvature of the spine with a Cobb's angle of at least 10°.

Scoliosis is typically classified as either congenital (caused by vertebral anomalies present at birth), idiopathic (cause unknown, subclassified as infantile, juvenile, adolescent, or adult, according to when onset occurred), or secondary to another primary condition. According to the method of the invention Scoliosis is Idiopathic Scoliosis or "IS".

The term "POC5" also known as "POC5 Centriolar Protein Homolog (Chlamydomonas)" and " C5orf37" means POC5 protein (NM_001099271 NP_001099271) which is essential for the assembly of the distal half of centrioles, required for centriole elongation. Belongs to the POC family. 3 isoforms of the human protein are produced by alternative splicing. The whole sequence of human *POC5* gene is referenced as Gene ID: 134359.

"Risk" in the context of the present invention, relates to the probability that an event will occur over a specific time period, as in the conversion to Scoliosis disease, and can mean a subject's "absolute" risk or "relative" risk. Absolute risk can be measured with reference to either actual observation post-measurement for the relevant time cohort, or with reference to

index values developed from statistically valid historical cohorts that have been followed for the relevant time period. Relative risk refers to the ratio of absolute risks of a subject compared either to the absolute risks of low risk cohorts or an average population risk, which can vary by how clinical risk factors are assessed. Odds ratios, the proportion of positive events to negative events for a given test result, are also commonly used (odds are according to the formula p/(l-p) where p is the probability of event and (1- p) is the probability of no event) to no conversion. Alternative continuous measures which may be assessed in the context of the present invention include time to Scoliosis disease conversion and therapeutic Scoliosis disease conversion risk reduction ratios.

"Risk evaluation," or "evaluation of risk" in the context of the present invention encompasses making a prediction of the probability, odds, or likelihood that an event or disease state may occur, the rate of occurrence of the event or conversion from one disease state to another, i.e., from a normal condition to a Scoliosis condition or to one at risk of developing a Scoliosis disease. Risk evaluation can also comprise prediction of future clinical parameters, traditional laboratory risk factor values, or other indices of Scoliosis disease, such as dopamine level detection, cellular population determination in peripheral tissues, in serum or other fluid (i.e. cerebrospinal fluid), either in absolute or relative terms in reference to a previously measured population. The methods of the present invention may be used to make continuous or categorical measurements of the risk of conversion to Scoliosis disease, thus diagnosing and defining the risk spectrum of a category of subjects defined as being at risk for a Scoliosis disease. In the categorical scenario, the invention can be used to discriminate between normal and other subject cohorts at higher risk for Scoliosis disease. In other embodiments, the present invention may be used so as to help to discriminate those having Scoliosis disease from normal.

"Clinical parameters or indicia" encompasses all non-sample or non-analyte biomarkers of subject health status or other characteristics, such as, without limitation, age (Age), geographical origin (Origin), gender (Sex), family history (FamHX), height (HT), weight (WT), waist (Waist) and body-mass index (BMI), as well as others such as clinical cardinal signs of Scoliosis disease (like the subject's spine is curved from side to side).

A "sample" in the context of the present invention is a biological sample isolated from a subject and can include, by way of example and not limitation, bodily fluids and/or tissue extracts such as homogenates or solubilized tissue obtained from a subject. Tissue extracts are obtained routinely from tissue biopsy and autopsy material. Bodily fluids useful in the present invention include blood, urine, saliva or any other bodily secretion or derivative thereof. In a

preferred embodiment, the sample to be tested is saliva or blood. As used herein "blood" includes whole blood, plasma, serum, circulating epithelial cells, constituents, or any derivative of blood. In a preferred embodiment the sample is a blood sample

According to the invention, POC5 polymorphism is a genomic polymorphism and is detected by using any type of body cell. In a preferred embodiment the cell is a blood cell.

According to the invention, the sample comprises exon 10 of *POC5* nucleic acid, wherein exon 10 of *POC5* nucleic acid is genomic DNA.

A "subject" in the context of the present invention is preferably a human.

The term "Allele" has the meaning which is commonly known in the art, that is, an alternative form of a gene (one member of a pair) that is located at a specific position on a specific chromosome which, when translated results in functional or dysfunctional (including non-existent) gene products.

The term "polymorphism" or "allelic variant" means a common sequence variation of a gene. Allelic variants can be found in the exons, introns, untranslated regions of the gene, or in the sequences that control expression of the gene. Complete gene sequencing often identifies numerous allelic variants (sometimes hundreds) for a given gene. The significance of allelic variants is often unclear until further study of the genotype and corresponding phenotype occurs in a sufficiently large population.

The term "Single nucleotide polymorphism" or "SNP" refers to a type of DNA polymorphism involving variation of a single base pair. There are millions of SNPs in the human genome. Most commonly, these variations are found in coding sequences of genes, non-coding regions of genes, or in intergenic regions between genes. When SNPs occur within a gene or in a regulatory region near a gene, they may play a more direct role in disease by affecting the gene's function.

The term "haplotype" refers to a set of alleles of closely linked loci on a chromosome that tend to be inherited together (means a 5' to 3' sequence of nucleotides found at a set of one or more polymorphic sites in a locus on a single chromosome from a single individual).

The SNPs pertaining to the invention (except NM_001099271:c.G1363C which was unknown) are known per se and sequences of them are publicly available from the data base http://www.ncbi.nlm.nih.gov/SNP/ or http://hapmap.ncbi.nlm.nih.gov/.

The SNPs studied at the 5q13.3 locus are described here after:

rs ID	HGVS name	Chr	Chromosome position	Gene	Transcription Orientation	Ac Am Change
rs34678567	NM_001099271:c.G1336A	5	5q13.3	POC5	-	A446T (exon 10)
Rs146984380	NM_001099271:c.C1286T	5	5q13.3	POC5	-	A429V (exon 10)
novel variant	NM_001099271:c.G1363C	5	5q13.3	POC5	-	A455P (exon 10)
rs190991771	NM_001099271:c.A673G	5	5q13.3	POC5	-	I225V (exon 6)
rs200926172	NM_001099271:c.G442A	5	5q13.3	POC5	-	D148N (exon 5)
novel variant	GRCh37/hg19 : 75013289 C>A	5	5q13.3	POC5	NA	5'UTR

Diagnostic method:

The present inventors have assayed for a statistical association between specific polymorphisms located at exon 10 (and at exon 5 and 6 and in 5'UTR region) of *POC5* nucleic acid and Idiopathic Scoliosis disease (SI) using a cohort of Scoliosis disease families, patients and controls. More precisely, the present inventors have assayed for a statistical association between specific polymorphisms contained in chromosome 5q13.3 locus of Scoliosis patients (multiplex families and cases) and controls.

By SNP genotyping and whole exome sequencing in IS patients belonging to a large family and subsequent Sanger sequencing in additional IS families, IS patients and controls, the inventors have found that specific SNPs associated with haplotype markers contained in the sequence of the 5q13.3 locus, more particularly in exon 10 of *POC5* nucleic acid cause idiopathic scoliosis in the corresponding patients.

As disclosed in the examples herein, the inventors have screened DNA blood samples of a well characterized cohort of IS families and patients and of controls to assess the genomic effects of single nucleotide polymorphisms (SNPs) at different loci. Evidence that IS is caused by 5q13.3 locus mutations in exon 10 (and at exon 5 and 6 and in 5'UTR region) of *POC5* nucleic acid in several families and patients have been provided.

More precisely, the inventors have now identified specific SNP biallelics marker located exon 10 of *POC5* nucleic acid, wherein the SNP biallelic marker selected from the group consisting of rs34678567, rs146984380, c.G1363C, was not only associated with IS but cause IS.

Indeed, the inventors have also performed functional analysis of POC5 mutations in zebrafish, and demonstrated spine deformities similar to those observed in patients and revealed brain-restricted expression during early development.

This results suggests that these POC5 mutations at exon 10, in particular the recurrent A446T POC5 mutation (SNP rs34678567) which is found in up to 10% of familial IS cases in this cohort, is a strong predictor of IS disease occurrence during adolescence. Furthermore, POC5 mutations at exon 10 could be a therapeutic target in young subjects.

Furthermore, the inventors have also identified specific SNP biallelics marker located exon 5 and 6 and in 5'UTR region of *POC5* nucleic acid, wherein the SNP biallelic marker selected from the group consisting of rs190991771, rs200926172, GRCh37/hg19 : 75013289 C>A was also associated with IS.

A first object of the invention is a method of identifying a subject having or at risk of having or developing a Scoliosis disease, comprising determining, in a sample obtained from said subject, the presence or absence of an allelic variant of single nucleotide polymorphism (SNP) located at *POC5* nucleic acid.

In preferred embodiment the SNP is located in exon 10 of the *POC5* nucleic sequence

In particular embodiment the SNP is selected from the group consisting of rs34678567, rs146984380, c.G1363C and wherein

- the presence of the allele (A) of SNP rs34678567 indicates an increased risk of having or being at risk of having or developing an idiopathic scoliosis disease, and
- the presence of the allele (T) of SNP rs146984380 indicates an increased risk of having or being at risk of having or developing an idiopathic scoliosis disease

- the presence of the allele (C) of SNP c.G1363C indicates an increased risk of having or being at risk of having or developing an idiopathic scoliosis disease.

According to the invention, the presence of the allele (A) of SNP rs34678567 indicates an increased risk of having or being at risk of having or developing a Scoliosis disease. As shown in the examples, allele T of SNP rs34678567 significantly increases the risk of IS compared with controls.

According to the invention, the presence of the allele (T) of SNP rs146984380 indicates an increased risk of having or being at risk of having or developing a Scoliosis disease. As shown in the examples, allele T of SNP rs146984380 significantly increases the risk of IS compared with controls.

According to the invention, the presence of the allele (C) of SNP c.G1363C indicates an increased risk of having or being at risk of having or developing a Scoliosis disease. As shown in the examples, allele C of SNP c.G1363C significantly increases the risk of IS compared with controls.

