5-29-2013

Developmental dysplasia of the hip: Linkage mapping and whole exome sequencing identify a shared variant in CX3CR1 in all affected members of a large multi-generation family.

George J Feldman
*Thomas Jefferson University, george.feldman@jefferson.edu*

Javad Parvizi
*Thomas Jefferson University, Javad.Parvizi@jefferson.edu*

Mark Levenstien
*The Children's Hospital of Philadelphia*

Kathryn Scott
*University of Pennsylvania*

Jill A Erickson
*University of Utah*

Let us know how access to this document benefits you

Follow this and additional works at: [http://jdc.jefferson.edu/orthofp](http://jdc.jefferson.edu/orthofp)

Part of the [Orthopedics Commons](http://jdc.jefferson.edu/orthofp)

Recommended Citation
Feldman, George J; Parvizi, Javad; Levenstien, Mark; Scott, Kathryn; Erickson, Jill A; Fortina, Paolo; Devoto, Marcella; and Peters, Christopher L, "Developmental dysplasia of the hip: Linkage mapping and whole exome sequencing identify a shared variant in CX3CR1 in all affected members of a large multi-generation family." (2013). *Department of Orthopaedic Surgery Faculty Papers*. Paper 44. [http://jdc.jefferson.edu/orthofp/44](http://jdc.jefferson.edu/orthofp/44)
Authors
George J Feldman, Javad Parvizi, Mark Levenstien, Kathryn Scott, Jill A Erickson, Paolo Fortina, Marcella Devoto, and Christopher L Peters
Developmental Dysplasia of the Hip: Linkage Mapping and Whole Exome Sequencing Identify a Shared Variant in CX3CR1 in All Affected Members of a Large Multi-Generation Family

Abstract:
Developmental Dysplasia of the Hip (DDH) is a debilitating condition characterized by incomplete formation of the acetabulum leading to dislocation of the femur, suboptimal joint function, and accelerated wear of the articular cartilage resulting in arthritis. DDH affects 1 in 1000 newborns in the United States with well defined “pockets” of high prevalence in Japan, Italy and other Mediterranean countries. Although reasonably accurate for detecting gross forms of hip dysplasia, existing techniques fail to find milder forms of dysplasia. Undetected hip dysplasia is the leading cause of osteoarthritis of the hip in young individuals causing over 40% of cases in this age group.

A 72 member, four generation affected family has been recruited, and DNA from its members retrieved. Genome-wide linkage analysis revealed a 2.561 Mb candidate region (38.7-41.31 Mb from the p term of chromosome 3) co-inherited by all affected members with a maximum LOD
score of 3.31. Whole exome sequencing and analysis of this candidate region in four severely affected family members revealed one shared variant, rs3732378, that causes a threonine (polar) to methionine (non-polar) alteration at position 280 in the trans-membrane domain of CX3CR1. This mutation is predicted to have a deleterious effect on its encoded protein which functions as a receptor for the ligand fractalkine. By Sanger sequencing this variant was found to be present in the DNA of all affected individuals and obligate heterozygotes. CX3CR1 mediates cellular adhesive and migratory functions and is known to be expressed in mesenchymal stem cells destined to become chondrocytes. A genetic risk factor that might to be among the etiologic factors for the family in this study has been identified, along with other possible aggravating mutations shared by 4 severely affected family members. These findings might illuminate the molecular pathways affecting chondrocyte maturation and bone formation.
Developmental Dysplasia of the Hip: Linkage Mapping and Whole Exome Sequencing Identify a Shared Variant in CX3CR1 in All Affected Members of a Large Multi-Generation Family

George J Feldman,1 Javad Parvizi,1,2 Mark Levenstien,3 Kathryn Scott,6 Jill A Erickson,7 Paolo Fortina,6
Marcella Devoto ,3,4,5 Christopher L Peters7

1 Thomas Jefferson University Division of Orthopaedic Research, Philadelphia PA, USA
2 Rothman Institute of Orthopaedics
3 Division of Genetics, The Children’s Hospital of Philadelphia, Philadelphia, PA, USA
4 Department of Pediatrics and Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
5 University of Rome “La Sapienza”, Department of Molecular Medicine, Rome, Italy
6 Department of Cancer Biology, Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, PA 19107, USA
7 University of Utah, University Orthopaedic Center, Salt Lake City, UT, USA
Pedigree of 71 member affected family. Filled in symbols denote individuals with 3 or more signs of DDH. Symbols containing question marks denote individuals with 1 or 2 signs of DDH.

236x93mm (300 x 300 DPI)
Results of multipoint model based genome wide linkage analysis for affected family for chromosome 3. No other chromosome showed a positive LOD score. Gray line between 0 and 4 on the y axis denotes a LOD of 3.0.

285x220mm (300 x 300 DPI)
Pedigree of 71 member affected family showing those members (*) whose DNA had the rs3732378 variant.

95x38mm (300 x 300 DPI)
Overview of Susceptibility Variant Search Strategy

182x224mm (300 x 300 DPI)
### Table 1. Shared Non-Synonymous SNPs in Candidate Region.

