

PPP6R2 is a Novel Candidate Gene for Developmental Dysplasia of the Hip as Revealed by Whole Exome Sequencing of 29 Families of Danish Descent

Maja Dembic (✉ Maja.Dembic@rsyd.dk)

Odense Universitets Hospital <https://orcid.org/0000-0003-2505-6863>

Lars van Brakel Andersen

Odense University Hospital

Martin Jakob Larsen

Odense University Hospital

Inger Mechlenburg

Aarhus University Hospital

Kjeld Søballe

Aarhus University Hospital

Jens Michael Hertz

Odense University Hospital

Primary research

Keywords: DDH, whole exome sequencing, PPP6R2, mechanotransduction, cilia

Posted Date: November 8th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1038868/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background

Developmental dysplasia of the hip (DDH) is a common condition involving instability of the hip with multifactorial etiology. Early diagnosis and treatment are critical as undetected DDH is an important cause of long-term hip complications. Better diagnostics may be achieved through genetic methods, especially for patients with positive family history. Several candidate genes have been reported but the exact molecular etiology of the disease is yet unknown.

Results

In the present study, we performed whole exome sequencing of DDH patients from 29 families with at least two affected first-degree relatives. We identified *PPP6R2* as a novel DDH gene as two rare missense mutations were identified in three families co-segregating with disease. Mutational burden analysis across the families identified 455 candidate genes, with many genes involved in mechanotransduction, in particular the cilia, the cytoskeleton, the extracellular matrix, and the Notch pathway.

Conclusions

Here we report for the first time a previously uncorrelated gene with DDH, *PPP6R2*. Taken altogether, the data suggest a polygenic mode of inheritance for DDH, and we propose that impaired mechanotransduction is involved in the etiopathological mechanism.

Introduction

Developmental dysplasia of the hip (DDH, OMIM #142700) is a common condition characterized by abnormal hip development during infancy and early development, with a wide spectrum of phenotypic presentations [1]. Clinical diagnosis of DDH in older children presents challenges as early features are non-specific, and the disorder might be at an advanced stage by the time of diagnosis. However, early diagnosis and treatment is critical to achieve the best possible outcome. Early treatment consists in obtaining a stable hip joint with minimal complications as early as possible, which is usually achieved with a brace that secures the hip in a flexed and abducted position [2]. Undetected hip dysplasia increases the risk of long-term complications, including pain, secondary osteoarthritis and surgery [2, 3].

The global incidence of DDH is estimated to be between 0.1-6.6 per 1000 live births with high prevalence among the Asians, Caucasians, Mediterranean, and American populations [3]. The etiology of DDH is multifactorial including genetic, mechanical, and environmental factors. Some of the risk factors that increase the occurrence of DDH include: female gender, breech position at birth, fetal position, oligohydramnios, a large birth weight, primiparity, and swaddling [3]. The influence of genetics in DDH is well-established from studies on families with multiple individuals affected by DDH [4, 5], and monozygotic twins [6]. Multiple genes have been related to the disorder. These are mainly genes coding for connective tissue proteins, genes regulating osteogenesis and chondrogenesis, formation of joint structures, and chemokine receptors [3]. An autosomal dominant mode of inheritance with incomplete penetrance is most often considered for DDH, although, some studies proposed a multiple gene inheritance [7, 8]. Given the phenotypic and genotypic complexity of DDH, it is possible that DDH is caused by a combined action of multiple genes.

The etiopathogenesis of DDH has not been fully elucidated, and wide genetic studies are needed to resolve the issue of heritability, which would lead to better preventive strategies, especially in newborns with positive family history. We performed a comprehensive genetic analysis on 29 families where at least two relatives were affected by DDH. We investigated single nucleotide variants (SNVs), inserts and deletions (indels), and copy number variations (CNV) within single families and we performed a mutational burden analysis across the families. We identify novel candidate genes and propose a biological mechanism for the disease.

Results

Sample collection and clinical description

After obtaining written informed consent, we collected blood samples from 66 patients (54 (82%) females and 12 (18%) males; all of Danish descent) diagnosed with DDH, belonging to a total of 29 families. Each of the families had at least two affected individuals. One family presented with five affected first-degree relatives, five families had three affected first-degree relatives, while the remaining 23 families had two affected first-degree relatives (the pedigrees are shown in Additional file 1).

Whole exome sequencing

To investigate the possibility of a shared genetic cause of DDH in our cohorts, we performed whole exome sequencing (WES) of the affected individuals. We reached a mean sequencing coverage of 140.5X. On average, 94.2% of the target regions had a sequencing depth of at least 30X (Additional file 2). There was only one fallout sample with a very low coverage, which we decided to keep as it belonged to a three-member family (family 28), and we could still use at least two related individuals for effective variant identification.

We performed genetic analysis of SNVs, indels, and CNVs. The diagram of the general structure of the study is shown in Fig. 1. After alignment and joint variant calling within each family, the resulting files contained on average 50243 variants (Additional file 2). We further filtered to include only coding sequence and splicing variants, with high quality signal as assessed by base call accuracy of at least 99%, a variant allele frequency (VAF) in reads of 25% or more, and by manual check of each variant. DDH is known for incomplete penetrance and we decided not to sequence healthy relatives as controls as it may have confounding effects. As control reference, we used variant frequencies in the general population as listed in the Genome Aggregation Database (gnomAD), based on data from 125,748 exomes and 15,708 whole-genomes of unrelated individuals from all over the world. We kept only the variants with a minor allele frequency (MAF) equal or under 0.01, and with less than five reported homozygotes in the overall global population according to gnomAD.

Variant distribution and annotation

We found a total of 3332 rare and shared variants in 2798 genes, carried by related individuals across the 29 families. The great majority of the variants were missense (3025), followed by a relatively small number of stop-gain variants (75), splicing mutations (72), and frameshift variants (70) (Additional file 3). Inframe indels accounted for 66 cases, and the rest of the variants involved either 5'UTR premature start codons (9), initiator codons (4), stop losses (4), or disruptive inframe deletions (3).

On average, we found 141 rare shared variants in families with only two members. As expected, families with three members shared a smaller number of variants on average (45) than two-member families (Fig. 2), suggesting a certain amount of false positives. Family 34 had five members and a relatively high number of rare variants in common between all the affected individuals: 14.

Surprisingly, there were two families, family 2 (two members), and family 35 (three members), which both had only two rare variants shared between the affected relatives. These were actually the same two missense mutations in the *PPP6R2* gene: c.1421G>A and c.2345G>A (Table 1). The *PPP6R2* gene encodes a regulatory subunit for the serine/threonine protein phosphatase-6 catalytic subunit (PPP6C). Phosphoprotein phosphatases (PPP) are multimeric holoenzymes that modulate mitotic entry and progression; usually their substrate specificity is defined through association of the catalytic subunit with regulatory subunits [9]. We searched for additional variants affecting genes coding for regulatory and catalytic subunits of PPPs in our data and we found that family 1 also bears the same two variants in the *PPP6R2* gene as family 2 and 35. In total, the two *PPP6R2* SNVs were detected in seven individuals from three unrelated families. The two SNVs were always inherited together, suggesting a *cis* positioning on the same allele.

Table 1 Genes encoding for phosphoprotein phosphatases (PPP) subunits and carrying SNVs

Gene	Protein	Mutation nucleotide	Mutation protein	Effect	Family	CADD score	Number of tools predicting damaging
<i>PPP6R2</i>	Regulatory subunit to the PPP6C catalytic domain	NM_001242898.1:c.1421G>A	NP_001229827.1:p.Arg474His	Missense	Fam1, Fam2, Fam35	23.5	4 out of 6
<i>PPP6R2</i>	Regulatory subunit to the PPP6C catalytic domain	NM_001242898.1:c.2345G>A	NP_001229827.1:p.Gly782Asp	Missense	Fam1, Fam2, Fam35	14.76	2 out of 6
<i>ANKRD28</i>	Regulatory subunit to the PPP6C catalytic domain	NM_001195098.1c.-325C>T		5'UTR premature start codon	Fam32	7.624	NA
<i>PPP2R3C</i>	Regulatory subunit to the PPP2C catalytic domain	NM_017917.3:c.173G>A,	NP_060387.2:p.Arg58Gln,	Missense	Fam13	22.6	2 out of 6
<i>PPP2R3C</i>	Regulatory subunit to the PPP2C catalytic domain	NM_017917.3:c.29G>A	NP_060387.2:p.Arg10His	Missense	Fam33	22.3	2 out of 6
<i>PPP1R13B</i>	Regulatory subunit to the PPP1C catalytic domain	NM_015316.2:c.1283C>T	NP_056131.2:p.Pro428Leu	Missense	Fam27	24.3	3 out of 6
<i>PPP1R18</i>	Regulatory subunit to the PPP1C catalytic domain	NM_133471.3:c.215C>T	NP_597728.1:p.Pro72Leu	Missense	Fam5	22.2	1 out of 6
<i>PPP1R37</i>	Regulatory subunit to the PPP1C catalytic domain	NM_019121.1:c.2059G>A	NP_061994.1:p.Gly687Arg	Missense	Fam25	24.1	3 out of 6
<i>PPP1R9A</i>	Regulatory subunit to the PPP1C catalytic domain	NM_001166160.1:c.125A>T	NP_001159632.1:p.Glu42Val	Missense	Fam4	25.6	5 out of 6

List of genes encoding for PPP subunits carrying rare variants that co-segregated with DDH within single families. All these genes encode for regulatory subunits of PPP enzymes, among which the protein phosphatases PP1, PP2, and PP6. In the *PPP6R2* gene, the same two missense variants were found in seven individuals from three unrelated families. In family 2 and family 35 these two mutations were the only rare variants shared between the affected individuals of the family. The CADD score is used to evaluate the deleteriousness of a variant. A CADD score of over 20 indicates that a variant is among the 1% most deleterious substitution in the genome, and thus is considered damaging. Additionally, we used other six tools for evaluating a negative functional impact (SIFT, Polyphen, MutTaster, MutAssessor, FATHMM Pred, and FATHMM MKL). The results of a voting system from these six algorithms is shown. A higher number of tools predicting a variant damaging gives a higher score in the voting system. Abbreviations: PPP: phosphoprotein phosphatases; CADD: combined annotation dependent depletion.