A further object of the invention is a method of identifying a subject having or at risk of having or developing a Scoliosis disease, comprising determining, in a sample obtained from said subject, the presence or absence of an allelic variant of single nucleotide polymorphism (SNP) located at exon 5 and 6 and in 5'UTR of *POC5* nucleic acid

In particular embodiment the SNP is selected from the group consisting of, rs190991771, rs200926172, GRCh37/hg19 : 75013289 C>A and wherein

the presence of the allele (G) of SNP rs190991771 indicates an increased risk of having or being at risk of having or developing an idiopathic scoliosis disease,

the presence of the allele (A) of SNP rs200926172 indicates an increased risk of having or being at risk of having or developing an idiopathic scoliosis disease.

the presence of the allele (A) of SNP GRCh37/hg19 : 75013289 indicates an increased risk of having or being at risk of having or developing an idiopathic scoliosis disease.

According to the invention, the presence of the allele (G) of SNP rs190991771 indicates an increased risk of having or being at risk of having or developing a Scoliosis disease.

According to the invention, the presence of the allele (A) of SNP rs200926172 indicates an increased risk of having or being at risk of having or developing a Scoliosis disease.

According to the invention, the presence of the allele (A) of SNP GRCh37/hg19 : 75013289 indicates an increased risk of having or being at risk of having or developing a Scoliosis disease.

In one embodiment of the invention, the method of identifying a subject having or at risk of having or developing a Scoliosis disease, comprising determining the presence or absence of an allelic variant in a single nucleotide polymorphism (SNP) located at exon 10 (or at exon 5 or 6 or in 5'UTR) of *POC5* (in 5q13.3 locus) in a blood sample obtained from said subject.

In one embodiment of the invention, the subject having or being at risk of having or developing a Scoliosis disease may be a substantially healthy subject, which means that the subject has not been previously diagnosed or identified as having or suffering from a Scoliosis disease, or that has not developed a Scoliosis disease.

In another embodiment, said subject may also be one that is asymptomatic for the Idiopathic Scoliosis disease. As used herein, an "asymptomatic" subject refers to a subject that does not exhibit IS symptoms, which are diagnosed, according to internationally validated criteria (Weinstein SL BJ, Weinstein SL. The Thoracolumbar spine. Turek's orthopaedics: principles and their application, 5th ed Philadelphia: JP Lippincott Company 1994:447-85.).

In another embodiment of the invention, said subject may be one that is at risk of having or developing a Scoliosis disease, as defined by clinical indicia such as for example: age, gender, clinical marker (like spine deformity), family history of Scoliosis disease.

According to the invention, the determination of the presence or absence of said SNPs may be determined by DNA sequencing, PCR analysis or any genotyping method known in the art. Examples of such methods include, but are not limited to, chemical assays such as allele specific hybridation, primer extension, allele specific oligonucleotide ligation, sequencing, enzymatic cleavage, flap endonuclease discrimination; and detection methods such as fluorescence, chemiluminescence, and mass spectrometry.

For example, the presence or absence of said polymorphism may be detected in a DNA sample, preferably after amplification. For instance, the isolated DNA may be subjected

to amplification by polymerase chain reaction (PCR), using specific oligonucleotide primers that are specific for the polymorphism or that enable amplification of a region containing the polymorphism. According to a first alternative, conditions for primer annealing may be chosen to ensure specific amplification; so that the appearance of an amplification product be a diagnostic of the presence of the polymorphism according to the invention. Otherwise, DNA may be amplified, after which a mutated site may be detected in the amplified sequence by hybridization with a suitable probe or by direct sequencing, or any other appropriate method known in the art.

Actually numerous strategies for genotype analysis are available (Cooper et al., 1991; Grompe, 1993). Briefly, the nucleic acid molecule may be tested for the presence or absence of a restriction site. When a base polymorphism creates or abolishes the recognition site of a restriction enzyme, this allows a simple direct PCR genotype of the polymorphism. Further strategies include, but are not limited to, direct sequencing, restriction fragment length polymorphism (RFLP) analysis; hybridization with allele-specific oligonucleotides (ASO) that are short synthetic probes which hybridize only to a perfectly matched sequence under suitably stringent hybridization conditions; allele-specific PCR; PCR using mutagenic primers; ligase-PCR, HOT cleavage; denaturing gradient gel electrophoresis (DGGE), temperature denaturing gradient gel electrophoresis (TGGE), single-stranded conformational polymorphism (SSCP) and denaturing high performance liquid chromatography (Kuklin et al., 1997). Direct sequencing may be accomplished by any method, including without limitation chemical sequencing, using the Maxam-Gilbert method ; by enzymatic sequencing, using the Sanger method ; mass spectrometry sequencing ; sequencing using a chip-based technology; and real-time quantitative PCR. Preferably, DNA from a subject is first subjected to amplification by polymerase chain reaction (PCR) using specific amplification primers. However several other methods are available, allowing DNA to be studied independently of PCR, such as the rolling circle amplification (RCA), the InvaderTMassay, or oligonucleotide ligation assay (OLA). OLA may be used for revealing base polymorphisms. According to this method, two oligonucleotides are constructed that hybridize to adjacent sequences in the target nucleic acid, with the join sited at the position of the polymorphism. DNA ligase will covalently join the two oligonucleotides only if they are perfectly hybridized to one of the allele.

Therefore, short DNA sequences, in particular oligonucleotide probes or primers, according to the present invention include those which specifically hybridize the one of the allele of the polymorphism.

Oligonucleotide probes or primers may contain at least 10, 15, 20 or 30 nucleotides. Their length may be shorter than 400, 300, 200 or 100 nucleotides.

The PCR primers that can be used in the present invention for amplification of a sequence comprising a SNP consisting of rs34678567, rs146984380, c.G1363C are :

Forward 5'CTTTTCATAAGGTGGGACCT3'; (SEQ ID N°1)

Reverse 5'TCCGATGCCCTTACCAG3' (SEQ ID N°2)

According to the invention, the determination of the presence or absence of said SNPs may also be determined by detection or not of the mutated POC5 protein (i.e. POC A446T, A429V, A455P, I225V or D148N) by any method known in the art. The presence of the protein of interest may be detected using standard electrophoretic and immunodiagnostic techniques, including immunoassays such as competition, direct reaction, or sandwich type assays. Such assays include, but are not limited to, Western blots; agglutination tests; enzyme-labelled and mediated immunoassays, such as ELISAs; biotin/avidin type assays; radioimmunoassays; immunoelectrophoresis; immunoprecipitation, etc. The reactions generally include revealing labels such as fluorescent, chemiluminescent, radioactive, enzymatic labels or dye molecules, or other methods for detecting the formation of a complex between the antigen and the antibody or antibodies reacted therewith. Labels are known in the art that generally provide (either directly or indirectly) a signal. As used herein, the term "labelled" with regard to the antibody or aptamer, is intended to encompass direct labelling of the antibody or aptamer by coupling (i.e., physically linking) a detectable substance, such as a radioactive agent or a fluorophore (e.g. fluorescein isothiocyanate (FITC) or phycoerythrin (PE) or indocyanine (Cy5), to the antibody or aptamer, as well as indirect labelling of the probe or antibody (e.g., horseradish peroxidise, HRP) by reactivity with a detectable substance. An antibody or aptamer may be also labelled with a radioactive molecule by any method known in the art. For example, radioactive molecules include but are not limited radioactive atom for scintigraphic studies such as I123, I124, In111, Re186 and Re188. The aforementioned assays generally involve separation of unbound protein in a liquid phase from a solid phase support to which antigen-antibody complexes are bound. Solid supports which may be used in the practice of the invention include substrates such as nitrocellulose (e.g., in membrane or microtiter well form); polyvinylchloride (e.g., sheets or microtiter wells); polystyrene latex (e.g., beads or microtiter plates); polyvinylidine fluoride; diazotized paper; nylon membranes; activated beads, magnetically responsive beads, etc.

More particularly, an ELISA method may be used, wherein the wells of a microtiter plate are coated with an antibody against the protein to be tested. A biological sample

containing or suspected of containing the marker protein is then added to the coated wells. After a period of incubation sufficient to allow the formation of antibody-antigen complexes, the plate (s) can be washed to remove unbound moieties and a detectably labelled secondary binding molecule added. The secondary binding molecule is allowed to react with any captured sample marker protein, the plate washed and the presence of the secondary binding molecule detected using methods well known in the art.

Alternatively, an immunohistochemistry (IHC) method may be used. IHC specifically provides a method of detecting a target in a biological sample or tissue specimen in situ. The overall cellular integrity of the sample is maintained in IHC, thus allowing detection of both the presence and location of the target of interest. Typically a biological sample is fixed with formalin, embedded in paraffin and cut into sections for staining and subsequent inspection by light microscopy. Current methods of IHC use either direct labelling or secondary antibody-based or hapten-based labelling. Examples of known IHC systems include, for example, EnVision™ (DakoCytomation), Powervision® (Immunovision, Springdale, AZ), the NBA™ kit (Zymed Laboratories Inc., South San Francisco, CA), HistoFine® (Nichirei Corp, Tokyo, Japan).

In particular embodiment, a tissue section (e.g. a liver tumor sample or biopsy) may be mounted on a slide or other support after incubation with antibodies directed against (the) protein(s) encoded by POC5 gene with SNPs located at exon 11. Then, microscopic inspections in the sample mounted on a suitable solid support may be performed. For the production of photomicrographs, sections comprising samples may be mounted on a glass slide or other planar support, to highlight by selective staining the presence of the protein of interest. Therefore IHC samples may include, for instance: (a) preparations comprising cell samples (b) fixed and embedded said cells and (c) detecting the protein of interest in said cell samples. In some embodiments, an IHC staining procedure may comprise steps such as: cutting and trimming tissue, fixation, dehydration, paraffin infiltration, cutting in thin sections, mounting onto glass slides, baking, deparaffination, rehydration, antigen retrieval, blocking steps, applying primary antibodies, washing, applying secondary antibodies (optionally coupled to a suitable detectable label), washing, counter staining, and microscopic examination.