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Position</th>
<th>Chr_Pos</th>
<th>Gene</th>
<th>Region</th>
<th>Type</th>
<th>Dbsnp</th>
<th>Sift Assessment</th>
<th>Polyphen 2 Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr3</td>
<td>38350519</td>
<td>chr3_38350519</td>
<td>SLC22A14</td>
<td>CDS</td>
<td>snv</td>
<td>rs818818</td>
<td>tolerated</td>
<td>benign</td>
</tr>
<tr>
<td>chr3</td>
<td>38350543</td>
<td>chr3_38350543</td>
<td>SLC22A14</td>
<td>CDS</td>
<td>snv</td>
<td>rs818817</td>
<td>tolerated</td>
<td>benign</td>
</tr>
<tr>
<td>chr3</td>
<td>38645420</td>
<td>chr3_38645420</td>
<td>SCN5A</td>
<td>CDS</td>
<td>snv</td>
<td>rs1805124</td>
<td>tolerated</td>
<td>benign</td>
</tr>
<tr>
<td>chr3</td>
<td>38739574</td>
<td>chr3_38739574</td>
<td>SCN10A</td>
<td>CDS</td>
<td>snv</td>
<td>rs6595241</td>
<td>tolerated</td>
<td>benign</td>
</tr>
<tr>
<td>chr3</td>
<td>39229000</td>
<td>chr3_39229000</td>
<td>XIRP1</td>
<td>CDS</td>
<td>snv</td>
<td>rs6805248</td>
<td>tolerated</td>
<td>benign</td>
</tr>
<tr>
<td>chr3</td>
<td>39307162</td>
<td>chr3_39307162</td>
<td>CX3CR1</td>
<td>CDS</td>
<td>snv</td>
<td>rs37322378</td>
<td>damaging</td>
<td>possibly damaging</td>
</tr>
<tr>
<td>chr3</td>
<td>41278119</td>
<td>chr3_41278119</td>
<td>CTNNB1</td>
<td>CDS</td>
<td>snv</td>
<td>rs77750814</td>
<td>tolerated</td>
<td>benign</td>
</tr>
<tr>
<td>chr3</td>
<td>41925398</td>
<td>chr3_41925398</td>
<td>ULK4</td>
<td>CDS</td>
<td>snv</td>
<td>rs1052501</td>
<td>tolerated</td>
<td>benign</td>
</tr>
<tr>
<td>chr3</td>
<td>41996136</td>
<td>chr3_41996136</td>
<td>ULK4</td>
<td>CDS</td>
<td>snv</td>
<td>rs2272007</td>
<td>tolerated</td>
<td>benign</td>
</tr>
</tbody>
</table>

Shared Non-Synonymous SNPs in Chromosome 3 Candidate region and their predicted effect on protein function

270x152mm (300 x 300 DPI)
<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Variants</th>
<th>Description</th>
<th>Frequency</th>
<th>Value</th>
<th>Delta</th>
<th>Shared Variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromosome</td>
<td>Position</td>
<td>Ref</td>
<td>Alt</td>
<td>Minor Allele</td>
<td>Major Allele</td>
<td>Genotype</td>
</tr>
<tr>
<td>------------</td>
<td>----------</td>
<td>-----</td>
<td>-----</td>
<td>-------------</td>
<td>-------------</td>
<td>----------</td>
</tr>
<tr>
<td>3</td>
<td>100000</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>3</td>
<td>200000</td>
<td>A</td>
<td>G</td>
<td>A</td>
<td>G</td>
<td>A</td>
</tr>
<tr>
<td>3</td>
<td>300000</td>
<td>G</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>G</td>
</tr>
<tr>
<td>3</td>
<td>400000</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>C</td>
</tr>
</tbody>
</table>

Variants outside chr 3 candidate region shared by 4 severely affected family members

268x38mm (300 x 300 DPI)
### Depth of Coverage for Whole Exome Sequencing

Sample numbers 001-01, 001-03, 001-09, 001-10 corresponds to sequencing results for individual 1, 3, 9, and 10 respectively in pedigree figure 1.

<table>
<thead>
<tr>
<th>Sample</th>
<th># Reads</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Mapped</td>
</tr>
<tr>
<td>001-01</td>
<td>89565560</td>
<td>88005637</td>
</tr>
<tr>
<td>001-03</td>
<td>103675132</td>
<td>101961295</td>
</tr>
<tr>
<td>001-09</td>
<td>101330158</td>
<td>99388838</td>
</tr>
<tr>
<td>001-10</td>
<td>94357648</td>
<td>92917554</td>
</tr>
</tbody>
</table>

243x199mm (300 x 300 DPI)
Haplotypes on Chromosome 3 of all affected individuals with 3 or more signs of DDH and obligate heterozygotes. Individuals 47 and 33 show crossovers which indicate the proximal and distal boundaries of the candidate region.

276x240mm (300 x 300 DPI)
Additional Sequencing Metrics for 4 Affected Family Members
Title: Developmental Dysplasia of the Hip: Linkage Mapping and Whole Exome Sequencing Identify a Shared Variant in CX3CR1 in All Affected Members of a Large Multi-Generation Family

George J Feldman,1 Javad Parvizi,1,2 Mark Levenstien,3 Kathryn Scott,6 Jill A Erickson,7 Paolo Fortina,6 Marcella Devoto,3,4,5 Christopher L Peters7

1 Thomas Jefferson University Division of Orthopaedic Research, Philadelphia PA, USA
2 Rothman Institute of Orthopaedics
3 Division of Genetics, The Children’s Hospital of Philadelphia, Philadelphia, PA, USA
4 Department of Pediatrics and Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
5 University of Rome “La Sapienza”, Department of Molecular Medicine, Rome, Italy
6 Department of Cancer Biology, Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, PA 19107, USA
7 University of Utah, University Orthopaedic Center, Salt Lake City, UT, USA