Among other genes encoding for PPP subunits, we found a 5'UTR premature start codon variant in the *ANKRD28* gene, which is another regulatory subunit of PPP6C (Table 1); two families bearing two missense *PPP2R3C* SNVs, and four families with four different missense

variants in genes encoding PPP1C regulatory subunits: *PPP1R13B*, *PPP1R18*, *PPP1R37*, and *PPP1R9A*. Most of these variants were predicted to be somewhat damaging as their CADD score for deleteriousness was over 20.

We searched the literature for known genes associated with DDH (Additional file 4) to see if we found any known variants or known genes in our study group. We found novel variants in previously reported DDH candidate genes in humans *CDH7*, *CX3CR1*, *DACH1*, *ESR1*, *MYH10*, *NOTCH2*, *PCNT*, *POLE*, *TBX4*, and *TENM3*; while the following genes previously described in dogs had variants in our study group: *COL6A3*, *EVC2*, *INPP5D*, *IQCA1*, *KLH17*, *LAMA2*, *NWD1*, and *OTOG*.

CNV analysis

Analysis of CNVs did not detect any common structural variants in regions involving previously reported DDH candidate genes (Additional file 5). Only one CNV variant was detected in multiple families (family 1 and family 22), co-segregating with the disease, and it involved the duplication of exons 19 to 22 of the *ANKRD24* gene. *ANKRD24*, similarly to some PPP6 regulatory subunits, has multiple ankyrin repeats domains; however, its function is currently unknown.

Functional analysis of the variants and candidate genes

To identify the most deleterious variants that may cause DDH in our cohort, we used *in silico* predictions to filter and select the most damaging, rare, and relevant variants. We included shared variants within families, and used a more stringent filtering for variant frequency in the total population. We selected only variants with a MAF equal or under 0.01, less than 5 homozygotes, and 590 alternate allele counts in the general population, which would be the expected number assuming an autosomal dominant mode of inheritance and an average DDH incidence of 1:3400.

Among the identified 3332 rare and shared variants from our study group, 825 of them were predicted to be highly damaging that either had a loss-of-function (LoF) mutation, or were assigned a very high prediction score (5 or 6 algorithms predicting deleteriousness, or a CADD score of over 25) (Additional file 6). Four of these variants occurred in known DDH candidate genes in humans: *DACH1*: c.2050C>T, *MYH10*: c.423G>A, *NOTCH2*: c.6094C>A, and *TBX4*: c.805G>A [10-13], while the *EVC2*: c.3061C>T, the *OTOG*: c.6943C>T, and the *SHC3*: c.875A>T gene variants occur in candidate genes previously described in dogs [14, 15].

As most of the families had many very damaging variants, we used the PhoRank algorithm to filter further the variants in an attempt to select for the most relevant ones. In Fig. 3 is shown the distribution of very damaging variants across the families and according to PhoRank. The PhoRank algorithm assigns a score, from 0 to 1, to each gene according to closeness to a user-specified phenotype (hip dysplasia in this case). For most of the variants the median value of PhoRank was low (under 0.325). However, few variants had a very high score: 35 genes carrying 36 very damaging variants had a PhoRank score over 0.75, indicating closeness to the DDH phenotype (Fig. 3, while Additional file 7 shows the whole list).

Among these, we found very damaging mutations in three collagen genes: *COL12A1*, *COL2A1*, and *COL6A2*. Mutations in *COL2A1* are known to cause spondyloepiphyseal dysplasia and *COL2A1* has previously been related to osteoarthritis secondary to DDH [16]. We also identified very damaging variants in four other genes involved in chondrodysplasias: *BMPR1B*, *KIF22*, *MMP9*, and *NANS*. Furthermore, we found three known pathogenic mutations in the genes *MAN2B1*, *IDUA*, and *ABCA3*, of which the *IDUA*: c.1829-1G>A is a LoF mutation causes mucopolysaccharidosis of type 1 [17], an autosomal recessive disorder that commonly involves the hips, with similar clinical and radiological presentation as DDH [18]. The *MAN2B1* mutation is involved in mannosidosis alpha, while the *ABCA3* mutation in pulmonary disease.

Importantly, most of the families presented with more than one good candidate variant that co-segregated in the affected members, leading to the assumption of a multiple genetic contribution to the phenotype. On the other hand, families with three and five affected members (family 5, family 8, family 20, family 28, and family 34), even though they had very damaging variants, none of these had a PhoRank score of over 0.75, indicating that genes currently not connected to DDH may be involved in the pathology in these families.

Mutational burden analysis

Since the number of identified DDH candidate genes was high in our analysis, we decided to undertake an alternative approach by investigating which genes were most frequently affected across the families. We filtered as previously for shared variants between the affected members within the families and low variant frequency in the general population, and then we looked for genes affected by these variants in more than one family. We found 455 genes to be affected in at least two families, carrying a total of 958 variants (Table 2 and Additional file 8 showing the whole list of genes).

Table 2 Number of genes bearing rare variants co-segregating with disease in more than one family.

Genes common to	Number of genes	Number of variants
9 families	1	11
6 families	1	6
5 families	7	31
4 families	16	52
3 families	62	163
2 families	368	673
Total	455	958

Titin (*TTN*) was the most affected gene in our study group, as we found nine families and eleven missense variants in the *TTN* gene, with two of the families carrying two variants. Next in line was cadherin 23 (*CDH23*), with six families affected and carrying six different variants, among which a LoF initiator codon variant. Seven genes were affected in five families, 16 genes among four families, and the remaining 62 and 368 genes were found carrying variants among three and two families, respectively. Figure 4 shows a list of the 87 most commonly hit genes as they were found carrying variants in three or more families.

Since our most affected gene (*TTN*) is also the longest in the genome, we ran a correlation test between gene length and the number of variants, and gene length and the number of families with variants for each gene. A correlation coefficient of $r = 0.36$ and $r = 0.33$ respectively, suggested that gene length by itself was not a relevant factor in determining how many variants we would find, or how many families would carry the variants.

Next, we looked in the literature to see if the list of our most affected genes contains known DDH candidate genes. We found *NOTCH2*, *TBX4*, and *TENM3* known candidate genes each to have variants in two families; while *DACH1*, *CHD7*, *CX3CR1*, *ESR1*, *MYH10*, *PCNT*, and *POLE* candidate genes had each variants only in a single family (Additional file 4).

Although our top hit genes have not been previously reported in relation to DDH, many of them belonged to the same protein family or had a common function with many known candidate genes. For example, *CDH23*, an essential protein for inner-ear mechanotransduction and hearing [19], is the second most affected gene in our study and it belongs to the cadherin superfamily of calcium-dependent cell-cell adhesion proteins, as well as the DDH candidate gene protocadherin9 (*PCDH9*) [11]. Similarly, we found two families with missense variants in the plasma membrane Ca²⁺ transporting 2 (*ATP2B2*) gene, which is closely related to the *ATP2B4* gene, whose variants were previously described to co-segregate with heparan sulfate proteoglycan 2 (*HSPG2*) variants in four related DDH individuals [8]. The semaphorin gene *SEMA4D* is a candidate gene in dogs and in humans, and *SEMA5D* was previously reported in dogs [20, 21]; although we did not find variants in *SEMA4D* or *SEMA5D*, we did find a *SEMA6C* variant in three families.

Biological function and candidate pathways

At this point, we reasoned that the DDH phenotype might be a consequence of a dysfunctional pathway, rather than of a dysfunctional single gene. To investigate more systematically which groups of genes and pathways were more frequently affected by variants in our cohort, we performed GO term and pathway analysis taking into consideration only genes affected in two or more families.