A second object of the invention is a kit for identifying whether a subject has or is at risk of having or developing a Scoliosis disease, comprising:

- at least a means for detecting the SNP selected from the group consisting of rs34678567, rs146984380, c.G1363C, and

- instructions for use

In one embodiment of the invention, the kit for identifying whether a subject has or is at risk of having or developing a Scoliosis disease, comprising:

- at least one primer and/or at least one probe for amplification of a sequence comprising a SNP consisting of rs34678567, rs146984380, c.G1363C and

- instructions for use

In one embodiment of the invention, the primer or probe may be labelled with a suitable marker. In another embodiment of the invention, the primer or probe may be coated on an array.

The PCR primers that can be used in the present invention for amplification of a sequence comprising a SNP consisting of rs34678567, rs146984380, c.G1363C are

Forward 5'CTTTTCATAAGGTGGGACCT3'; (SEQ ID N°1)

Reverse 5'TCCGATGCCCTTACCAG3' (SEQ ID N°2)

Therapeutic method

As previously the inventors demonstrate in the functional analysis of POC5 mutations in zebrafish model, spine deformities similar to those observed in patients compared to zebrafich control without any mutation in exon 10 of the *POC5* gene

Accordingly a third object of the present invention is a nuclease for use in treating Idiopathic Scoliosis and/or preventing progression of Idiopathic Scoliosis in a patient, wherein the presence of SNPs at exon 10 of the *POC5* nucleic sequence in a sample previously obtained from said patient, have been detected by a method of the invention previously described in order to repair genetic point mutations of the SNPs located at the exon 10 of POC5 nucleic sequence.

In particular embodiment the SNP is selected from the group consisting of rs34678567, rs146984380, c.G1363C and wherein

- the presence of the allele (T) of SNP rs34678567 indicates an increased risk of having or being at risk of having or developing a idiopathic scoliosis disease, and

- the presence of the allele (T) of SNP rs146984380 indicates a increased risk of having or being at risk of having or developing a idiopathic scoliosis disease

- the presence of the allele (C) of SNP c.G1363C indicates a increased risk of having or being at risk of having or developing a idiopathic scoliosis disease.

A man skilled in the art, know as to design a specific nuclease in order to repair of genetic point mutations like SNPs located at the exon 10 of POC5 nucleic sequence , namely rs34678567, rs146984380, c.G1363C.

The term "nuclease" or "endonuclease" means synthetic nucleases consisting of a DNA binding site, a linker, and a cleavage module derived from a restriction endonuclease which are used for gene targeting efforts. The synthetic nucleases according to the invention exhibit increased preference and specificity to bipartite or tripartite DNA target sites comprising DNA binding (i.e. TALE recognition site(s)) and restriction endonuclease target site while cleaving at off-target sites comprising only the restriction endonuclease target site is prevented.

Restriction endonucleases (also called restriction enzymes) as referred to herein in accordance with the present invention are capable of recognizing and cleaving a DNA molecule at a specific DNA cleavage site between predefined nucleotides. In contrast, some endonucleases such as for example FokI comprise a cleavage domain that cleaves the DNA unspecifically at a certain position regardless of the nucleotides present at this position. Therefore, preferably the specific DNA cleavage site and the DNA recognition site of the restriction endonuclease are identical. Moreover, also preferably the cleavage domain of the chimeric nuclease is derived from a restriction endonuclease with reduced DNA binding and/or reduced catalytic activity when compared to the wildtype restriction endonuclease.

According to the knowledge that restriction endonucleases, particularly type II restriction endonucleases, bind as a homodimer to DNA regularly, the chimeric nucleases as referred to herein may be related to homodimerization of two restriction endonucleases subunits. Preferably, in accordance with the present invention the cleavage modules referred to herein have a reduced capability of forming homodimers in the absence of the DNA recognition site, thereby preventing unspecific DNA binding. Therefore, a functional homodimer is only formed upon recruitment of chimeric nucleases monomers to the specific DNA recognition sites. Preferably, the restriction endonuclease from which the cleavage module of the chimeric nuclease is derived is a type IIP restriction endonuclease. The preferably palindromic DNA recognition sites of these restriction endonucleases consist of at least four or up to eight contiguous nucleotides. Preferably, the type IIP restriction endonucleases cleave the DNA within the recognition site which occurs rather frequently in

the genome, or immediately adjacent thereto, and have no or a reduced star activity. The type IIP restriction endonucleases as referred to herein are preferably selected from the group consisting of: PvuII, EcoRV, BamHI, BcnI, BfaSORF1835P, BfiI, BglI, BglII, BpuJI, Bse6341, BsoBI, BspD6I, BstYI, Cfr10I, Ecl18kI, EcoO109I, EcoRI, EcoRII, EcoRV, EcoR124I, EcoR124II, HinP11, HinclI, HindIII, Hpy99I, Hpy188I, MspI, MunI, MvaI, NaeI, NgoMIV, NotI, OkrAI, PabI, PacI, PspGI, Sau3AI, SdaI, SfiI, SgrAI, ThaI, VvuYORF266P, DdeI, Eco57I, HaeIII, HhaII, HindII, and NdeI.

Other nuclease for use in the present invention are disclosed in WO 2010/079430, WO2011072246, WO2013045480, Mussolino C, et al (Curr Opin Biotechnol. 2012 Oct;23(5):644-50) and Papaioannou I. et al (Expert Opinion on Biological Therapy, March 2012, Vol. 12, No. 3 : 329-342) all of which are herein incorporated by reference.

Another object of the present invention is a method of treating Idiopathic Scoliosis in a subject comprising the steps of :

a) providing a biological sample from a subject,

b) detecting in a biological sample obtained at step a) the presence or absence of an allelic variant of single nucleotide polymorphism (SNP) located at the exon 10 of POC5 nucleic sequence of ; and

if a SNPs is detected,

treating the subject with an nuclease in order to repair genetic point mutations of the SNPs located at the exon 10 of POC5 nucleic sequence.

In particular embodiment the SNP is selected from the group consisting of rs34678567, rs146984380, c.G1363C and wherein

- the presence of the allele (T) of SNP rs34678567 indicates an increased risk of having or being at risk of having or developing a idiopathic scoliosis disease, and
- the presence of the allele (T) of SNP rs146984380 indicates a increased risk of having or being at risk of having or developing a idiopathic scoliosis disease
- the presence of the allele (C) of SNP c.G1363C indicates a increased risk of having or being at risk of having or developing a idiopathic scoliosis disease.

The invention will be further illustrated by the following figures and examples. However, these examples and figures should not be interpreted in any way as limiting the scope of the present invention.

FIGURES:

Figure 1- Pedigrees of IS families and sporadic cases with exon 10 *POC5* mutations

A) and **B**) IS families F2, F19, F35 and F41, and sporadic cases SC39, SC58 and SC83 show the same *POC5* missense mutation c.1336G>A listed in dbSNP135 (rs34678567). Sample chromatograms for a healthy individual, without the mutation (upper left panel), and an affected individual, with the mutation (lower left panel) are shown. The mutation position is indicated by an arrow above the chromatogram. Pedigrees for families F2, F19, F35 and F41, and sporadic cases SC39, SC58 and SC83 are shown. **C**) Pedigrees of sporadic IS cases SC1, SC77, SC137, SC149 and SC150 harbouring the *POC5* missense mutation c.1286C>T listed in dbSNP135 (rs146984380), and corresponding chromatograms. **D**) Pedigree of IS family F31 harbouring the *POC5* missense mutation c.1363G>C, and corresponding chromatograms. Cobb's angles and genotype at the mutated position are indicated below symbols corresponding to individuals; U: uncertain status, ND: status not determined.

Figure 2- Mutant POC5 expression leads to scoliosis in zebrafish

A) POC5 was expressed in zebrafish injected with myc-tagged wild-type (wt-POC5) or mutated POC5 (mut-POC5) mRNAs **(i)** but not in non-injected wild-type (WT) fish. Embryos injected with mut-POC5 mRNAs (A446T (n=153), A429V (n=130) and A455P (n=149)) exhibited mild to severe curvature of the body axis at 3 dpf compared to non-injected wild-type (n=234) and wt-POC5 (n=122) injected embryos. **(ii)** Bar graph illustrating the frequency of axis curvature defects in WT, wt-POC5 and mut-POC5 fish. **(iii)** Examples of the four phenotypic classes observed: unaffected, mild (highlighted by an arrow), moderate, severe. **B**) Lateral view of juvenile zebrafish expressing wild-type **(i)** and mutPOC5 **(ii-iii)**. With calcein staining, wild-type fish **(i)** showed no curvature, while mut-POC5 fish **(ii-iii)** revealed spine curvature phenotypes. Some mut-POC5 fish exhibited a curve occurring primarily in the sagittal plane of the fish (downward curve) **(ii)** others exhibited deviation of the spine on the coronal plane (lateral curve) with slight rotation of the spine **(iii)**. Spinal deformities are highlighted by black arrows.

Figure 3 - Genetic refinement of the 5q13.3 idiopathic scoliosis interval in family F2

An IlluminaOmniExpress chip was used for whole-genome genotyping of IS family F2, revealing a minimum IS interval of 5.582 Mb on chromosome 5. Haplotypes with only

some of these SNP are shown on a simplified pedigree. Arrowheads show centromeric and telomeric recombination events. The refined 5q13.3 IS interval is limited by rs300263 and rs4704627 as its centromeric and telomeric boundaries, respectively.

EXAMPLE:

Material & Methods

Patients

191 independent IS families (41 multiplex families and 150 sporadic cases) participated in this study. Multiplex IS families included 153 affected individuals, 52 individuals with uncertain status and 120 unaffected individuals. 6/41 multiplex families included 5-11 affected individuals, and 35/41 multiplex families were composed of 2-4 affected individuals. 103 control individuals consisted in healthy parents of individuals presenting intellectual disabilities.