Keywords: Developmental Dysplasia of the Hip, linkage analysis, whole exome sequencing

Introduction

Developmental Dysplasia of the Hip (DDH) is a debilitating condition characterized by incomplete formation of the acetabulum leading to dislocation of the femur, suboptimal joint
function, and accelerated wear of the articular cartilage resulting in arthritis. (1) DDH affects 1 in 1000 newborns in the United States with well defined “pockets” of high prevalence in Japan, Italy and other Mediterranean countries. (2) Because of its high prevalence and undesirable consequences, screening programs involving manipulation of the femur or ultrasound imaging of the hip in infants are in place in most countries. (3) Although reasonably accurate for detecting gross forms of hip dysplasia, existing techniques fail to find milder forms of dysplasia. (4) Undetected hip dysplasia is the leading cause of osteoarthritis of the hip in young individuals causing over 40% of cases in this 20-40 year old age group(4) A sensitive and specific test for DDH has remained a desirable yet elusive goal in orthopaedic medicine for a long time.

DDH is a complex disorder having an etiology that is both environmental and genetic. (5-7) Environmental risk factors include breech presentation (with feet toward cervix), oligohydramnios (deficiency of amniotic fluid), and primiparity (first born). (3,8,9) Evidence for a genetic cause is well established and includes a higher concordance between monozygotic (41%) than dizygotic (2.8%) twins, and a 12-fold increase of DDH among first degree relatives of those affected by the disorder. (10,11) DDH appears to be transmitted in an autosomal dominant fashion in several families and, perhaps because of its complex etiology, exhibits incomplete penetrance.(12) Our hypothesis is that DDH affected individuals have mutation(s) or genetic variants that make them susceptible to the disorder. The goal of this study is to identify genetic susceptibility factors for DDH, and in so doing, lay the foundation for a genetic test to accurately identify susceptible newborns so that intervention with a device such as a Pavlik harness can be used to allow complete development of the acetabular labrum. Should this goal be attained our understanding of molecular pathways responsible for acetabular development will be enhanced.
Methods

Clinical diagnostic criteria

Before the initiation of this study, informed consent was obtained from each participant and approval was obtained from the Institutional Review Board of Thomas Jefferson University. We have recruited a large family from Utah that shows transmission of DDH through four generations and isolated DNA from 72 family members (Figure 1). Individual family members were diagnosed using clinical exams and supine anterior posterior radiographs of the pelvis. Imaging of the hips was evaluated by three orthopaedic surgeons, with clinical opinions of two additional surgeons elicited in any cases of disagreement. Shenton’s line (disrupted=affected), Center Edge Angle (<20 degrees=affected), Tonnis Angle (>10 degrees=affected), Extrusion Distance (>10mm=affected) and Femoral Neck angle were measured in each radiograph and compared to control values derived from 11 independent studies. Detailed clinical evaluation and criteria for diagnosis of this family has been described. In general those individual family members with one or two signs of DDH were deemed questionable, those with three or more signs affected.

Linkage analysis

Genome-wide genotyping was performed on DNA isolated from venous blood or cheek epithelial cells obtained from members of this family. The samples were analyzed using the Affymetrix GeneChip Human Mapping 250K Nsp Array following Affymetrix’s protocol. Briefly, total genomic DNA (250 ng per sample) at 50 ng/µl is digested with the restriction enzyme, Nsp I. Adaptors that recognize the cohesive 4 bp overhangs from the digestion reactions were ligated to the products. Following the ligation reaction, a PCR primer,
complementary to the ligated adaptor sequence, was used to amplify the DNA fragments. The PCR products are verified and analyzed on a 2% agarose gel. The products from the separate PCR reactions are pooled together and purified using AMPure XP beads (Beckman-Coulter Genomics, Danvers, MA). Purified products were quantitated using the NanoDrop ND-1000, fragmented and analyzed on a 4% agarose gel. The samples were then labeled, hybridized to the 250K Nsp Arrays, washed on a Fluidics Station 450, and scanned according to Affymetrix’s protocols on a GeneChip Scanner 3000G with Autoloader. Quality control assessment of the sample files was performed using Genotyping Console v4.0.

Using 261,961 SNPs as a starting point, only autosomal markers were used because male to male transmission within the pedigree ruled out X-linked inheritance. Individual samples were removed if the proportion of missing SNPs was greater than 0.1. The following filters were then applied 1) SNPs with a missing genotype rate greater than 0.01 were excluded. 2) SNPs with a minor allele frequency less than 0.1 were excluded. 3) SNPs with a Mendelian error rate greater than 0 (as defined by PLINK version 1.07 (24) were excluded. Using the HapMap Phase 2 CEU population, SNPs were further pruned to remove those in high linkage disequilibrium. Specifically, we used a method that considered SNPs pairwise in a sliding window (50 base pairs wide and 5 base pair step size), and removed one from each pair with a correlation coefficient of 0.1 or more. Finally, we excluded SNPs showing additional Mendelian inconsistencies detected by MERLIN version 1.12 (25). Thus the final number of SNPs analyzed was 18,778.

Model-based and model-free multipoint LOD score analysis were performed using the software MERLIN. Because of the size of the pedigree and the limitations of MERLIN, genome-
wide analysis was performed by including only the affected individuals and their parents. Once a likely candidate region was identified, analysis of the informative markers in the candidate region was repeated using LINKMAP (26) in the entire pedigree. In both cases, an affected-only model-based analysis was run assuming a dominant model with no phenocopies and a very low disease allele frequency of 0.0001. All individuals with three or more signs of disease were coded as affected. All individuals with two signs or less, and unaffected individuals who are in the blood line (not married into the family) were coded as unknown, while unaffected individuals who married into the family were coded as unaffected.