The number one hit in the analysis of the enrichment of GO biological process terms was detection of mechanical stimulus, followed by limb morphogenesis (Fig. 5 and Additional file 9). Among other significantly enriched GO terms there were many belonging to the extracellular matrix, microtubules, or the cytoskeleton topology and function. GO cellular component terms were: dynein and collagen complexes, the costamere, the basement membrane, collagen-containing extracellular matrix, cell-substrate junctions, myofibrils, and the cilium. While enriched molecular function terms were ATP-dependent microtubule motor activity, dynein binding, microtubule and motor activity, calcium ion transporting and binding activity, and actin binding.

Pathway analysis revealed among significantly enriched KEGG pathways: ABC transporters and the Extracellular matrix (ECM)-receptor interactions pathway (Fig. 6 and Additional file 9). When we used the Reactome database, significant enrichment of the Notch pathway was highlighted as we found variants in the following five genes: *NOTCH2*, *CREBBP*, *EP300*, *MAML1*, and *NOTCH4*. Other enriched pathways were the Laminin interactions (FDR=0.0012), and the Non-integrin membrane-ECM interactions pathway (FDR=0.0260) due to many variants found in the laminin (*LAMA1*, *LAMA2*, *LAMA5*, *LAMC3*, *AGRN*) and collagen (*COL4A5*, *COL7A1*, *COL4A1*, *COL4A3*) genes.

Significantly enriched keywords associated with our group of genes were Usher syndrome as we curiously found variants in four genes known to be causative of the disease including *USH2A*, *CDH23*, *GPR98*, and *HARS*. Other enriched keywords included: laminin EGF-like domain, basement membrane, dynein, motor protein, actin-binding, cilium, and extracellular matrix in confirm to what already found with GO term and pathway analysis.

Protein interaction analysis

To investigate which protein networks the identified genes belonged to, we used STRING (<https://string-db.org/>) and performed protein interaction network analysis using as input the list of 455 genes affected in at least two families. The resulting protein network was rather large, shown in Fig. 7 (Additional file 10 for high resolution), with several clearly observable smaller clusters, most of which have already been identified by GO term and pathway analysis such as laminins and collagens, which together form a bigger network of extracellular matrix proteins, adhesion and cytoskeletal proteins, and the Notch signalling pathway. On the figure, there were also visible two clusters with members of the Wnt signalling pathway and the E3-ubiquitin protein ligase complex.

In the overall, the resulting protein interaction network had significantly more interconnections than expected, as the protein interaction network enrichment p value was $4.6e-14$ (string-db.org). The value was calculated for 453 number of nodes (proteins) in the network, and reaching to 613 edges, as compared to 446 expected edges for a network of a random set of proteins of similar size (the average node degree was 2.71 with an average node clustering coefficient of 0.34). This means that the genes from our list form a larger group of biologically connected proteins than what a random set of genes of the same dimensions would make.

Discussion

DDH is a common musculoskeletal condition influenced by genetic, mechanical, and environmental factors, resulting in a wide range of phenotypic manifestations. The earlier that DDH is diagnosed, the simpler and more effective are the treatments. However, late detection of instable hips is a persistent problem, and the phenotypic heterogeneity of DDH poses a serious challenge for diagnostics, especially in adults [3, 22]. Predictive genetic testing may improve significantly diagnostic methods and patient management in families with a positive family history of DDH.

Until now, 34 genes have been related to DDH in humans [8, 10-13, 20, 21, 23-44], sometimes with contradictory results [45]. However, no tight phenotype-genotype correlations have been established and the molecular mechanisms of the disease have just started to be elucidated. To this end, canine hip dysplasia is a relevant natural animal model for DDH [46]. Studies focusing on genetic markers for hip dysplasia in dogs have unraveled over 50 candidate genes, with little overlap with human candidate genes [14, 15, 47-51]. It is generally thought that sequence variants in individual genes are the cause of DDH, although a simple single-gene inheritance model is very unlikely to be able to explain the whole complexity of the phenotypic and genetic heterogeneity in DDH [3]. A multiple gene model of inheritance was already proposed in an initial study [7]. Modern sequencing techniques allow now for genome-wide investigations, and several studies using NGS reported many loci/genes in relation to DDH. A two-gene model identified by exome sequencing has been proposed for a family carrying *HSPG2* and *ATP2B4* variants [13]; while Zhu et al. used genome sequencing and found 1344 genes in 10 patients possibly carrying DDH-causative mutations, many of which are involved in cytoskeleton structure and function [30].

In the present study, we investigated exonic and splicing variants in family relatives affected by DDH from 29 families. We identified a novel gene, *PPP6R2*, involved in cell cycle entry and progression, to carry two missense variants co-segregating with the disease in three families. Since the two mutations are inherited together, they are most probably situated *in cis*, suggesting a founder allele model.

Besides two families carrying only *PPP6R2* variants, in most of the other families we identified several genes that may be imputable for the disease, suggesting a polygenic mode of inheritance.

In an attempt to select the genes that are more relevant to DDH, we looked for genes carrying variants in multiple families. This approach identified two novel genes, *TTN* and *CDH23*, as the two most affected genes in our study, presenting with variants in more families than any other gene. Although neither of these genes have been previously related to DDH, Zhu et al. [30] lists *TTN* among the DDH-specific genes in their study group, while *CDH23* was associated to DDH in a GWAS study from Yan et al. [52]. Among the DDH-closest genes (according to PhoRank) with the most damaging variants in our study, we found that *EVC2*, *FRAS1*, *CDC45*, and *NOTCH2* have variants co-segregating with the disease in two families each, while *ABCA3*, *EP300*, *TBX4*, and *IDUA* had variants in three families, further strengthening the notion that these genes may be involved in the etiopathogenesis of DDH.

If we considered group of similar genes, the number of families with variants would increase considerably. For example, *NOTCH2*, which encodes a transmembrane surface receptor, together with other three Notch receptors plays a fundamental role in bone morphogenesis [53]. Besides the two families carrying variants in known DDH candidate gene *NOTCH2*, we also found three families carrying variants in *NOTCH4*,

and one family with a variant in *NOTCH3*. Similarly, we found two missense variants in the known DDH gene *TBX4*, but we also found two families with *TBX5* variants. The myosin heavy chain 10 (*MYH10*) gene is another previously reported DDH gene [10]; we found one family with a very damaging variant in *MYH10*, but also three families in *MYH7B*, two families in *MYH14*, and one family in *MYH3*. Laminin 2 (*LAMA2*) is a candidate extracellular matrix gene described in hip dysplasia in dogs [48]. Besides finding two families with *LAMA2* variants, we also found four families with *LAMA1* variants, two families with *LAMA5* variants, and one family with a *LAMA3* variant.

We applied pathway analysis to provide insights into which groups of genes are the most commonly affected. As expected, we found the Notch transcription pathway, extracellular matrix-receptor interactions, and laminins to be among the significantly enriched pathways. Notch is an ancient pathway that regulates critical processes during embryological development and it is involved in skeletal development and homeostasis [53]. It is the third most important pathway for osteoblastogenesis, next to the Wnt signalling pathway and the bone morphogenetic protein (BMP) pathway, and these pathways interact between each other in osteoblast differentiation [54]. The Notch pathway is also involved in mechanotransduction [55, 56], as mechanical force is required for ligand-induced Notch activation [57, 58].

Many enriched GO-terms we found were related to the cytoskeleton and microtubules, similarly to what previously described by Zhu et al. [30]. More specifically, we identified the following GO molecular function terms as dominant: ATP-dependent microtubule motor activity, dyneins, actin-binding, and calcium binding. Collagens, the extracellular matrix, the basement membrane, the costamere, the axoneme, adherens junctions, the cilium, myofibrils, and cytoskeleton were most frequent topological GO term associations. Interestingly, the most enriched biological process GO term was detection of mechanical stimulus.

Mechanical forces influence the growth and shape of virtually every tissue and organ in our bodies, and through mechanotransduction these forces are translated into molecular processes and gene expression. Integrins and cadherins mediate the adhesion of cells to the extracellular matrix and their distortion activates key molecules on the cell membrane, such as ion channels and other signal-transduction molecules, which trigger signalling cascades leading eventually to changes in gene expression. An alternative model suggests also that mechanical stress can be conveyed directly through the cytoskeleton to a nucleic scaffold of actin, myosin, and nuclear titin, which detects the nuclear envelope stress and activates genes within seconds [59, 60].

Mechanotransduction occurs through an ensemble of membrane proteins, channels, structural cytoskeletal proteins, and mechanosensory organelles, such as the cilium and the cell nucleus. Cilia are specialized structures of the membrane localized on the surface of the majority of cell types including bone cells [61]. They have a microtubule-based core called the axoneme, with a centriolar anchor. Dysfunction of ciliary proteins generate a wide range of disorders named ciliopathies [62]. Many ciliopathies involve diverse skeletal malformation, suggesting that cilia are indispensable for regulating bone development and homeostasis. The ciliary membrane hosts many receptors for detecting various stimuli including Notch receptors. Cilia are an enriched GO term in our study as we find 26 genes in this category to have variants co-segregating with disease in more than two families including the previously reported candidate gene *EVC2*. Furthermore, variants in ciliary protein genes *RP1L1* and *PKD1* were found in five different families co-segregating with disease, as well as variants in four genes causing Usher syndrome involved in proper organization of stereocilia (*CDH23*, *GPR98*, *HARS*, *USH2A*), among which is our second most frequently affected gene *CDH23*.