IS was diagnosed by combining clinical examination of the spine, including the forward bending test (Adams test), and measurement of Cobb's angle on X-rays .

This study was approved by the local ethics committee and was promoted by the Hospices Civils de Lyon, France.

DNA sample collection and lymphoblast cell lines

Blood samples were collected from each participant (IS patients and their affected or unaffected relatives) after providing them with an information sheet describing the aim of the research and obtaining their written informed consent for participation in this study. DNA was extracted from peripheral blood using a QIAmp DNA Blood Midi Kit (Qiagen), according to the manufacturer's instructions. Lymphoblast cell lines were established for at least one affected individual in each multiplex IS family.

SNP genotyping and genome-wide linkage scan

DNA samples (500 ng) were hybridised on 700k Illumina HumanOmniExpress SNP arrays (Illumina) as described in the manufacturer's protocol. Genotypes were analysed and the locations of the recombination events were identified using Merlin° software. Haplotypes were drawn using Haplopainter° software. Critical chromosomal regions were defined as segments (haplotypes) where SNPs are identical in all affected relatives (affected only method).

Exome capture and sequencing

Enrichment of exomes was performed with an Agilent SureSelect all exome kit (V2 optimised for ABI SOLiD sequencing and V4 optimised for Illumina HiSEQ sequencing), using 2 µg of subjects' genomic DNA. This enrichment is designed to cover approximately 50 Mb of genomic sequences, mainly protein coding sequences. Exon-enriched DNA libraries from 7 affected members of the same family were sequenced individually, the first four samples on a single ABI SOLiD 4 sequencer slide, and the remaining three on a lane in an Illumina HiSEQ 2000 platform in accordance with the manufacturer's instructions. This produced 50-base-pair end reads and 100-base-pair end reads (for SOLiD and HISeq, respectively).

Read mapping and variant calling

The Burrows-Wheeler Aligner (BWA)[27] was used as the main aligner for mapping against the human genome (hg19). The human genome was indexed using the bwtsw algorithm included with BWA. Alignment was performed using a maximum mismatch penalty of three. All other BWA parameters were left at their default values. The alignment was generated in pair-end mode, and SAMTOOLS[28] was used for storage. All PCR duplicates were removed from alignments. The "Best Practice Variant Detection with GATK v2" was used for SNP and indel calling[29,30].

Sanger sequencing of the recurrent *POC5* variant

A DNA fragment from all individuals from IS family F2, from one proband from each of the remaining 40 IS families (F1 and F3-F41) and from 150 IS cases (C1-C150) - containing the c.G1336A (p.A446T) *POC5* rare SNV identified by whole-exome sequencing - was PCR-amplified for classic Sanger sequencing. This amplicon was then sequenced in all available individuals from IS families F19, F31, F35 and F41, in which the c.G1336A (p.A446T) *POC5* rare SNV was detected. The PCR primers used were as follows: Forward 5'CTTTTCATAAGGTGGGACCT3'; Reverse 5'TCCGATGCCCTTACCAG3'. PCR was performed on a FlexCycler (AnalytikJena). PCR products were purified using commonly

applied methods before analysis of amplicons on an ABI 3730xl DNA Analyzer (Applied Biosystems).

***POC5*-targeted high-throughput sequencing**

A Fluidigm Access Array° device (IntegraGen) was used to study whole exonic, flanking intronic and regulatory *POC5* sequences in a proband from 40 of the multiplex IS families (F1/ F3-F41), 150 sporadic cases (SC1-150) and 103 control individuals (206 control chromosomes).

POC5 Western blot

Total protein lysates from lymphoblast cell lines derived from two patients, members of family F2, and two healthy controls were resolved on a NuPAGE® Novex® Gel (4-12% Bis-Tris) (Life Technologies). After migration, proteins were transferred to a Hybond-ECL Nitrocellulose membrane (Amersham Biosciences). The membrane was incubated with antibodies for POC5 (1:300; Abcam) and GAPDH (1:3000; Abcam). POC5 amounts were determined by semi-quantitative analysis, using Scion-image° software according to the recommendations in the software handbook.

Functional validation in zebrafish

Zebrafish were raised from a colony maintained according to established procedures. All procedures described here were carried out in compliance with the guidelines published by the Canadian Council for Animal Care.

RNA-driven expression of POC5 in injected embryos was performed using a myc-tagged ORF clone of human POC5 (Origene). Mutations were introduced into this vector by site-directed mutagenesis using a QuikChange® XL Site-Directed Mutagenesis Kit (Agilent). Messenger RNAs were transcribed from linearised constructs using the mMESSAGE mMACHINE kit (Ambion). Messenger RNAs were injected into one- or two-cell stage embryos using a Picospritzer. The final injection volume was ~1.5 nl, at a concentration of 50 nM messenger RNA. Injected and non-injected embryos were then incubated in appropriate media at 28.5 °C for 24 h, after which they were assessed for viability. Morphological differences between injected and non-injected embryos were assessed under an Olympus SZX12 stereoscope.

POC5 expression in injected embryos was confirmed by Western blot of a total protein extract (40 μg) from 3 days-post-fertilization (dpf) mut-POC5 zebrafish. Blots were probed with antibodies for myc (rabbit polyclonal; at a dilution of 1:2000; Sigma) or γ tubulin as loading control (Sigma; 1:5000).

Whole-mount zebrafish bone staining

Calcein staining was performed on juvenile fish as previously described in Patten et al.[31]. Briefly, fish were first anaesthetised in 0.6 mM MS-222, buffered to pH 7.0, before immersion for 10 min in 0.2% calcein in 10% Hank's solution, buffered to pH 7.2. After staining, fish were washed three times in Hank's solution for 10 min. Vertebral mineralization was assessed after calcein staining using a Leica DMR microscope.

***In situ* hybridisation**

Hybridisation using sense and anti-sense probes targeting the zebrafish orthologue of POC5 was used to observe the endogenous localisation of poc5 mRNA. Embryos at 3 dpf and 5 dpf were processed for in situ hybridisation as previously described[31].

Results

Our initial hypothesis was that, whereas multifactorial inheritance is likely to account for a large number of IS cases, monogenic inheritance -i.e. transmission of a rare mutation with a strong effect- might explain at least a subgroup of multiplex IS families for which the disease appears to be transmitted as an autosomal dominant trait. Our recent linkage analysis in a large multiplex IS family (family F2) compatible with autosomal dominant inheritance revealed disease marker co-segregation at two locations, 3q12.1 and 5q13.3, in the eleven affected members of this family[19]. Direct Sanger sequencing of a number of candidate genes from these two regions did not reveal any causative mutation. The lack of high-throughput methods to investigate all the candidate genes from the large linkage regions hampered progress on this family at the time of this previous study. Now, however, exome sequencing has emerged as a rapid and efficient high-throughput tool to screen for mutations[32]. In the present study, we therefore performed exome sequencing after genome-wide linkage using the Illumina HumanOmniExpress chip in IS family F2 to refine the 3q12.1 and 5q13.3 candidate regions in which the disease locus lies to chromosome 3: 95,083,093-107,153,338 and chromosome 5: 73,905,694–79,488,191, respectively (NCBI hg19; Fig. 3).

Whole-exome sequencing was then performed on seven affected individuals from IS family F2. Sequencing data were of good quality, with 86% of the exons in the regions tested proving sufficient for variant calling. Variant calling was performed independently on all samples, and very stringent filter thresholds were applied to include the following potential mutations: non-synonymous, splice, frameshift and nonsense variants, and variants that were not listed, or were present at less than 5% frequency in the 1000 genomes, EVS and in-house exome variants databases. This analysis revealed ~550–760 single-nucleotide variants (SNVs) in the 3q12.1 and 5q13.3 chromosomal regions in the 7 subjects (4,564 SNVs in total). After applying the filters, only one non-synonymous SNV was found in all seven affected individuals (Table S1). This SNV, c.G1336A, is located in the *POC5* gene (NM_001099271). *POC5* codes for a centriolar protein homologue and the variant results in a single amino acid change, p.A446T. This variant is known (rs34678567) and has an allelic frequency of ~ 1.6% in the European population represented in the EVS database. To evaluate the potential association between this variant and the disease, we sequenced all available affected and unaffected individuals from IS family F2 and from 40 additional multiplex IS families. The c.G1336A *POC5* variant was found in all affected members of family F2 (Fig. 1A), and in all affected members of 3 additional multiplex IS families (F19, F35 and F41; Fig. 1B), providing strong evidence for its association with IS. To identify additional modifications to *POC5*, we sequenced the whole exonic, flanking intronic and regulatory *POC5* sequences of this gene in all 40 IS families (families F1, F3-41), 150 sporadic IS cases (SC1-150) and 103 control individuals (206 control chromosomes). The c.G1336A *POC5* variant was found in 3/150 sporadic cases, but never in control individuals (Fig. 1A and 1B). Another rare *POC5* variant, c.C1286T (p.A429V; rs146984380, allelic frequency of ~ 0.9% in 1000 genomes and EVS databases), was found in 5/150 sporadic IS cases (Fig. 1C), but once again never in control individuals. We also identified a *POC5* missense mutation, c.G1363C (p.A455P), in a single IS family (Fig. 1D). This mutation was undetected in sporadic IS cases, control individuals and in the 1000 genomes, EVS and in-house exome variants databases. To determine whether POC5-protein levels are perturbed in IS, lymphoblast cell lines were generated from two affected members of IS family F2, and their unaffected relatives (not bearing the c.G1336A *POC5* SNV). This revealed a slight decrease in the amount of POC5 protein expressed in IS lymphoblasts.