To test whether the linkage signal on chromosome 3p can be explained by the variant identified by exome sequencing, linkage analysis conditional on rs3732378 was performed by means of the approach described by Sun and Cox including informative markers in the candidate region. (37) This approach uses the T2 statistical test which is often used for studies of complex traits. It utilizes multiple SNP markers simultaneously and considers the effects of multiple disease susceptibility loci. The rationale for this approach is that if a variant in a candidate linkage region influences susceptibility to the disease, there should not be additional evidence for linkage when conditioning on the genotypes at the candidate variant. The software Stepc (http://genemed.bsd.uchicago.edu/old/software.html) implements this approach as a test of the null hypothesis that the variant tested is a disease susceptibility variant. A z-score is calculated before and after conditioning on the candidate SNP to evaluate the evidence for linkage in a given region. A non-significant result after conditioning, which indicates no residual evidence for linkage, is in agreement with the hypothesis that the SNP tested is a disease susceptibility variant.
Four severely affected family members with three or more signs of the disorder were selected for whole exome sequencing and for analysis of shared exonic variants. Control reference sequence was derived from the 1000 Genome project (http://www.1000genomes.org) and from the GRCh37 assembly of NCBI. Exome capture on each individual was performed using SureSelect Human All Exome kit v.1 (Agilent technologies) and sequencing was performed using a Genome Analyzer IIx (Illumina Inc). Sequence reads were mapped to the reference genome using the Burrows Wheller Aligner Tool (BWA). An average of 97.25 million reads were available for each individual of which 92.5% aligned to the targeted region. An average of 91.6% targeted regions were covered at a read depth of at least 20x (Figure S1). Ts/Tv ratios were 2.1, 2.099, 2.171 and 2.098 for individuals 1, 3, 9 and 10 respectively. See figure S3 for additional sequencing metrics.

The list of variants from all four affected individuals was filtered and the following variants were retained: 1) variants shared by all four affected individuals with a quality score greater than 20. Quality scores measure the probability that a base is called incorrectly. The quality score of a given base, Q, is defined by the equation $Q = -10 \log_{10}(e)$ where e is the estimated probability of the base call being wrong. Thus, a higher quality score indicates a smaller probability of error. For example a quality score of 20 represents an error rate of 1 in 100, with a corresponding call accuracy of 99%.

2) shared variants that mapped to the linkage candidate region 3) new mutations (those not found in dbSNP135), non-synonymous SNPs and splice site causing variants within subset. Non-synonymous variants shared by all four affected individuals were evaluated for functional effects on
the protein they encoded using PolyPhen2 and SIFT. Polyphen 2 uses structural features of the protein as well as information on evolutionary conservation to predict whether a mutation affects protein function. SIFT sorts nucleotide changes that are tolerated from those that are not based on their evolutionary conservation. Variants were screened for biological significance using by determining Genomic Evolutionary Rate Profiling (GERP) scores. GERP identifies evolutionarily constrained elements in multiple alignments by quantifying substitution deficits. These deficits represent substitutions that would have occurred if the element were neutral DNA, but did not occur because the element has been under functional evolutionary constraint.

Validation of the sequencing results in the four severely affected members and in other family members and in-laws (individuals married-into the family) was performed using Sanger DNA sequencing.

Filtering whole exome variants for relevance to bone using Ingenuity software: Starting with 1,518,628 variants spanning 20,018 genes, variants were: excluded that are 1) observed with an allele frequency greater than or equal to 0.1% of the genomes in the 1000 genomes project or 2) greater than or equal to 0.1% of the public Complete Genomics genomes or 3) greater than or equal to 0.1% of the National Heart Lung and Blood Institute Exome Sequencing Project (NHLBI ESP) exomes. Retained data were those 1) experimentally observed to be associated with a phenotype: pathogenic, possibly pathogenic, unknown significance or 2) established gain of function in the literature or gene fusions or inferred activating mutations by Ingenuity or 3) predicted gain of function by SIFT, a program that Sorts Intolerant From Tolerant variants (31) or 4) resided in a microRNA binding site or 5) were non-synonymous and not predicted to be innocuous by SIFT or 6) disrupt splice site or 7) were deleterious to a microRNA. Retained data also included those data that are associated with gain of function or were
compound heterozygous or homozygous or heterozygous or haploinsufficient or hemizygous and have call quality greater than 20 and occur in at least 4 of the case samples at the variant level in the case samples. Finally data passing the previously described screens were filtered for those variants that are known or predicted to directly affect bone formation.