Curiously, PPP6, whose regulatory subunit *PPP6R2* gene we identify as the novel DDH candidate gene, is a specific regulator of the G0 mitotic kinase Aurora A (*AURKA*) [63]. PPP6 inactivates *AURKA* through de-phosphorylation during spindle formation. However, an emerging role of *AURKA* is the regulation of ciliary assembly [64]. *AURKA* is abundant in G0 where it triggers ciliary resorption and *AURKA* removal is a likely prerequisite of effective ciliogenesis. An inappropriately active *AURKA* could impair the proper resorption of the cilia prior to mitosis and possibly disrupt the cell cycle.

Conclusion

We identify *PPP6R2* as novel DDH candidate gene, and we find very damaging variants in the following known candidate genes from human and canine hip dysplasia: *DACH1*, *MYH10*, *NOTCH2*, *TBX4*, *EVC2*, *OTOG*, and *SHC3*, which has important implications for the development of diagnostic methods for DDH. Our study also supports the hypothesis that multiple genes contribute to the development of DDH, as we found many families with several high damaging variants co-segregating with the disease. Pathway analysis of more frequently affected genes in our study group identified transduction of mechanical stimulus as the primary enriched biological process, and we found a remarkable number of variants in genes involved in cilia. We also found variants in a significant number of genes belonging to the Notch pathway, as well as the extracellular matrix proteins, microtubules, cytoskeleton, collagens, and laminins. Most of these proteins participate to the transduction of the mechanical stimulus at some level. This provides a link that could explain in molecular terms how environmental and mechanical stimuli, e.g. tight swaddling or a harness, determine the insurgence and/or the treatment of DDH. We conclude that DDH is a polygenic disorder, and we propose that impaired mechanotransduction is involved in the pathophysiological mechanism of the disease. Further studies are needed to confirm and define more in detail the biological significance of these variants respect to disease.

Methods

General overview of the project

We carried out genetic analysis of 66 affected individuals of Danish origin, from a total of 29 families, with confirmed DDH diagnosis through radiological and clinical examination, ascertained through the Danish periacetabular osteotomy (PAO) Database at Aarhus University Hospital [65]. We focused only on affected individuals within the families, as DDH is known for incomplete penetrance and including healthy relatives as controls may have confounding effects. As control we used MAFs of each variant in the total population according to the gnomAD database v2.1.1 (GRCh37/hg19), previously known as the Exome Aggregation Consortium (ExAC).

Whole exome sequencing

After obtaining written informed consent, we collected blood EDTA samples and performed whole exome sequencing (WES). Genomic DNA was isolated using standard methods. The libraries were prepared on a Beckman Coulter Biomek 4000 Laboratory Automation Workstation using Roche NimbleGen SeqCap EZ MedExome Enrichment Kits. Targeted 2x75bp pair end sequencing was performed on an Illumina NextSeq 550 platform.

Data analysis

Raw reads were processed using the Burrows-Wheeler Alignment tool (BWA-MEM) v. 0.7.12, and aligned to reference sequence (human NCBI build 37). GATK Best Practice pipeline v.4.1.4.1 was used for variant calling [66], while VarSeq software (Golden Helix) was used for variant annotation. Variant selection was done through a cascade of filtering steps in VarSeq. The following quality cut-offs were used: (1) accuracy base call of at least 99% (Q20), and (2) variant allele frequency (VAF) in reads \geq 25%. The variants were also checked manually at a later stage and any of them located in multimapped regions or regions with high background noise and low mean quality (<20) according to VarSeq were excluded. We excluded as well variants with less than 20x coverage, or originating from alignment or PCR mistakes. To select for DDH-specific, rare, and functional variants in single families we filtered coding and splicing variants through the following pipeline: (1) presence in all affected individuals in the family, (2) $MAF \leq 0.01$ (gnomAD database), (3) and less than 5 homozygotes. Splicing region variants, residing outside of the GT-AG dinucleotide, were assessed by Alamut Visual v.2.8 software and included only when significantly affecting splicing. We used VarSeq for CNVs calling and analysis. We selected CNVs that cover at least four exons, occurred in all affected individuals of a family, and presented with low frequency rates in the gnomAD database and the database of the genomic variants (DGV) (≤ 0.02).

Functional analysis of variants

To select for the most damaging and DDH-related variants we further filtered the variants keeping only those with less than 590 alternate allele counts (the expected allele count assuming an autosomal dominant mode of inheritance and given the average DDH incidence of 1:3400 [3]) in gnomAD. We used the Combined Annotation Dependent Depletion (CADD Scores 1.4) method to score the variants for deleteriousness, PHRED scaled scores were used [67]. The CADD algorithm includes conservation metrics, functional genomic data, exon-intron boundaries, and protein functionality scores. A score equal or greater of 20 indicates that a variant is predicted to be among the 1% of the most deleterious substitutions. We used a CADD score of over 25 to select for the most damaging variants in our study group. We further used a dbNSFP tool [68] that incorporates six different algorithms to evaluate the functional impact of a variant, including: SIFT, Polyphen, MutTaster, MutAssessor, FATHMM Pred, and FATHMM MKL. The most damaging variants were selected based on a voting system where five or all six of these algorithms predicted a negative functional impact. To estimate which variants were more likely to be involved in DDH we used the PhoRank algorithm and hip dysplasia as user-specified phenotype. PhoRank assigns a score to each gene based on their proximity to a user-specified phenotype, using Gene Ontology (GO) and Human Phenotype Ontology (HPO)[69]. The genes are ranked from 0 to 1 where 1 indicates the closest direct, or through a shared gene, relationship.

Genetic overlap, GO term, pathway, and STRING analysis

We created an R-script to make a list of all the genes and variants we found in the families and to filter for those that occur in multiple families. For term and pathway enrichment analysis, we used the list of 455 genes with variants in multiple families and the following databases (STRING v.11.0b, <https://string-db.org/>). Gene Ontology (GO) (<http://geneontology.org/>) is a comprehensive database of current knowledge of biological systems, where molecular function, biological process, and cellular component are the three aspects used for classification of the genes. Kyoto Encyclopedia of Genes and Genomes (KEGG) (<https://www.genome.jp/kegg/pathway.html>) is a collection of maps representing known molecular interactions and networks, while Reactome (<https://reactome.org/>) is a pathway database. The enrichments are calculated as strength measured with the formula: $\text{Log}_{10}(\text{observed} / \text{expected})$. Only terms with a false discovery rate (FDR) of under 0.05 are reported. They were calculated using the Benjamini–Hochberg procedure for multiple testing correction within each category. The protein interaction network was analyzed and visualized using STRING. STRING is a database of physical and functional, known and predicted, protein-protein

interactions [70]. The STRING online resource performs network and enrichment analysis, visualizes protein networks, and reports a statistical analysis of enrichments of connections between proteins, and functional subsystems.

Declarations

Ethics approval and consent to participate

The study was performed in line with the principles of the Declaration of Helsinki, and it has been approved by The Regional Committee on Health Research Ethics for Southern Denmark (S-20170055). Written informed consent was obtained from each patient included in the study prior to sample collection.

Consent for publication

Not applicable.

Data availability statement

In order to comply with the ethical approval, anonymized data supporting the findings of this study is available within the main article and its supplementary data, containing all the rare and shared variants within the families. Raw sequencing data are not available for broad data sharing because of the ethical and legal restrictions on data usage, since the patients did not provide informed consent that covers deposition to public repositories, sharing, and additional use of their genetic data. Please contact the corresponding author (Maja Dembic, Maja.Dembic@rsyd.dk), if you would like to request the raw data; it will however require an additional application to the ethical committee. All software used is described and cited throughout the manuscript. Custom scripts are available at <https://github.com/MaDemb/DDH>.

Competing interests

The authors declare no conflict of interest.

Funding

Funding support for this work has been received from 'The Danish Rheumatism Association' (grant number R101-A1997) and the "Bevica Fonden".

Authors' contributions

MD performed data analysis, genetic data interpretation, and wrote custom scripts and the manuscript. LvBA performed data analysis, and wrote custom scripts. MJL provided data analysis support, and participated in the discussions on the overall and detailed study plan. IM and KS performed clinical examinations and did the selection of patients; JH participated in the patient collection, designed the research project and supervised it. All authors discussed the results and commented on the manuscript.

Acknowledgements

Not applicable.