POC5 protein is highly conserved across species[33]. However, very little is known about its biological function in vertebrates. In view of the high conservation between the

protein orthologues of zebrafish (*Danio rerio*) and human (58% amino acid similarity), we used a zebrafish model system to determine the functional significance of the variants A446T, A429V and A455P. We first designed a morpholino anti-sense oligomer targeting the ATG of the zebrafish *poc5* orthologue of *POC5* (XM_685988). Injection of this morpholino caused curved notochord formation in a minority (<10%) of embryos. Next, we used the zebrafish to develop models for the A446T, A429V and A445P human *POC5* changes observed in familial and sporadic IS cases. We injected a myc-tagged synthetic mRNA into zebrafish embryos. This mRNA corresponds to either the wild-type *POC5* (wt-*POC5*) coding sequence, or mutated versions of *POC5* engineered with variants A446T, A429V and A455P. Expression of these exogenous mRNAs in zebrafish was confirmed by Western blot (Fig. 2A(i)). All four constructs resulted in a similar protein expression level. Embryos injected with mut-POC5 mRNAs showed an overt phenotype from 3 dpf, with mild to severe curvature of the body axis in at least 45% of injected embryos compared to wild-type fish ($p<0.001$; Fig. 2A(ii-iii)). Body axis defects consisted of either downward or lateral curvature. Phenotypes were classed as unaffected, mildly defective (minor downward curve in the notochord), moderately defective (curved body axis with downward curvature), or severely defective (lateral curvature of body axis usually accompanied by rotation) (Fig. 2A(iii)). When mut-POC5 fish developed to juvenile stages (20-30 dpf), we observed downward (Fig. 2B(ii)) or lateral (Fig. 2B(iii)) curvature of the fully mineralised vertebral column, reminiscent of the phenotype observed in IS patients. Injection of wt-POC5 mRNA at the same concentrations and producing a similar level of expressed protein did not have any obvious effect. Altogether, the genetic studies and zebrafish experiments presented here suggest that the A446T, A429V and A445P *POC5* variants act in a dominant negative fashion to cause spinal deformities in patients.

In humans, POC5 protein localises to the distal portion of centrioles. This protein is recruited to procentrioles for full maturation of the centriole, and normal cell-cycle processing[33]. However, whole-mount *in situ* hybridisation of zebrafish embryos revealed expression of POC5 mRNA to be mainly restricted to the brain, with low-level labeling near the heart and gut regions at 3 and 5 dpf. At 3 dpf, a particularly high level of expression was evident at the midbrain-hindbrain boundary; a crucial organizing center of neural patterning in the brain[34]. These data therefore suggest that the IS phenotype in zebrafish likely originates from a brain defect having consequences on spinal morphology.

Three additional *POC5* SNVs, including 2 rare missense SNVs and a novel 5'UTR SNV were identified (Table 2). The pathogenicity and possible role of these mutations in IS will be determined.

Discussion

In this invention, we describe detection of IS-causing variants in the *POC5* gene. The rare A446T *POC5* variant was found in all affected patients from IS family F2 and from 3/40 additional IS families and 3/150 sporadic IS cases, but was undetected in 206 control chromosomes. This recurrent mutation may therefore be found in up to 10% of multiplex IS families and a significant number of sporadic IS cases. Given the high prevalence (3-4%) and incomplete penetrance of IS in the general population, it is very likely that individuals sequenced or genotyped as part of the HapMap or genome project consortia (including the 1000 Genomes) could be affected with, or predisposed to IS. We therefore cannot rule out that known single-nucleotide variants with very low minor allelic frequency could be pathogenic. Along the same lines, among the 4 IS families where this mutation was detected, 4/23 individuals with the mutation were classed as either unaffected or uncertain (Fig. 1), supporting the suggestion that this variant is associated with incomplete disease penetrance. Moreover, a spine deformity similar to that observed in human individuals was observed in ~60% of A446T *POC5* mutant zebrafish, but never in wild-type fish. Thus, we have independent corroborative evidence indicating the pathology of this rare variant. Henceforth, when this mutation is found in a given IS family, it should be possible to use a presymptomatic test on young at-risk members of the family to identify those who may go on to develop the disease, and thus requiring specific orthopaedic follow-up. We believe that pre-adolescents harbouring the A446T *POC5* rare variant should receive regular clinical orthopaedic assessment until bone maturation is completed (~ age 16). Although our zebrafish experiments also showed that the *POC5* p.A429V rare variant and p.A455P mutation both cause spine deformity, the penetrance of the disease caused by these mutations is as yet unknown in humans, making it unwise to provide clinical recommendations until additional familial data become available.

In summary, with this invention we have identified the first ever causative gene for IS. The present work is likely to pave the way to deciphering the genetic causes of familial and sporadic IS and provides a molecular diagnostic test for IS. Since *POC5* is highly expressed in the zebrafish brain during early embryonic development, we speculate that IS could be a brain-associated disorder having connections with neurogenic, left-right axis determination, proprioceptive or metabolic pathophysiological hypotheses.

Table S1- Summary of Single Nucleotide Variants (SNVs) for Exome captured samples

(i) Whole exome sequence filtering data for the seven exome-sequenced individuals

Pooled Variants in targeted regions 3 and 5 from all patients	4,564
Nonsynonymous/nonsense/splice/frame shift in all patients	1260
Variants common in all affected individuals	30
Heterozygous variants	7
Variants not in dbSNP	0*
< 5% allelic frequency in 1000 genomes	3
< 5% allelic frequency <in EVS and in-house exome database	1

Variant calling was made independently on samples and SNVs were then pooled and filtered. *All the SNVs were annotated in the dbSNP. We therefore re-filter the 7 remaining heterozygous variants to keep SNVs with a minor allelic frequency <5% in European ancestry of 1000 genomes, EVS and in our house-exome database.

(ii) Details of the remaining 1 SNV (i) that is shared by the seven exome-sequenced individuals

Gene	Full name of Protein	Chromosome	rs number	Genomic	cDNA	Protein
POC5	POC5 centriolar protein homolog	5	rs34678567	g.74981103	c.G1336A	p.A446T

Table 2- Additional *POC5* SNVs Identified Using High-Throughput *POC5* Sequencing

IS Patients	Position (Mb) GRCh37/hg19	dbSNP138	DNA Change	Mutation	AA change	Transcript ID Ensembl	Protein ID Ensembl	Protein UniProtKB	AA Position	Allele frequency (dbSNP, 1000 Genomes)	Effect Prediction (SNPnexus)
C10	74990497	rs190991771	T>C	Missense	I/V	ENST00000428202	ENSP00000410216	Q8NA72	225	No allele frequency	Benign
						ENST00000380475	ENSP00000369842	Q8NA72-2	108		Benign
						ENST00000446329	ENSP00000399481	Q8NA72-3	200		Benign
C39, F41 (IL1, IL2)	74998501	rs200926172	C>T	Missense	D/N	ENST00000428202	ENSP00000410216	Q8NA72	148	No allele frequency	Benign
						ENST00000380475	ENSP00000369842	Q8NA72-2	31		Possibly damaging
						ENST00000446329	ENSP00000399481	Q8NA72-3	123		Benign
C128	75013289	Novel	C>A	5'UTR							CpG:41 Island Change

REFERENCES:

Throughout this application, various references describe the state of the art to which this invention pertains. The disclosures of these references are hereby incorporated by reference into the present disclosure.

1. Weinstein SL BJ, Weinstein SL. The Thoracolumbar spine. Turek's orthopaedics: principles and their application, 5th ed Philadelphia: JP Lippincott Company 1994:447-85.

2. Lonstein JE, Bjorklund S, Wanninger MH, Nelson RP. Voluntary school screening for scoliosis in Minnesota. The Journal of bone and joint surgery American volume 1982;64:481-8.

3. Kapoor M, Laham SG, Sawyer JR. Children at risk identified in an urban scoliosis school screening program: a new model. Journal of pediatric orthopedics Part B 2008;17:281-7.

4. Sahlstrand T, Petruson B. A study of labyrinthine function in patients with adolescent idiopathic scoliosis. I. An electro-nystagmographic study. Acta orthopaedica Scandinavica 1979;50:759-69.

5. Rousie D, Hache JC, Pellerin P, Deroubaix JP, Van Tichelen P, Berthoz A. Oculomotor, postural, and perceptual asymmetries associated with a common cause. Craniofacial asymmetries and asymmetries in vestibular organ anatomy. Annals of the New York Academy of Sciences 1999;871:439-46.

6. Barrack RL, Wyatt MP, Whitecloud TS, 3rd, Burke SW, Roberts JM, Brinker MR. Vibratory hypersensitivity in idiopathic scoliosis. Journal of pediatric orthopedics 1988;8:389-95.

7. Bylund P, Jansson E, Dahlberg E, Eriksson E. Muscle fiber types in thoracic erector spinae muscles. Fiber types in idiopathic and other forms of scoliosis. Clinical orthopaedics and related research 1987:222-8.

8. Slager UT, Hsu JD. Morphometry and pathology of the paraspinous muscles in idiopathic scoliosis. Developmental medicine and child neurology 1986;28:749-56.

9. Birchall D, Hughes DG, Hindle J, Robinson L, Williamson JB. Measurement of vertebral rotation in adolescent idiopathic scoliosis using three-dimensional magnetic resonance imaging. Spine 1997;22:2403-7.

10. Pedrini-Mille A, Pedrini VA, Tudisco C, Ponseti IV, Weinstein SL, Maynard JA. Proteoglycans of human scoliotic intervertebral disc. The Journal of bone and joint surgery American volume 1983;65:815-23.

11. Machida M, Dubousset J, Imamura Y, Miyashita Y, Yamada T, Kimura J. Melatonin. A possible role in pathogenesis of adolescent idiopathic scoliosis. Spine 1996;21:1147-52.

12. Kindsfater K, Lowe T, Lawellin D, Weinstein D, Akmakjian J. Levels of platelet calmodulin for the prediction of progression and severity of adolescent idiopathic scoliosis. The Journal of bone and joint surgery American volume 1994;76:1186-92.

13. Cowell HR, Hall JN, MacEwen GD. Genetic aspects of idiopathic scoliosis. A Nicholas Andry Award essay, 1970. Clinical orthopaedics and related research 1972;86:121-31.