Versions of software used in this analysis were: Ingenuity Variant Analysis version 1.2.20121206 Content versions: Ingenuity Knowledge Base (Vega 121015.000), COSMIC (v61), dbSNP (Build 137), 1000 Genome Frequency (v3), TargetScan (v6.1), EVS (ESP6500), JASPAR (10/12/2009), PhyloP hg18 (11/2009), PhyloP hg19 (01/2009), Vista Enhancer hg18 (10/27/2007), Vista Enhancer hg19 (12/26/2010), CGI Genomes (11/2011), SIFT (06/2012), BSIFT (06/2012), TCGA (5/14/2012)

Results

Description of the affected family: Diagnostic criteria for the family recruited for this study have been previously described in detail.(13) Briefly, this four generation, 72 member family from Utah shows transmission of DDH in a manner consistent with an autosomal dominant mode of inheritance with incomplete penetrance (Figure 1). Eleven patients had 3 or more signs of DDH and were considered to be unequivocably affected (see Methods). Thirteen individuals had one or two signs of DDH and had questionable diagnoses. Originally individual 27, an adolescent, was classified as affected. Subsequent review by a panel of orthopaedic surgeons changed this diagnosis to questionable. She was therefore classified as unknown for the purpose of linkage
analysis in the current study. Poor quality of DNA in affected individuals 2 and 6 resulted in missing genotype rate >0.1 and caused them to be excluded from the analysis.

**Linkage analysis:** Model-free and model-based analyses were performed using the linkage analysis program MERLIN. Both genome-wide analyses revealed an identical candidate region on chromosome 3 with both analyses producing a maximum LOD score of 3.31 in an interval delimited by SNPs rs4481097 and rs4626072 at 38.91 to 40.66 Mb from the p terminal end of chromosome 3 (Figure 2). There was no other region in the genome that yielded LOD scores greater than or equal to 0.55. Crossovers in individuals 47 and 33 determined the proximal and distal boundaries of the candidate region to be 38.7Mb and 41.31Mb respectively from the p terminus of chromosome 3. (figure S2). Model-based affected-only analysis in the candidate region using the software LINKMAP that allowed inclusion of all individuals resulted in a maximum LOD score of 3.28.

**Whole exome analysis** of four severely affected family members (individuals 1, 3, 9, and 10 in figure 1) was performed. Table 2 shows the 96 shared variants that mapped to the candidate region including 7 novel mutations. All 7 shared new mutations were intronic or found in untranslated regions except for the CATG/C indel found in the coding region of the *DLEC* gene. This indel was found to reside in a region of the *DLEC* gene that is not conserved in evolution. No shared splice-site mutations or variants within the candidate region were found. (Table 2) Shared non-synonymous variants that mapped to the chromosome 3 candidate region were further evaluated for functional significance by Polyphen 2 and SIFT. Of the 9 non-synonymous
shared SNPs in the candidate region, only one (rs3732378) was predicted to have a deleterious effect on its encoded protein (Table 1). Polyphen 2 predicted that this variant was “possibly damaging”. SIFT predicted that this variant was damaging (score of 0.01 with a range of 0 damaging-0.05 tolerated). This SNP is in the coding sequence of the chemokine (C-X3-C motif) receptor 1 (CX3CR1) which functions as a receptor for the ligand fractalkine which mediates cellular adhesive and migratory functions. Fractalkine or CX3CL1 is a 373-amino acid protein that has a chemokine domain located on top of a mucin-like stalk (27). Many cell types secrete CX3CL1 (34), which exists in both membrane-bound and soluble forms. The membrane-bound CX3CL1 can serve as an adhesion molecule for cells expressing the fractalkine receptor (CX3CR1) (28). The C to T missense variant in the CX3CR1 receptor causes a threonine (polar) to methionine (non-polar) alteration at position 280 (hence referred to as M280) in the trans-membrane domain of the protein.

A number of regulatory elements were found near rs3732378 from the ENCODE database. (32) The histone methylation sites H3K4Me1 and H3K4Me2 were found to overlap the position of the variant rs3732378 on chromosome 3 DNA. These sites are often located near regulatory elements. This DNA site is also involved in binding the transcription factors GATA2 and JunD in the K562 cell line. The RNA binding protein PABPC1 (polyA binding protein C1) also binds to this site in the GM12878 cell line. (data not shown). (32,33) This cell line is lymphoblastoid with a relatively normal karyotype. Both this cell line and cartilage are of mesodermal origin. PABPC1 mRNA and protein are expressed in the cartilage of 6-12 year old children and adolescents and this gene is expressed in hypertrophic growth plate cartilage. (38)
Validation by Sanger sequencing: The presence of SNP rs3732378 was validated by Sanger sequencing in the four affected individuals and in the DNA of all 13 other affected individuals and obligate heterozygotes within the family. Sanger sequencing was also performed to discover the presence of this SNP in individuals with fewer signs of the DDH as well as in seemingly unaffected in-laws (see figure 3). Individuals 8, 18, 32, 38, 44, 45, and 59 who appeared to have fewer signs of DDH were found to carry this variant. This variant was also present in the DNA of married-in individuals 21, 46, 14 and 64. (figure 3)

Conditional linkage analysis: To test whether SNP rs3732378 explained the linkage signal on chromosome 3p, we used the T2 test implemented in Stepc (see Sun et al,2002 ref 37) to evaluate the overall evidence for linkage in the critical region before and after conditioning on the genotype at rs3732378. While evidence for linkage before conditioning was significant with a z-score of 2.3849 (p-value=0.0085), analysis conditional on genotypes at rs3732378 resulted in a z-score of 0.9399 (p-value=0.1736), thus not rejecting the null hypothesis that rs3732378 is a disease susceptibility variant.