References

1. Yang S, Zusman N, Lieberman E, Goldstein RY. Developmental Dysplasia of the Hip. *Pediatrics*. 2019;143. doi: 10.1542/peds.2018-1147.
2. Dezateux C, Rosendahl K. Developmental dysplasia of the hip. *Lancet*. 2007;369:1541-52. doi: 10.1016/S0140-6736(07)60710-7.
3. Harsanyi S, Zamborsky R, Krajciová L, Kokavec M, Danisovic L. Developmental Dysplasia of the Hip: A Review of Etiopathogenesis, Risk Factors, and Genetic Aspects. *Medicina (Kaunas)*. 2020;56. doi: 10.3390/medicina56040153.
4. Basit S, Hannan MA, Khoshhal KI. Developmental dysplasia of the hip: usefulness of next generation genomic tools for characterizing the underlying genes - a mini review. *Clin Genet*. 2016;90:16-20. doi: 10.1111/cge.12755.
5. Gkiatas I, Boptsi A, Tserga D, Gelalis I, Kosmas D, Pakos E. Developmental dysplasia of the hip: a systematic literature review of the genes related with its occurrence. *EFORT Open Rev*. 2019;4:595-601. doi: 10.1302/2058-5241.4.190006.
6. Geiser M, Buri B, Buri P. Congenital dislocation of the hip in identical twins. *J Bone Joint Surg Br*. 1959;41-B:314-8. doi: 10.1302/0301-620X.41B2.314.
7. Wynne-Davies R. Acetabular dysplasia and familial joint laxity: two etiological factors in congenital dislocation of the hip. A review of 589 patients and their families. *J Bone Joint Surg Br*. 1970;52:704-16. <https://www.ncbi.nlm.nih.gov/pubmed/5487570>.

8. Basit S, Albalawi AM, Alharby E, Khoshhal KI. Exome sequencing identified rare variants in genes HSPG2 and ATP2B4 in a family segregating developmental dysplasia of the hip. *BMC Med Genet.* 2017;18:34. doi: 10.1186/s12881-017-0393-8.
9. Nasa I, Kettenbach AN. Coordination of Protein Kinase and Phosphoprotein Phosphatase Activities in Mitosis. *Front Cell Dev Biol.* 2018;6:30. doi: 10.3389/fcell.2018.00030.
10. Tuzovic L, Yu L, Zeng W, Li X, Lu H, Lu HM, et al. A human de novo mutation in MYH10 phenocopies the loss of function mutation in mice. *Rare Dis.* 2013;1:e26144. doi: 10.4161/rdis.26144.
11. Mabuchi A, Nakamura S, Takatori Y, Ikegawa S. Familial osteoarthritis of the hip joint associated with acetabular dysplasia maps to chromosome 13q. *Am J Hum Genet.* 2006;79:163-8. doi: 10.1086/505088.
12. Wang K, Shi D, Zhu P, Dai J, Zhu L, Zhu H, et al. Association of a single nucleotide polymorphism in Tbx4 with developmental dysplasia of the hip: a case-control study. *Osteoarthritis Cartilage.* 2010;18:1592-5. doi: 10.1016/j.joca.2010.09.008.
13. Basit S, Alharby E, Albalawi AM, Khoshhal KI. Whole genome SNP genotyping in a family segregating developmental dysplasia of the hip detected runs of homozygosity on chromosomes 15q13.3 and 19p13.2. *Congenit Anom (Kyoto).* 2018;58:56-61. doi: 10.1111/cga.12235.
14. Zhou Z, Sheng X, Zhang Z, Zhao K, Zhu L, Guo G, et al. Differential genetic regulation of canine hip dysplasia and osteoarthritis. *PLoS One.* 2010;5:e13219. doi: 10.1371/journal.pone.0013219.
15. Sanchez-Molano E, Woolliams JA, Pong-Wong R, Clements DN, Blott SC, Wiener P. Quantitative trait loci mapping for canine hip dysplasia and its related traits in UK Labrador Retrievers. *BMC Genomics.* 2014;15:833. doi: 10.1186/1471-2164-15-833.
16. Granchi D, Stea S, Sudanese A, Toni A, Baldini N, Giunti A. Association of two gene polymorphisms with osteoarthritis secondary to hip dysplasia. *Clin Orthop Relat Res.* 2002;108-17. doi: 10.1097/00003086-200210000-00018.
17. Cobos PN, Steglich C, Santer R, Lukacs Z, Gal A. Dried blood spots allow targeted screening to diagnose mucopolysaccharidosis and mucopolipidosis. *JIMD Rep.* 2015;15:123-32. doi: 10.1007/8904_2014_308.
18. Oussoren E, Wagenmakers M, Link B, van der Meijden JC, Pijnappel W, Ruijter GJG, et al. Hip disease in Mucopolysaccharidoses and Mucopolipidoses: A review of mechanisms, interventions and future perspectives. *Bone.* 2021;143:115729. doi: 10.1016/j.bone.2020.115729.
19. Jaiganesh A, Narui Y, Araya-Secchi R, Sotomayor M. Beyond Cell-Cell Adhesion: Sensational Cadherins for Hearing and Balance. *Cold Spring Harb Perspect Biol.* 2018;10. doi: 10.1101/cshperspect.a029280.
20. Shi D, Dai J, Zhu P, Qin J, Zhu L, Zhu H, et al. Association of the D repeat polymorphism in the ASPN gene with developmental dysplasia of the hip: a case-control study in Han Chinese. *Arthritis Res Ther.* 2011;13:R27. doi: 10.1186/ar3252.
21. Sekimoto T, Ishii M, Emi M, Kurogi S, Funamoto T, Hamada H, et al. Segmental copy number loss in the region of Semaphorin 4D gene in patients with acetabular dysplasia. *J Orthop Res.* 2013;31:957-61. doi: 10.1002/jor.22310.
22. Clinical practice guideline: early detection of developmental dysplasia of the hip. Committee on Quality Improvement, Subcommittee on Developmental Dysplasia of the Hip. *American Academy of Pediatrics. Pediatrics.* 2000;105:896-905. doi: 10.1542/peds.105.4.896.
23. Kenanidis E, Gkekas NK, Karasmani A, Anagnostis P, Christofilopoulos P, Tsiridis E. Genetic Predisposition to Developmental Dysplasia of the Hip. *J Arthroplasty.* 2020;35:291-300 e1. doi: 10.1016/j.arth.2019.08.031.
24. Zhao L, Zhou Z, Wang S, Jiao Q, Wu J, Ma F, et al. A recurrent mutation in bone morphogenetic proteins-2-inducible kinase gene is associated with developmental dysplasia of the hip. *Exp Ther Med.* 2017;13:1773-8. doi: 10.3892/etm.2017.4191.
25. Zhao L, Tian W, Pan H, Zhu X, Wang J, Cheng Z, et al. Variations of the COL1A1 gene promoter and the relation to developmental dysplasia of the hip. *Genet Test Mol Biomarkers.* 2013;17:840-3. doi: 10.1089/gtmb.2013.0179.
26. Feldman GJ, Parvizi J, Levenstien M, Scott K, Erickson JA, Fortina P, et al. Developmental dysplasia of the hip: linkage mapping and whole exome sequencing identify a shared variant in CX3CR1 in all affected members of a large multigeneration family. *J Bone Miner Res.* 2013;28:2540-9. doi: 10.1002/jbmr.1999.
27. Li L, Wang X, Zhao Q, Wang E, Wang L, Cheng J, et al. CX3CR1 polymorphisms associated with an increased risk of developmental dysplasia of the hip in human. *J Orthop Res.* 2017;35:377-80. doi: 10.1002/jor.23294.
28. Kapoor B, Dunlop C, Wynn-Jones C, Fryer AA, Strange RC, Maffulli N. Vitamin D and oestrogen receptor polymorphisms in developmental dysplasia of the hip and primary protrusio acetabuli—a preliminary study. *J Negat Results Biomed.* 2007;6:7. doi: 10.1186/1477-5751-6-7.
29. Hatzikotoulas K, Roposch A, Consortium DDHCC, Shah KM, Clark MJ, Bratherton S, et al. Genome-wide association study of developmental dysplasia of the hip identifies an association with GDF5. *Commun Biol.* 2018;1:56. doi: 10.1038/s42003-018-0052-4.
30. Zhu LQ, Su GH, Dai J, Zhang WY, Yin CH, Zhang FY, et al. Whole genome sequencing of pairwise human subjects reveals DNA mutations specific to developmental dysplasia of the hip. *Genomics.* 2019;111:320-6. doi: 10.1016/j.ygeno.2018.02.006.
31. Feldman G, Kappes D, Mookerjee-Basu J, Freeman T, Fertala A, Parvizi J. Novel mutation in Teneurin 3 found to co-segregate in all affecteds in a multi-generation family with developmental dysplasia of the hip. *J Orthop Res.* 2019;37:171-80. doi: 10.1002/jor.24148.