14. Garland HG. Hereditary Scoliosis. British medical journal 1934;1:328.

15. Miller NH. Genetics of familial idiopathic scoliosis. Clinical orthopaedics and related research 2007;462:6-10.

16. 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.

17. Harrington PR. The etiology of idiopathic scoliosis. Clinical orthopaedics and related research 1977:17-25.

18. Mongird-Nakonieczna J, Kozlowski B. [Familial occurrence of idiopathic scoliosis]. Chirurgia narzadow ruchu i ortopedia polska 1976;41:161-5.

19. Edery P, Margaritte-Jeannin P, Biot B, et al. New disease gene location and high genetic heterogeneity in idiopathic scoliosis. Eur J Hum Genet 2011;19:865-9.

20. Ocaka L, Zhao C, Reed JA, et al. Assignment of two loci for autosomal dominant adolescent idiopathic scoliosis to chromosomes 9q31.2-q34.2 and 17q25.3-qtel. Journal of medical genetics 2008;45:87-92.

21. Chan V, Fong GC, Luk KD, et al. A genetic locus for adolescent idiopathic scoliosis linked to chromosome 19p13.3. American journal of human genetics 2002;71:401-6.

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

23. Salehi LB, Mangino M, De Serio S, et al. Assignment of a locus for autosomal dominant idiopathic scoliosis (IS) to human chromosome 17p11. Human genetics 2002;111:401-4.

24. Miller NH, Justice CM, Marosy B, et al. Intra-familial tests of association between familial idiopathic scoliosis and linked regions on 9q31.3-q34.3 and 16p12.3-q22.2. Human heredity 2012;74:36-44.

25. 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.

26. 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.

27. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754-60.

28. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078-9.

29. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature genetics 2011;43:491-8.

30. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297-303.

31. Patten SA, Jacobs-McDaniels NL, Zaouter C, Drapeau P, Albertson RC, Moldovan F. Role of Chd7 in zebrafish: a model for CHARGE syndrome. PloS one 2012;7:e31650.

32. Gonzaga-Jauregui C, Lupski JR, Gibbs RA. Human genome sequencing in health and disease. Annual review of medicine 2012;63:35-61.

33. Azimzadeh J, Hergert P, Delouvee A, et al. hPOC5 is a centrin-binding protein required for assembly of full-length centrioles. The Journal of cell biology 2009;185:101-14.

34. Wassef M, Joyner AL. Early mesencephalon/metencephalon patterning and development of the cerebellum. Perspectives on developmental neurobiology 1997;5:3-16.

CLAIMS:

1. A method of identifying a subject having or at risk of having or developing a idiopathic scoliosis, comprising determining, in a sample obtained from said subject, the presence or absence of an allelic variant of single nucleotide polymorphism (SNP) located in *POC5* nucleic sequence.

2. The method according to claim 1, wherein the SNP is located in exon 10 of the *POC5* nucleic sequence

3. The method according to claim 2, wherein the SNP is selected from the group consisting of rs34678567, rs146984380, c.G1363C and wherein

- the presence of the allele (A) of SNP rs34678567 indicates an increased risk of having or being at risk of having or developing an idiopathic scoliosis disease, and
- the presence of the allele (T) of SNP rs146984380 indicates a increased risk of having or being at risk of having or developing an idiopathic scoliosis disease
- the presence of the allele (C) of SNP c.G1363C indicates a increased risk of having or being at risk of having or developing an idiopathic scoliosis disease.

4. The method according to claim 1, wherein the SNP is located in exon 5, 6 or in 5'UTR of the *POC5* nucleic sequence.

5. The method according to claim 4, wherein the SNP is selected from the group consisting of rs190991771, rs200926172, GRCh37/hg19 : 75013289 C>A and wherein

- the presence of the allele (G) of SNP rs190991771 indicates an increased risk of having or being at risk of having or developing an idiopathic scoliosis disease,

- the presence of the allele (A) of SNP rs200926172 indicates an increased risk of having or being at risk of having or developing an idiopathic scoliosis disease.

- the presence of the allele (A) of SNP GRCh37/hg19 : 75013289 indicates an increased risk of having or being at risk of having or developing an idiopathic scoliosis disease.

6. The method according to claim 1 to 5, wherein the sample is a blood sample

7. The method according to claim 1 to 6, wherein the presence or absence of said SNP is determined by nucleic acid sequencing or by PCR analysis.

8. A kit for identifying whether a subject has or is at risk of having or developing a idiopathic scoliosis, comprising :

- at least a means for detecting the SNP selected from the group consisting of rs34678567, rs146984380; c.G1363C and
- instructions for use.

9. A kit according to claim 8, comprising:

- at least one primer and/or at least one probe for amplification of a sequence comprising a SNP consisting of rs34678567, rs146984380, c.G1363C,
- instructions for use.

10. A nuclease for use in treating Idiopathic Scoliosis and/or preventing progression of Idiopathic Scoliosis in a patient, wherein the presence of SNPs in exon 10 of the *POC5* nucleic sequence in a sample previously obtained from said patient, have been detected by a method according to claim 1 to 3 and claim 6 to 7, said nuclease is used in order to repair genetic point mutations of the SNPs located at the exon 10 of POC5 nucleic sequence.

11. A nuclease for use according to claim 7, wherein the SNP is selected from the group consisting of rs34678567, rs146984380, c.G1363C and wherein

- the presence of the allele (T) of SNP rs34678567 indicates an increased risk of having or being at risk of having or developing an idiopathic scoliosis disease, and

- the presence of the allele (T) of SNP rs146984380 indicates a increased risk of having or being at risk of having or developing an idiopathic scoliosis disease

- the presence of the allele (C) of SNP c.G1363C indicates a increased risk of having or being at risk of having or developing an idiopathic scoliosis disease.

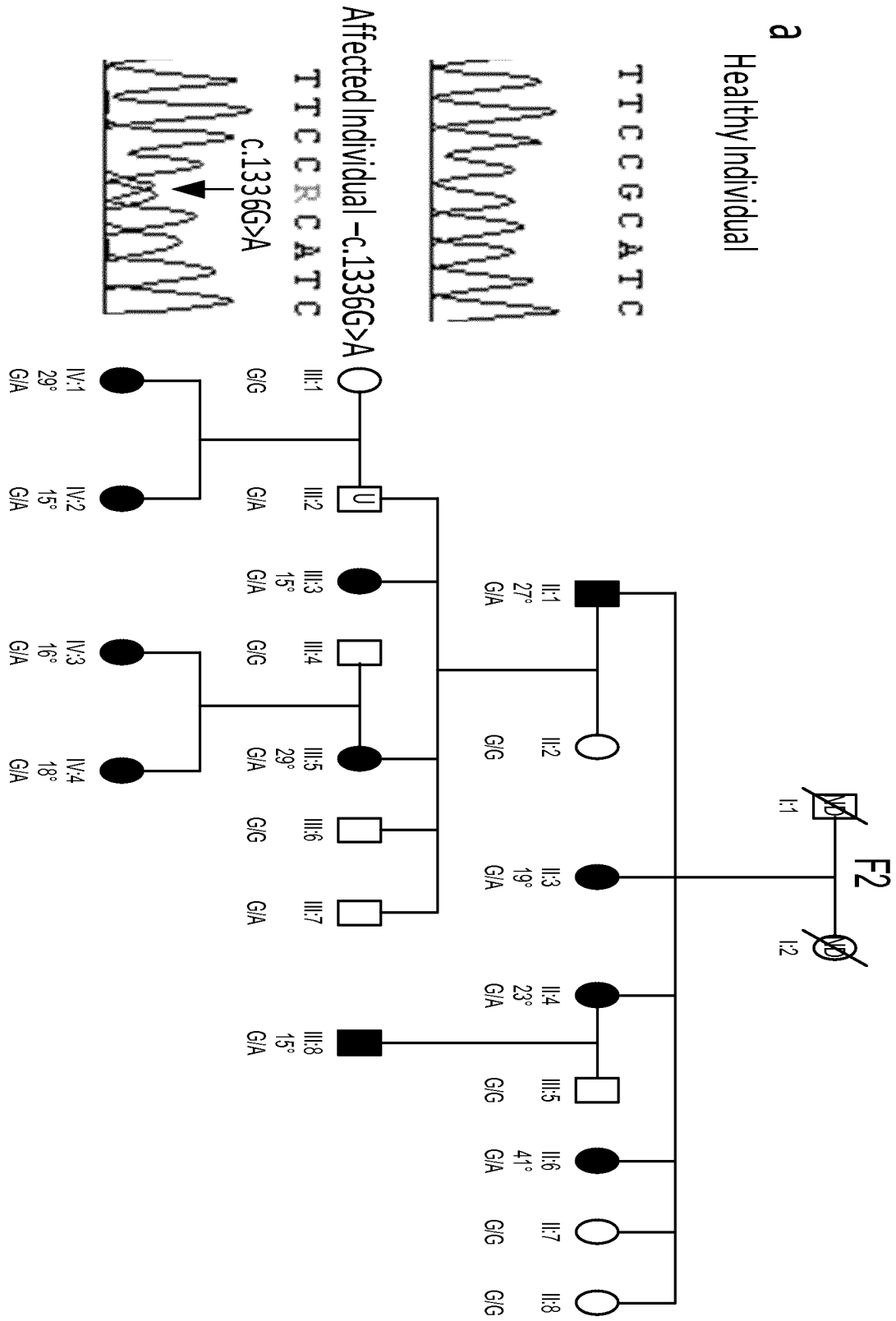
a
Healthy Individual
T T C C G C A T C
Affected Individual –c.1336G>A
T T C C R C A T C
c.1336G>A
I:1
I:2
F2
II:1 27° G/A
II:2 G/G
II:3 19° G/A
II:4 23° G/A
II:5 G/G
II:6 41° G/A
II:7 G/G
II:8 G/G
III:1 G/G
III:2 G/A
III:3 15° G/A
III:4 G/G
III:5 29° G/A
III:6 G/G
III:7 G/A
III:8 15° G/A
IV:1 29° G/A
IV:2 15° G/A
IV:3 16° G/A
IV:4 18° G/A