Screening shared variants outside candidate region for relevance to bone formation: Using Ingenuity software, variants were filtered for relevance to bone formation (see Methods for detailed description). After validation by Sanger sequencing, 6 variants were found to pass all filters. (Table 3) Four of these base changes were previously documented variants, 2 of them are novel mutations. Three of the 6 variants involve insertions or deletions of a triplet codon encoding an amino acid. One variant on chromosome 17p11.2 causes a frameshift mutation in RAI 1. Another novel mutation on chromosome 17q21.2, a deletion of 12 bases, resides in the
promoter region of one gene (TMEM99) and in the coding region of another (KRT10). Within this deleted segment are 3 CpG “islands” from 38975330-38975338 bp from the p terminal end of chromosome 17. (data not shown) One variant found in intron 1 of BMP8b on chromosome 1p34.2 is predicted to be activating. Activating mutations are those predicted by the SIFT algorithm to have the greatest potential for increasing reproductive fitness. Data for the predicted effects of the other known variants in this group was unavailable in either the SIFT or the Polyphen 2 databases.

Discussion

We have identified and retrieved DNA from one of the largest documented families showing inter-generational transmission of DDH. The goal of this study is to identify the molecular basis of the disease in this family using the approach of genome-wide linkage analysis together with whole exome sequencing illustrated schematically in Figure 4. By performing genome-wide linkage analysis, which makes no assumptions about where a mutation might reside, a 2.61 Mb candidate region on chromosome 3p22.2 has been identified with a high degree of certainty. However, classical linkage analysis requires unambiguous knowledge of who is affected in a given pedigree. Because this disorder is complex, with both environmental and genetic causes, and shows incomplete penetrance, not all individuals who appear unaffected will, in fact, be free of the disease allele. One method that we have chosen to address the problem of unknown, incomplete penetrance in our analysis is to define as affected only those family members whose diagnosis is certain (i.e. those having three or more signs of DDH). Penetrance for these affected individuals is irrelevant and unaffected individuals are scored as having an
unknown phenotype. Using this “affected only” model-based analysis assuming an autosomal dominant mode of inheritance, linkage analysis was performed and the results compared to model-free linkage analysis in which no assumption about the mode of inheritance was made. In both analyses the same candidate region on chromosome 3 resulted with no other genomic region producing lod scores higher than 0 in the model-based or 2.055 in the model-free analysis.

The maximum LOD score of 3.31 was found in an interval spanning 38.91 to 40.66 Mb from the p terminal end of chromosome 3 (Figure 2). This 2.61 Mb region has within it several candidate genes that might explain many of the signs seen in individuals affected with DDH. To further scrutinize these candidates, exomes of four severely affected family members were sequenced. Results of sequencing for the chromosome 3 candidate region shown in Table 1 reflect a spectrum of variants including synonymous and non-synonymous SNPs, variants that mapped to the untranslated region of various genes, deletions and SNPs in intronic and intergenic regions, as well as seven novel mutations. Novel mutations are those that appear in Table 2 without a dbSNP designation and are not found in the single nucleotide database 135. One of these novel mutations, an indel in the DLEC gene, was located in a part of the coding sequence found to be highly variable in evolution. The 6 other novel mutations occurred in intronic regions not related to mRNA splicing or in untranslated regions of their respective genes. None of these were in conserved genomic regions and it is difficult to evaluate their significance at this time.

In order to predict whether shared non-synonymous variants might be deleterious to the function of the respective genes, they were analyzed by both PolyPhen 2 and SIFT. Among the subset of 9 shared non-synonymous variants, only one was predicted to be detrimental to protein function. The variant rs3732378 causes a C to T transition in the coding region of the chemokine receptor (CX3CR1). This missense mutation converts a threonine (polar) to methionine (non-
polar) at position 280 in the trans-membrane domain. Polyphen 2, which uses structural features of the protein as well as information on evolutionary conservation, predicted that this mutation was “possibly damaging”. SIFT predicts that this mutation is damaging (score of 0.01 with a range of 0 damaging-0.05 tolerated). (Table 2)

Sequence conservation in evolution is often considered a measure of relative biologic importance. Genomic Evolutionary Rate Profiling (GERP) is a method for producing position-specific estimates of evolutionary constraint using maximum likelihood evolutionary rate estimation.(17) None of the candidate region variants (with the exception of rs3732378) were found to localize in regions of the genome that were highly evolutionarily conserved. At position 280 in CX3CR1, the threonine to methionine mutation, shared by all affected individuals in this family, has a GERP score of 3.98. Any positive score >2 is considered evidence of evolutionary constraint.(17) Threonine 280 is conserved in all mammals, the chicken and the D. rerio species of fish lending support to this analysis.(16)

The presence of this SNP was validated by Sanger sequencing in the four affected individuals and in the DNA of all 13 other affected individuals and obligate heterozygotes within the family. Sanger sequencing was also performed to discover the presence of this SNP in individuals with fewer signs of DDH as well as in seemingly unaffected married-in individuals (see figure 3). Individuals 8, 18, 32, 38, 44, 45, and 59 who appeared to have fewer signs of DDH were found to carry this variant. Interestingly, only 1 questionable child (59) was found to have the disease variant, while in the DNA of the other adolescents, who had been thought to have signs of DDH (27, 30, 31, 62 and 63), the mutation was missing. Five married in individuals (14, 21, 46, 58, 64) were also found to have this mutation in their DNA, supporting
the fact that the prevalence of this allele is very high in the indigenous Utah population from which this family originates. Analysis of genome-wide SNPs did not show any evidence of high levels of homozygosity in the genomes of the family members which would have been suggestive of inbreeding, and none of the spouses appeared to be related by SNP analysis.