32. Sun Y, Wang C, Hao Z, Dai J, Chen D, Xu Z, et al. A common variant of ubiquinol-cytochrome c reductase complex is associated with DDH. *PLoS One*. 2015;10:e0120212. doi: 10.1371/journal.pone.0120212.
33. Liu S, Tian W, Wang J, Cheng L, Jia J, Ma X. Two single-nucleotide polymorphisms in the DKK1 gene are associated with developmental dysplasia of the hip in the Chinese Han female population. *Genet Test Mol Biomarkers*. 2014;18:557-61. doi: 10.1089/gtmb.2014.0044.
34. Sadat-Ali M, Al-Habdan IM, Bubshait DA. Genetic Influence in Developmental Dysplasia of the Hip in Saudi Arabian Children Due to GDF5 Polymorphism. *Biochem Genet*. 2018;56:618-26. doi: 10.1007/s10528-018-9864-7.
35. Rouault K, Scotet V, Autret S, Gaucher F, Dubrana F, Tanguy D, et al. Evidence of association between GDF5 polymorphisms and congenital dislocation of the hip in a Caucasian population. *Osteoarthritis Cartilage*. 2010;18:1144-9. doi: 10.1016/j.joca.2010.05.018.
36. Dai J, Shi D, Zhu P, Qin J, Ni H, Xu Y, et al. Association of a single nucleotide polymorphism in growth differentiate factor 5 with congenital dysplasia of the hip: a case-control study. *Arthritis Res Ther*. 2008;10:R126. doi: 10.1186/ar2540.
37. Hao Z, Dai J, Shi D, Xu Z, Chen D, Zhao B, et al. Association of a single nucleotide polymorphism in HOXB9 with developmental dysplasia of the hip: a case-control study. *J Orthop Res*. 2014;32:179-82. doi: 10.1002/jor.22507.
38. Tian W, Zhao L, Wang J, Suo P, Wang J, Cheng L, et al. Association analysis between HOXD9 genes and the development of developmental dysplasia of the hip in Chinese female Han population. *BMC Musculoskelet Disord*. 2012;13:59. doi: 10.1186/1471-2474-13-59.
39. Ma W, Zha Z, Chen K, Chen H, Wu Y, Ma J, et al. Genetic association study of common variants in TGFB1 and IL-6 with developmental dysplasia of the hip in Han Chinese population. *Sci Rep*. 2017;7:10287. doi: 10.1038/s41598-017-11185-1.
40. Cengic T, Trkulja V, Pavelic SK, Ratkaj I, Markova-Car E, Mikolaucic M, et al. Association of TGFB1 29C/T and IL6 -572G/C polymorphisms with developmental hip dysplasia: a case-control study in adults with severe osteoarthritis. *Int Orthop*. 2015;39:793-8. doi: 10.1007/s00264-015-2675-0.
41. Jia J, Li L, Zhao Q, Zhang L, Ru J, Liu X, et al. Association of a single nucleotide polymorphism in pregnancy-associated plasma protein-A2 with developmental dysplasia of the hip: a case-control study. *Osteoarthritis Cartilage*. 2012;20:60-3. doi: 10.1016/j.joca.2011.10.004.
42. Qiao L, Yan W, Dai J, Yao Y, Chen D, Xu Z, et al. A novel missense variant in TXNDC3 is associated with developmental dysplasia of the hip in Han Chinese population. *Int J Clin Exp Pathol*. 2017;10:10483-8. <https://www.ncbi.nlm.nih.gov/pubmed/31966386>.
43. Watson CM, Crinnion LA, Gleghorn L, Newman WG, Ramesar R, Beighton P, et al. Identification of a mutation in the ubiquitin-fold modifier 1-specific peptidase 2 gene, UFPS2, in an extended South African family with Beukes hip dysplasia. *S Afr Med J*. 2015;105:558-63. doi: 10.7196/SAMJnew.7917.
44. Zhang J, Yan M, Zhang Y, Yang H, Sun Y. Association analysis on polymorphisms in WISP3 gene and developmental dysplasia of the hip in Han Chinese population: A case-control study. *Gene*. 2018;664:192-5. doi: 10.1016/j.gene.2018.04.020.
45. Rubini M, Cavallaro A, Calzolari E, Bighetti G, Sollazzo V. Exclusion of COL2A1 and VDR as developmental dysplasia of the hip genes. *Clin Orthop Relat Res*. 2008;466:878-83. doi: 10.1007/s11999-008-0120-z.
46. Pascual-Garrido C, Guilak F, Rai MF, Harris MD, Lopez MJ, Todhunter RJ, et al. Canine hip dysplasia: A natural animal model for human developmental dysplasia of the hip. *J Orthop Res*. 2018;36:1807-17. doi: 10.1002/jor.23828.
47. Ginja M, Gaspar AR, Ginja C. Emerging insights into the genetic basis of canine hip dysplasia. *Vet Med (Auckl)*. 2015;6:193-202. doi: 10.2147/VMRR.S63536.
48. Lavrijsen IC, Leegwater PA, Martin AJ, Harris SJ, Tryfonidou MA, Heuven HC, et al. Genome wide analysis indicates genes for basement membrane and cartilage matrix proteins as candidates for hip dysplasia in Labrador Retrievers. *PLoS One*. 2014;9:e87735. doi: 10.1371/journal.pone.0087735.
49. Fels L, Marschall Y, Philipp U, Distl O. Multiple loci associated with canine hip dysplasia (CHD) in German shepherd dogs. *Mamm Genome*. 2014;25:262-9. doi: 10.1007/s00335-014-9507-1.
50. Pfahler S, Distl O. Identification of quantitative trait loci (QTL) for canine hip dysplasia and canine elbow dysplasia in Bernese mountain dogs. *PLoS One*. 2012;7:e49782. doi: 10.1371/journal.pone.0049782.
51. Fels L, Distl O. Identification and validation of quantitative trait loci (QTL) for canine hip dysplasia (CHD) in German Shepherd Dogs. *PLoS One*. 2014;9:e96618. doi: 10.1371/journal.pone.0096618.
52. Yan W, Hao Z, Tang S, Dai J, Zheng L, Yu P, et al. A genome-wide association study identifies new genes associated with developmental dysplasia of the hip. *Clin Genet*. 2019;95:345-55. doi: 10.1111/cge.13483.
53. Zieba JT, Chen YT, Lee BH, Bae Y. Notch Signaling in Skeletal Development, Homeostasis and Pathogenesis. *Biomolecules*. 2020;10. doi: 10.3390/biom10020332.
54. Lin GL, Hankenson KD. Integration of BMP, Wnt, and notch signaling pathways in osteoblast differentiation. *J Cell Biochem*. 2011;112:3491-501. doi: 10.1002/jcb.23287.

55. LaFoya B, Munroe JA, Mia MM, Detweiler MA, Crow JJ, Wood T, et al. Notch: A multi-functional integrating system of microenvironmental signals. *Dev Biol.* 2016;418:227-41. doi: 10.1016/j.ydbio.2016.08.023.
56. Stassen O, Ristori T, Sahlgren CM. Notch in mechanotransduction - from molecular mechanosensitivity to tissue mechanostasis. *J Cell Sci.* 2020;133. doi: 10.1242/jcs.250738.
57. Gordon WR, Zimmerman B, He L, Miles LJ, Huang J, Tiyanont K, et al. Mechanical Allosteric: Evidence for a Force Requirement in the Proteolytic Activation of Notch. *Dev Cell.* 2015;33:729-36. doi: 10.1016/j.devcel.2015.05.004.
58. Chowdhury F, Li IT, Ngo TT, Leslie BJ, Kim BC, Sokoloski JE, et al. Defining Single Molecular Forces Required for Notch Activation Using Nano Yoyo. *Nano Lett.* 2016;16:3892-7. doi: 10.1021/acs.nanolett.6b01403.
59. Ingber DE. Tensegrity I. Cell structure and hierarchical systems biology. *J Cell Sci.* 2003;116:1157-73. doi: 10.1242/jcs.00359.
60. Wang N, Tytell JD, Ingber DE. Mechanotransduction at a distance: mechanically coupling the extracellular matrix with the nucleus. *Nat Rev Mol Cell Biol.* 2009;10:75-82. doi: 10.1038/nrm2594.
61. Kaku M, Komatsu Y. Functional Diversity of Ciliary Proteins in Bone Development and Disease. *Curr Osteoporos Rep.* 2017;15:96-102. doi: 10.1007/s11914-017-0351-6.
62. Reiter JF, Leroux MR. Genes and molecular pathways underpinning ciliopathies. *Nat Rev Mol Cell Biol.* 2017;18:533-47. doi: 10.1038/nrm.2017.60.
63. Zeng K, Bastos RN, Barr FA, Gruneberg U. Protein phosphatase 6 regulates mitotic spindle formation by controlling the T-loop phosphorylation state of Aurora A bound to its activator TPX2. *J Cell Biol.* 2010;191:1315-32. doi: 10.1083/jcb.201008106.
64. Pan J, Seeger-Nukpezah T, Golemis EA. The role of the cilium in normal and abnormal cell cycles: emphasis on renal cystic pathologies. *Cell Mol Life Sci.* 2013;70:1849-74. doi: 10.1007/s00018-012-1052-z.
65. Simonsen PO, Hertz JM, Soballe K, Mechlenburg I. Association between periacetabular osteotomy and hip dysplasia among relatives: a cross-sectional study. *Hip Int.* 2019;29:424-9. doi: 10.1177/1120700018785293.
66. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics.* 2013;43:11 0 1- 0 33. doi: 10.1002/0471250953.bi1110s43.
67. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* 2019;47:D886-D94. doi: 10.1093/nar/gky1016.
68. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet.* 2014;46:310-5. doi: 10.1038/ng.2892.
69. Singleton MV, Guthery SL, Voelkerding KV, Chen K, Kennedy B, Margraf RL, et al. Phevor combines multiple biomedical ontologies for accurate identification of disease-causing alleles in single individuals and small nuclear families. *Am J Hum Genet.* 2014;94:599-610. doi: 10.1016/j.ajhg.2014.03.010.
70. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;47:D607-D13. doi: 10.1093/nar/gky1131.