Figure 1a

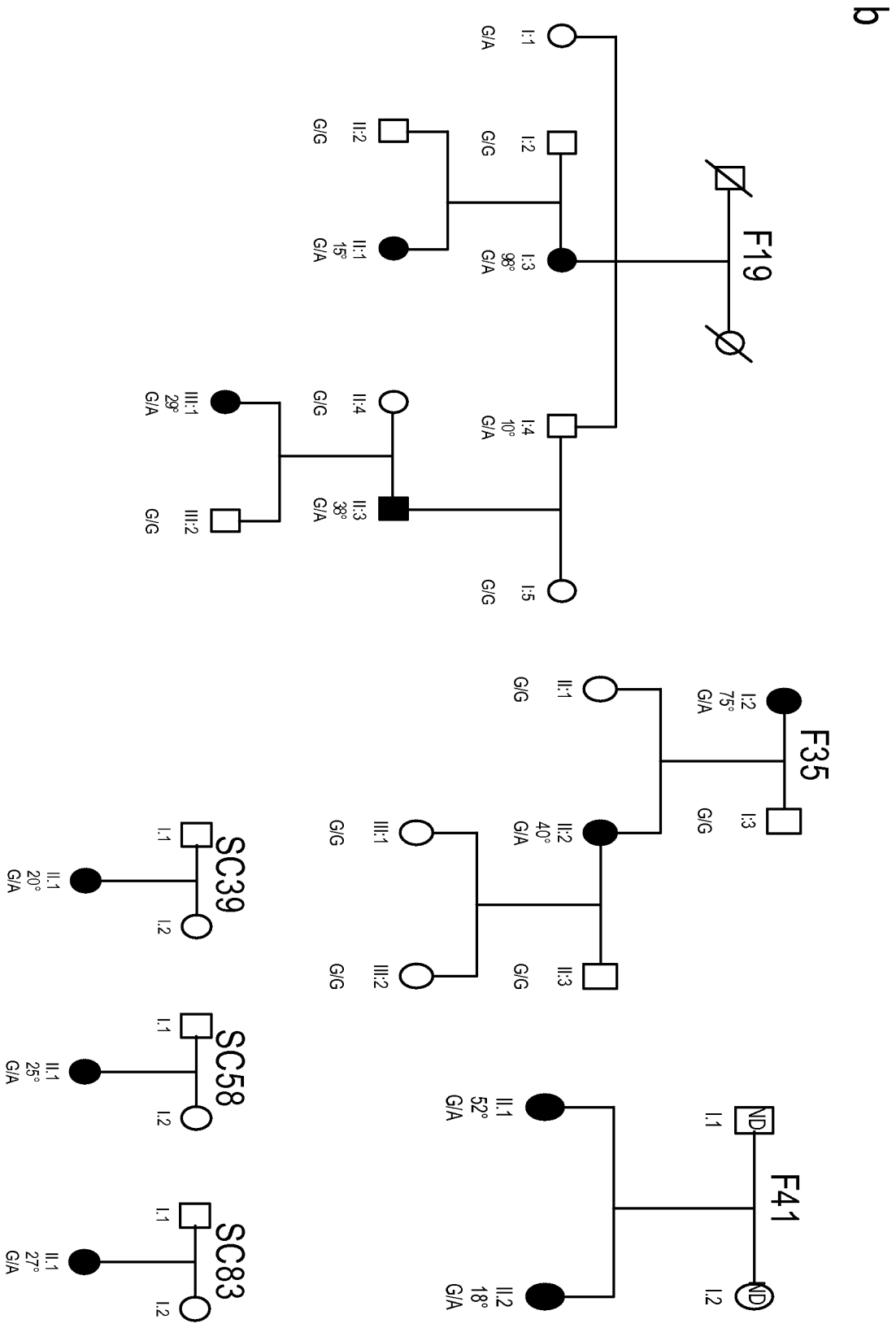
b
F19
I:1
G/A
I:2
G/G
I:3
98°
G/A
I:4
10°
G/A
I:5
G/G
II:1
15°
G/A
II:2
G/G
II:3
38°
G/A
II:4
G/G
III:1
29°
G/A
III:2
G/G
F35
I:2
75°
G/A
I:3
G/G
II:1
G/G
II:2
40°
G/A
II:3
G/G
III:1
G/G
III:2
G/G
F41
I.1
ND
I.2
ND
II.1
52°
G/A
II.2
18°
G/A
SC39
I.1
I.2
II.1
20°
G/A
SC58
I.1
I.2
II.1
25°
G/A
SC83
I.1
I.2
II.1
27°
G/A