The incomplete penetrance of DDH seen in some members of this family is understandable in light of the fact that DDH is a complex disorder with multiple genetic, epigenetic and environmental causes. Indeed the allele frequency of this relatively common variant varies from 3% in the Han Chinese to 8% in Caucasians and higher in other populations, including this one (Utah residents with Northern and Western European ancestry) whose allele frequency is 0.173, one of the highest recorded in the United States. (18) As found in most other multifactorial disorders, a number of individuals including the obligate heterozygotes and some “married-in” individuals in this family carry this allele but manifest no signs of DDH. This supports the idea that in this family this variant may be necessary but not sufficient by itself to cause the disorder. We are vigorously pursuing other etiologic factors that might explain the severity of DDH in this family and address this issue at a later point in this study.

Epigenetic factors that may cause altered gene expression of CX3CR1 include the histone modifications H3k4Me1 and Me2. These methylation sites, which are often located near regulatory elements, span this variant sequence in 7 cell lines. This DNA site also includes a binding site for the transcription factor Jun D in the K562 cell line. Jun D is required for BMP4 induced hematopoiesis in Xenopus. The significance of the presence of a PABPC1 (polyA binding protein C1) spanning this DNA segment is not clear.

While this study appears to have identified a significant genetic risk factor shared by all affected members of this family, it is not known how prevalent this variant will be in the overall
DDH patient population. We are in the process of validating this mutation in our DNA databank of over 30 dysplastics. Additionally, exome sequencing does not detect mutations that could exist in inter-genic regions within the chromosome 3 candidate region. Important regulatory sequences that affect gene expression could reside there. Not known also is whether the association of the CX3CR1 variant with DDH susceptibility will prove to be an etiologic factor. In vitro and in vivo experiments are currently underway to answer this question. Finally, pleiotropy with age related macular degeneration and cardiovascular disease cannot be ruled out or confirmed in this family.

This receptor, CX3CR1, which is involved in cell adhesion and migration, is known to play a role in chondrocyte maturation being expressed at higher levels in mesenchymal stem cells and becoming down regulated during the process of chondrocyte development. Recently in human mesenchymal stem cells that were induced to differentiate to chondrocytes, Djouad and his colleagues observed that CX3CR1 was present at higher levels in the stem cell and not expressed in the differentiated chondrocyte. (19) Additionally Cristino and her colleagues using a 3 dimensional hyaluronic acid scaffold have found that the chemokine receptors CXCR4 and CXCR5 were modulated during in vitro chondrogenic differentiation suggesting a role for the CXC class of chemokines in the differentiation and maturation of a cartilage-like structure in vitro.(20) The ability of mesenchymal cells to migrate and to receive regulatory signals from their environment via chemokine receptors may play a role in the developing structure of the cartilage anlage of the acetabulum.

The specific variant found in the DNA of all affected family members in this study is known to have significant biologic effects in a number of organ systems all related to CX3Cr1’s capacity to regulate cell migration or adhesion. Combadiere et al in 2007 found that
homozygosity for the CX3CR1 M280 allele was consistently more frequent in age-related macular degeneration (AMD) patients compared to controls observing that impaired migration of microglial cells occurred in affected patients. (21) They also found that chemotaxis of monocytes from individuals with homozygous M280 was impaired in the presence of bound Cx3CR1. Cx3CR1 is an HIV co-receptor as well as a leukocyte chemotactic/adhesion receptor for fractalkine and is overexpressed in the lymph nodes of HIV patients. Individuals with the M280 variant are more susceptible to HIV infection due to compromised migration of immune cells. (22) Finally, in cardiovascular disease the M280 variant provided a protective effect against carotid and coronary artery disease and stroke by decreasing cellular adhesion. (23)

All severely and moderately affected individuals were heterozygous for the M280 variant with a 50% complement of unmutated CX3CR1. The effect of this possibly decreased amount of fully functional receptor on the development of the acetabular labrum is unknown and is currently being investigated in animal models. The lack of effect of the variant on other organ systems in which it is present could be due to lack of sensitivity during a particular phase of development and to the fact that they can function within their biological limits on a possibly lower concentration of this protein. Additionally, the unique morphology of the developing human hip might account for the tissue specific phenotypic effect of this heterozygous mutant.

To explain the severity DDH in some individuals but not in others we hypothesized that severely affected individuals might have a second aggravating mutation. To determine which variants these might be, we searched throughout the human exome for variants shared by 4 severely affected individuals that caused various kinds of mutations (see Methods) and that resided in pathways related to bone formation. The variants described in Table 3, map outside
the chromosome 3 candidate region and for this reason are unlikely to be shared by all affecteds and by obligate heterozygotes as linkage analysis has shown that there is likely to be only one co-inherited genetic locus. One rare variant (rs77857664 allele frequency 0.361%), a T to C transition, was found in intron 1 of BMP8B, a gene known to be involved in development of the skeletal system. OXCT2 is a relatively small gene located entirely within intron 1 of BMP8B and has a function unrelated to bone metabolism. This T to C transition which also causes glutamic acid to glycine amino acid change in the OXCT2 protein is located in a transcriptionally active region, one with possible regulatory elements, as shown by a DNase I hypersensitivity cluster and an H3k27 acetylation mark found within this region by the Encode consortium. (32) SIFT analysis predicts this variant to have the potential to increase reproductive fitness.