Figures

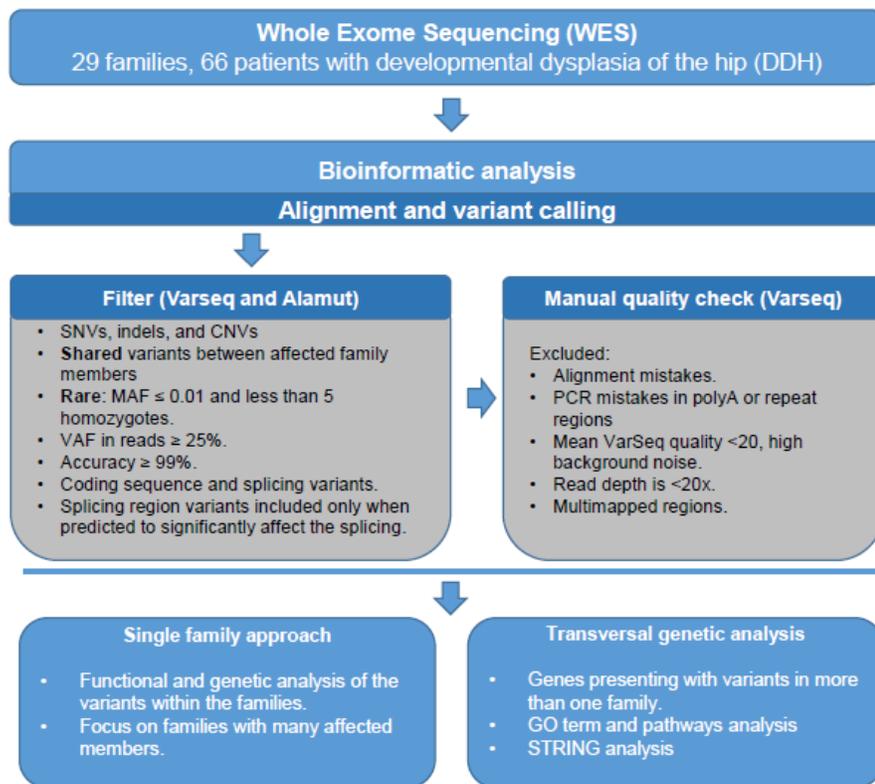


Figure 1
General overview of the project We performed whole exome sequencing of 66 patients from 29 families. After alignment and variant calling, the variants were filtered according to frequency and quality, and were manually checked for artefacts. We undertook a single family approach performing genetic analysis of the variants within each of the family, and a transversal approach across families identifying genes occurring with variants in more than one family. Abbreviations. SNVs: single nucleotide variants; InDels: inserts and deletions; CNVs: copy number variations; MAF: minor allele frequency; UTR: untranslated region; VAF: variant allele frequency (in reads). GO: Gene Ontology.

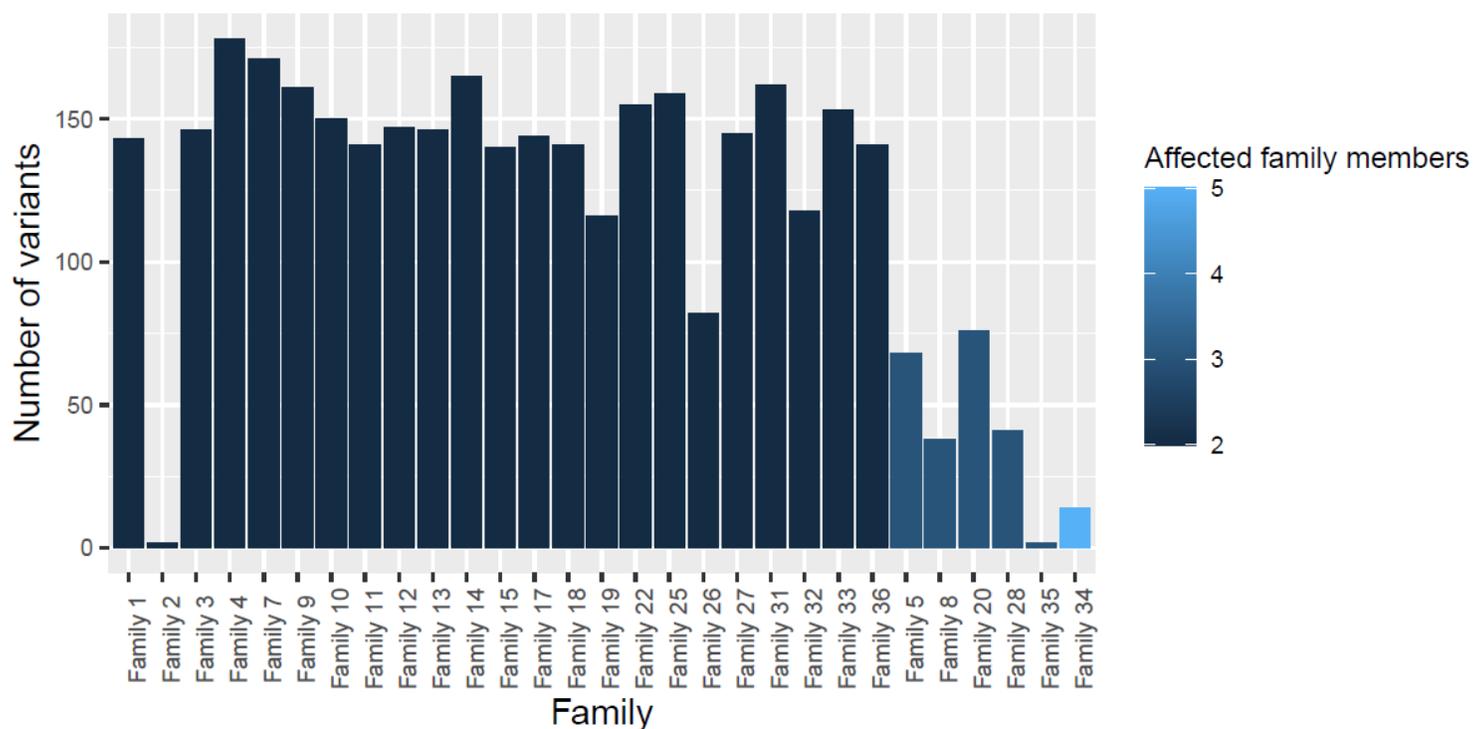


Figure 2
Page 14/19

Distribution of the variants across the families In the figure is shown a histogram representing the distribution of 3332 rare variants across the families. These included all splicing and coding sequence variants, with a MAF ≤ 0.01 , and shared between the affected individuals within a family. Most of the families (dark blue) had only two members, five families had three members (lighter blue), while family 34 had five members. Family 2 and 35 had only two variants, discovered to be the same ones.