Figure 1b

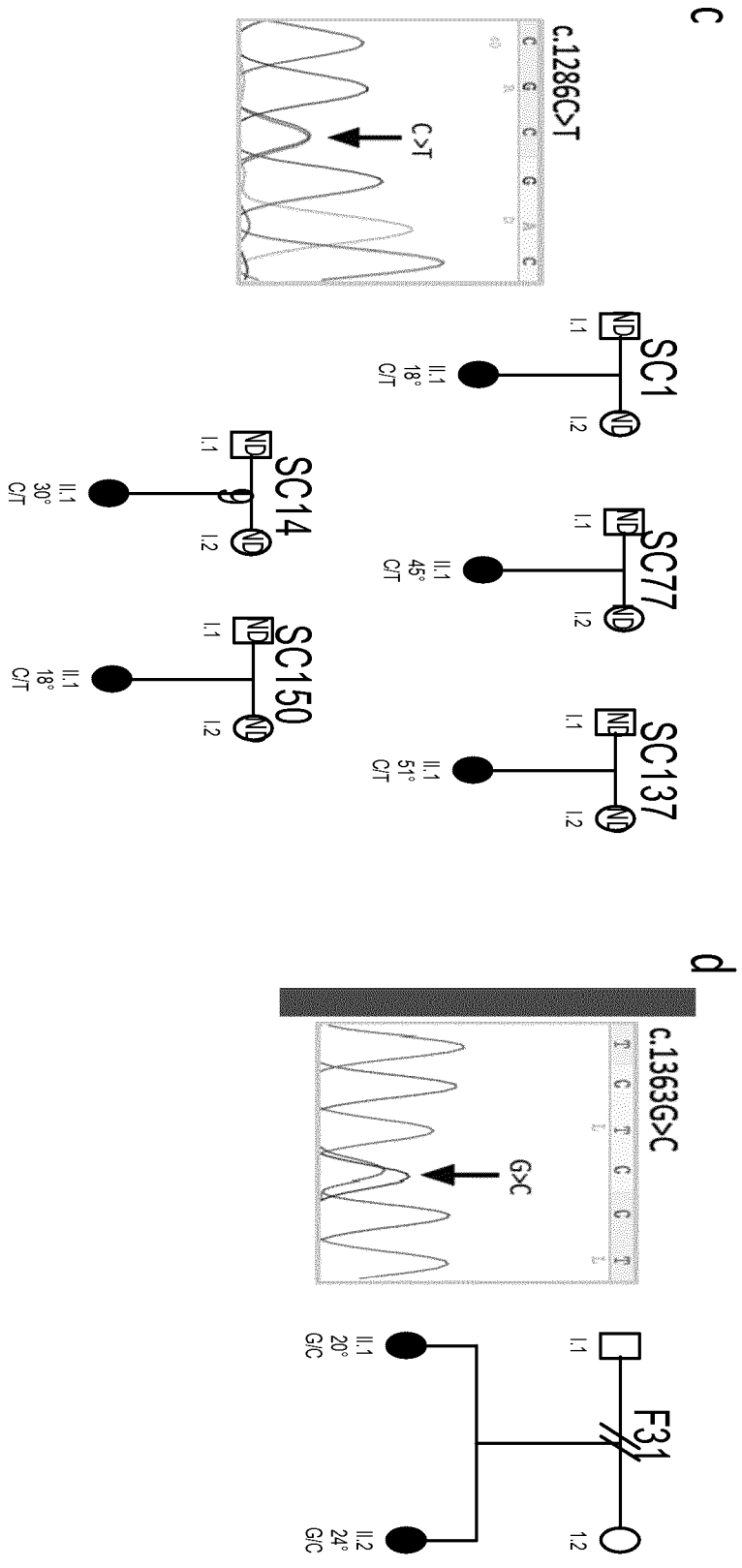

Figure 1 c et 1d

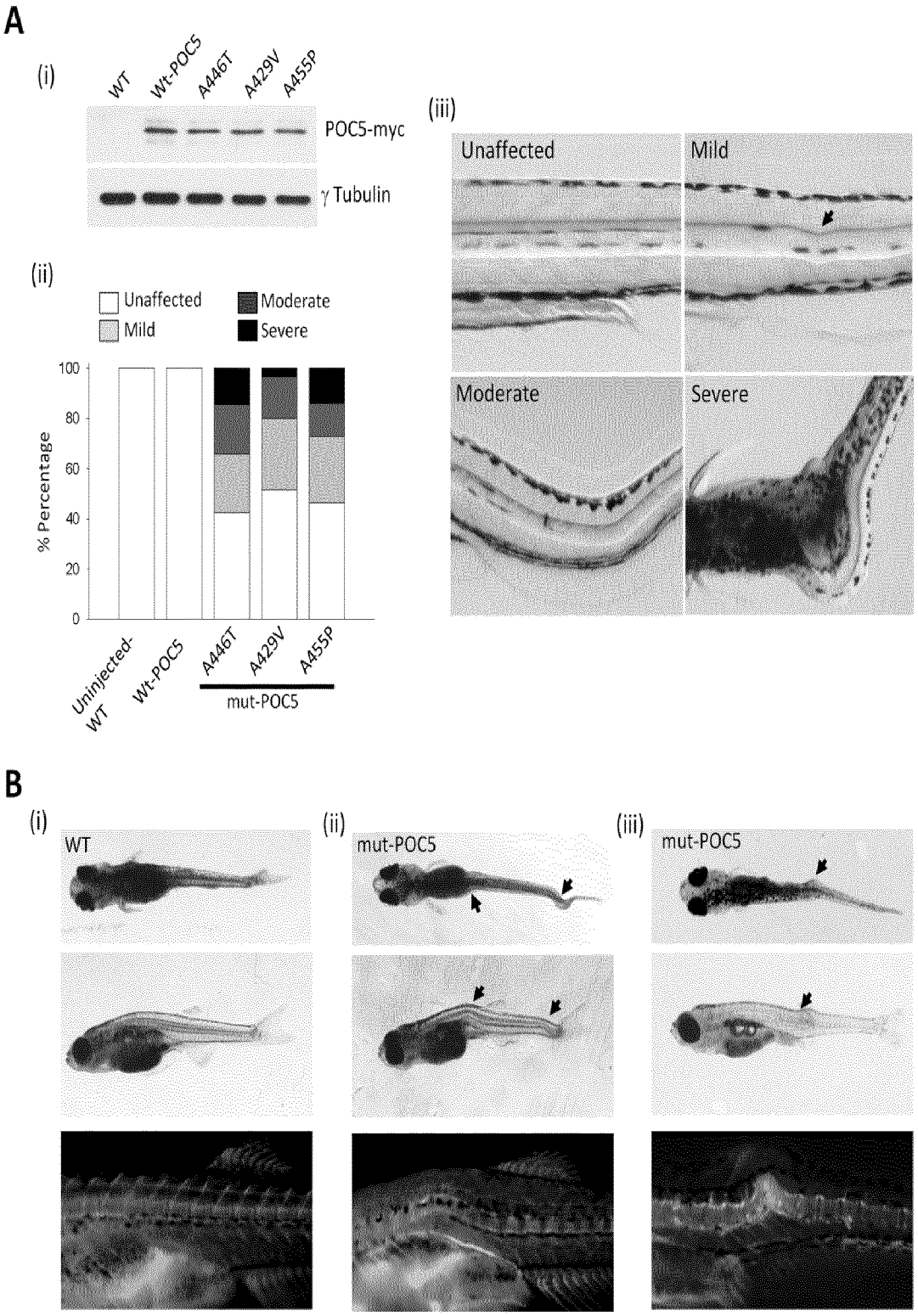
A
(i)
WT
Wt-POC5
A446T
A429V
A455P
POC5-myc
γ Tubulin
(ii)
Unaffected
Moderate
Mild
Severe
% Percentage
100
80
60
40
20
0
Uninjected-WT
Wt-POC5
A446T
A429V
A455P
mut-POC5
(iii)
Unaffected
Mild
Moderate
Severe
B
(i)
WT
(ii)
mut-POC5
(iii)
mut-POC5

Figure 2

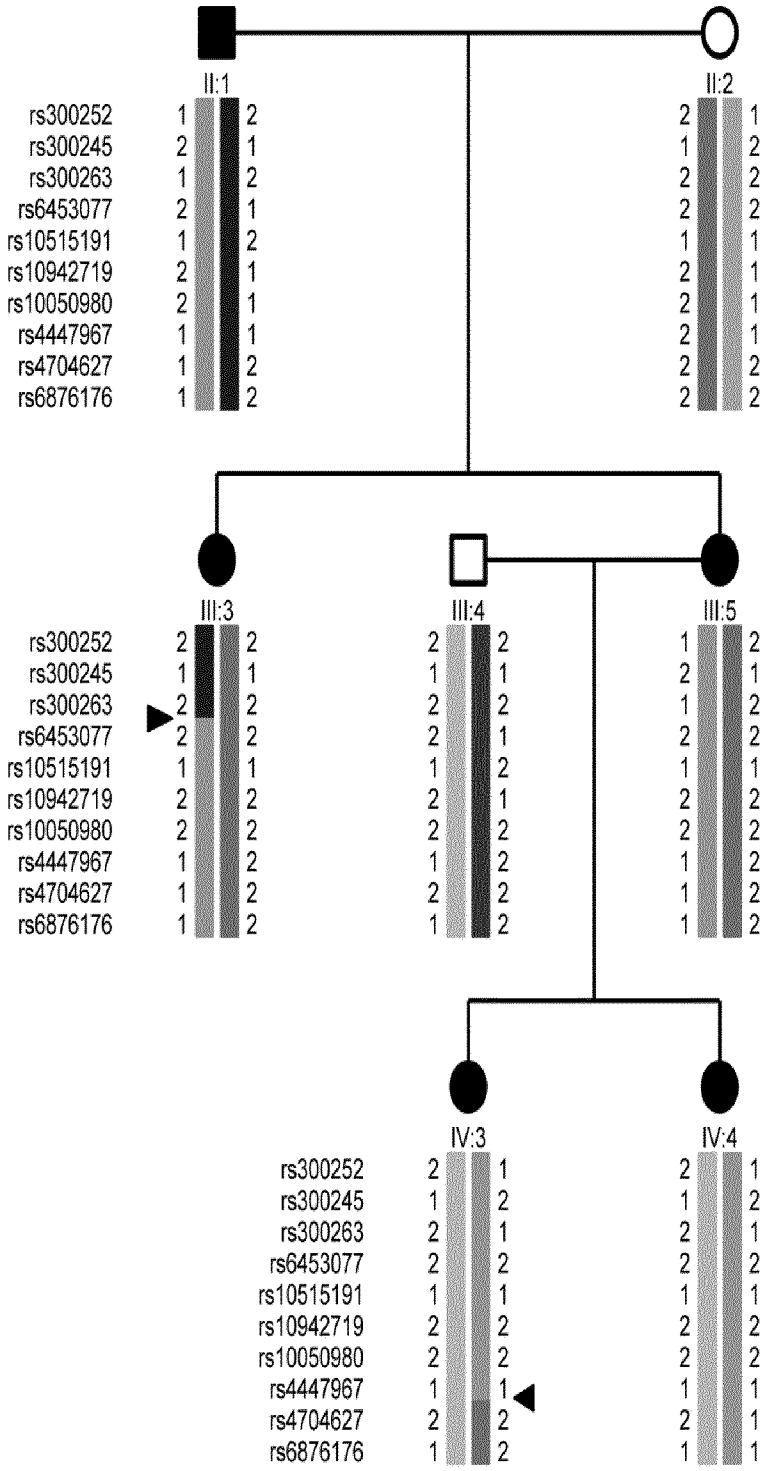
II:1
II:2
rs300252 1 2 2 1
rs300245 2 1 1 2
rs300263 1 2 2 2
rs6453077 2 1 2 2
rs10515191 1 2 1 1
rs10942719 2 1 2 1
rs10050980 2 1 2 1
rs4447967 1 1 2 1
rs4704627 1 2 2 2
rs6876176 1 2 2 2
III:3
III:4
III:5
rs300252 2 2 2 2 1 2
rs300245 1 1 1 1 2 1
rs300263 2 2 2 2 1 2
rs6453077 2 2 2 1 2 2
rs10515191 1 1 1 2 1 1
rs10942719 2 2 2 1 2 2
rs10050980 2 2 2 2 2 2
rs4447967 1 2 1 2 1 2
rs4704627 1 2 2 2 1 2
rs6876176 1 2 1 2 1 2
IV:3
IV:4
rs300252 2 1 2 1
rs300245 1 2 1 2
rs300263 2 1 2 1
rs6453077 2 2 2 2
rs10515191 1 1 1 1
rs10942719 2 2 2 2
rs10050980 2 2 2 2
rs4447967 1 1 1 1
rs4704627 2 2 2 1
rs6876176 1 2 1 1

Figure 3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2014/063780

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application and necessary to the claimed invention, the international search was carried out on the basis of:

   a. (means)

   ☐ on paper

   ☒ in electronic form

   b. (time)

   ☒ in the international application as filed

   ☐ together with the international application in electronic form

   ☐ subsequently to this Authority for the purpose of search

2. ☐ In addition, in the case that more than one version or copy of a sequence listing and/or table relating thereto has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2014/063780

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	BIO-RAD: "iProof High-Fidelity PCR Kit", , 2008, pages 1-2, XP002712295, Retrieved from the Internet: URL:http://www.bio-rad.com/webroot/web/pdf/lsr/literature/10002300B.pdf [retrieved on 2013-09-05]	8
A	the whole document	1-7,9-11

X	J. AZIMZADEH ET AL: "hPOC5 is a centrin-binding protein required for assembly of full-length centrioles", EUROPEAN JOURNAL OF CELL BIOLOGY, vol. 59, no. 2, 6 April 2009 (2009-04-06), pages 425-114, XP055077785, ISSN: 0171-9335, DOI: 10.1083/jcb.93.3.938	8,9
A	the whole document p. 112, para. bridging col. 1-2	1-7,10, 11

	-/--	

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
5 August 2014	13/08/2014

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Sauer, Tincuta

2

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2008/131877 A1 (WISE CAROL [US]) 5 June 2008 (2008-06-05)	10,11
A	the whole document para. 7, claims 1-13	1-9

Y	EP 2 573 173 A1 (UNIV GIESSEN JUSTUS LIEBIG [DE]) 27 March 2013 (2013-03-27)	10,11
A	the whole document para. 23-27	1-9

A	PATRICK EDERY ET AL: "New disease gene location and high genetic heterogeneity in idiopathic scoliosis", EUROPEAN JOURNAL OF HUMAN GENETICS, vol. 19, no. 8, 1 August 2011 (2011-08-01) , pages 865-869, XP055077791, ISSN: 1018-4813, DOI: 10.1038/ejhg.2011.31 the whole document	1-11

2

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2014/063780

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2008131877 A1	05-06-2008	US 2008131877 A1	05-06-2008
		US 2010143928 A1	10-06-2010
EP 2573173 A1	27-03-2013	EP 2573173 A1	27-03-2013
		WO 2013045480 A1	04-04-2013