Three of the observed variants shared by 4 severely affected family members involve insertion or deletions of repetitive amino acid(s). One of these in-frame variants in the transcription factor E2F4, on chromosome 16q22.1, is a novel mutation involving the insertion of the nucleotide triplet CAG in a highly conserved DNA sequence. This transcription factor participates in the transforming growth factor-beta superfamily mediated signaling pathway and deficiency of the E2F4 transcript leads to abnormal turbinate bone formation.(29) Another variant that causes an in-frame insertion of 2 leucines occurs is LTBP3 (latent transforming growth factor beta binding protein) on chromosome 11q12. This protein which forms a complex with transforming growth factor beta (TGF-beta), may be involved in their subcellular localization. Dabovic and his colleagues created an Ltbp3-null mutation in the mouse by gene targeting. (30) Mice homozygous for the mutation developed craniofacial malformations by day 10. At 2 months, there was a pronounced rounding of the cranial vault, extension of the mandible beyond the maxilla, and kyphosis. Between 6 and 9 months of age, mutant mice also developed
osteosclerosis and osteoarthritis. The pathologic changes were consistent with perturbed TGF-beta signaling in the skull and long bones. The third variant causes an in-frame deletion of glutamine in the RAI 1 on chromosome 17p11.2(retinoic acid inducible gene 1). Inactivation of RAI1 in the mouse recapitulates phenotypes seen in Smith Magenis syndrome. Bone related signs of this syndrome are deficient rib and nose cartilage formation.

The last two variants shared by 4 severely affected family members are potentially more disruptive. A deleted G in exon 2 of the RAI 1 gene causes a frameshift mutation at position 280 in the protein. Two individuals (3 and 9) are homozygous for this deletion. Finally the last variant found is a novel mutation, a deletion of a 12 bp segment in the promoter of the TMEM99 gene on chromosome 17q21.2. This segment of chromosome 17 located within exon 6 of the KRT10 gene co-encodes the repetitive amino acid motif glycine-histidine in the protein sequence of connective tissue protein, Keratin 10. Individual 1 is homozygous for this mutation the other 3 family members are heterozygous. While the function of TMEM99 is unknown it (and many other members of the TMEM gene family) are strongly expressed in cartilage.(35) This deleted promoter sequence encodes at least 3 predicted CpG methylation sites.(33) These CpG "islands" are often methylated in promoters of genes that are inactivated. (34) This finding might provide insight into the mechanism of incomplete penetrance often seen in familial DDH. This TMEM promoter is also the binding site of C/EBP beta, a transcription factor known to be involved in bone regulation. Lastly and interestingly this novel promoter mutation in TMEM maps to chromosome 17q21 very near the proximal border of the candidate region of another smaller DDH family analyzed by Feldman and his colleagues (36). Further validation of these novel mutations/variants is on-going.
In summary, a novel role is being proposed for an existing polymorphic nucleotide that is known to be biologically significant causing both beneficial and detrimental effects in the cardiovascular and ocular systems respectively. Gene dosage, timing of expression, and the unique morphology of the developing hip all create the micro-environment in which this variant acts, possibly explaining the phenotype of those with deficient labrum formation seen in DDH. Variants and new mutations that may contribute to the DDH phenotype severity have been found. Along with other genetic risk factors that we expect to be found in other families (since hip formation is a complex process involving the timed interaction of many proteins) these findings might illuminate the molecular pathways affecting chondrocyte maturation and bone formation.
Abstract

Developmental Dysplasia of the Hip (DDH) is a debilitating condition characterized by incomplete formation of the acetabulum leading to dislocation of the femur, suboptimal joint function, and accelerated wear of the articular cartilage resulting in arthritis. DDH affects 1 in 1000 newborns in the United States with well defined “pockets” of high prevalence in Japan, Italy and other Mediterranean countries. Although reasonably accurate for detecting gross forms of hip dysplasia, existing techniques fail to find milder forms of dysplasia. Undetected hip dysplasia is the leading cause of osteoarthritis of the hip in young individuals causing over 40% of cases in this age group. A sensitive and specific test for DDH has remained a desirable yet elusive goal in orthopaedics for a long time.

A 72 member, four generation affected family has been recruited, and DNA from its members retrieved. Genome-wide linkage analysis revealed a 2.61 Mb candidate region (38.7-41.31 Mb from the p term of chromosome 3) co-inherited by all affected members with a maximum LOD score of 3.31. Whole exome sequencing and analysis of this candidate region in four severely affected family members revealed one shared variant, rs3732378, that causes a threonine (polar) to methionine (non-polar) alteration at position 280 in the trans-membrane domain of CX3CR1. This mutation is predicted to have a deleterious effect on its encoded protein which functions as a receptor for the ligand fractalkine. By Sanger sequencing this variant was found to be present in the DNA of all affected individuals and obligate heterozygotes. CX3CR1 mediates cellular adhesive and migratory functions and is known to be expressed in mesenchymal stem cells destined to become chondrocytes. A genetic risk factor that might to be among the etiologic
factors for the family in this study has been identified, along with other possible aggravating mutations shared by 4 severely affected family members. These findings might illuminate the molecular pathways affecting chondrocyte maturation and bone formation.

REFERENCES


14. (genetics.bwh.harvard.edu/pph2/)


23
16. http://genome.ucsc.edu/cgi-bin/hgTracks?position=chr3:3930691239307412&hgsid=300630135&SNP135Common=pack&hgFind.matches=rs3732378,


accumulation is associated with cardinal features of age-related macular degeneration J Clin Invest. 2007;117(10):2920–2928.


24. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007 Sep;81(3):559-75.