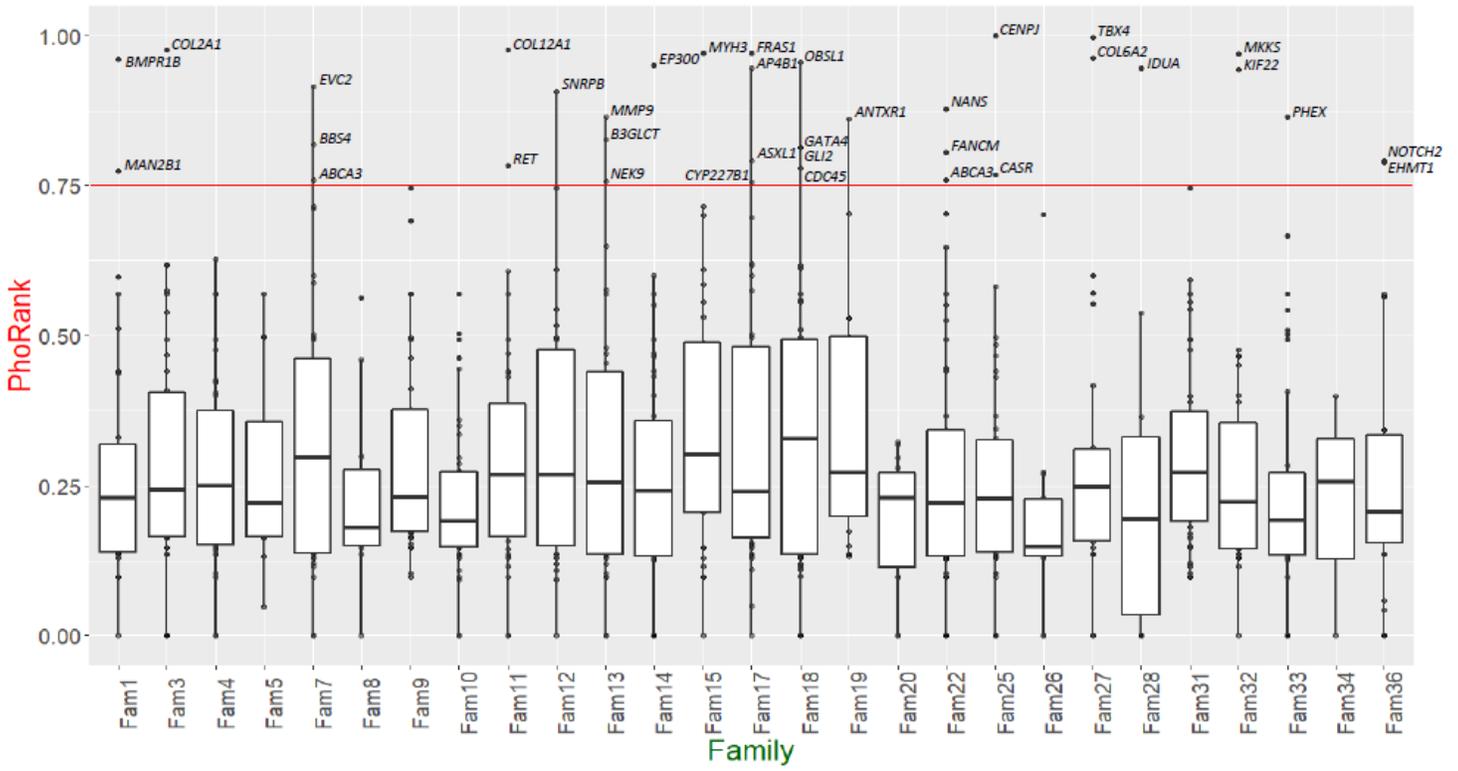


Figure 3

DDH relevant genes with the most damaging mutations In the figure is shown a box-plot of the 825 most damaging variants and their distribution across the families. The most damaging variants found in the affected individuals from 29 families were selected based on a high prediction of deleteriousness. The variants were either loss-of-function (LoF), had a CADD score of over 25, or were predicted to be damaging by 5 or more in silico tools. The box-plot shows the distribution of PhoRank values of the variants across the families. A PhoRank of over 0.75 predicts a related function of the gene respect to hip dysplasia phenotype using Gene Ontology (GO) and Human Phenotype Ontology (HPO). The median line shows the median, the hinges are the first and third quartiles, and the whiskers extend to the largest and smallest value, but not further than $1.5 * IQR$ (first and third interquartile range) from the hinge. The names of the genes carrying very damaging variants and a PhoRank score of over 0.75 are visible over the red line.

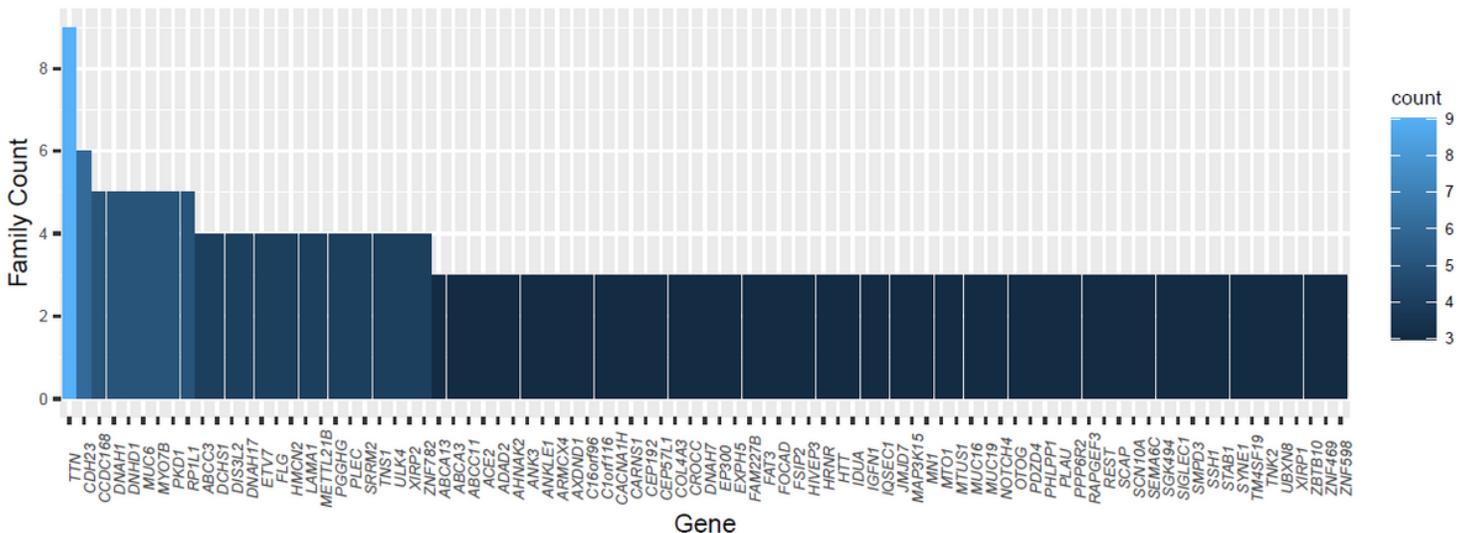


Figure 4

The most frequently affected genes in our study Histogram with the 87 genes bearing rare variants co-segregating with DDH in multiple families and the respective family count. TTN is the most affected gene occurring with variants in nine families. For simplicity, only the genes having variants in three or more families are shown.

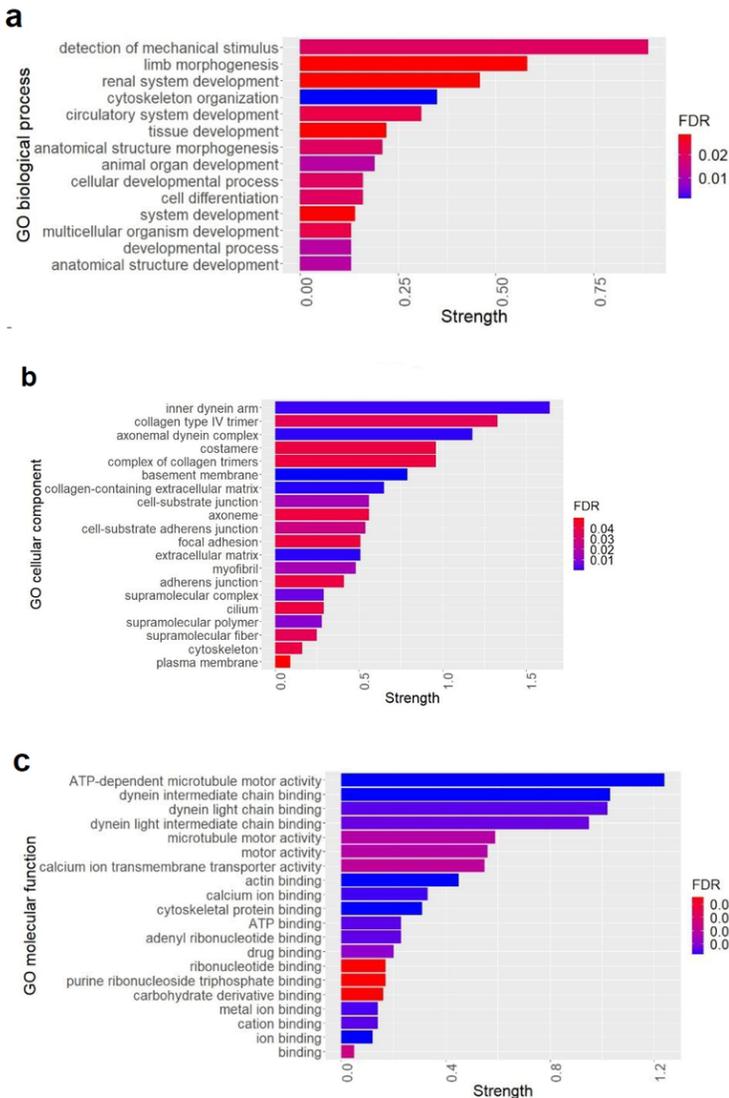


Figure 5

GO term analysis of the most affected genes in our study In the figure is shown an enrichment analysis made with STRING (string-db.org) of the GO terms within biological process (a), cellular component (b), and molecular function (c) of the 455 genes carrying variants in two or more families. The strength describes the enrichment effect and it is measured as $\text{Log}_{10}(\text{observed} / \text{expected})$. The ratio is between the number of observed genes from our network in a specific category, and the expected number of genes falling in the same category from a random network of the same size. False discovery rates (FDR) are p-values corrected for multiple testing using the Benjamini–Hochberg procedure. Enrichments with an FDR of under 0.05 are considered significant and displayed in the figure.

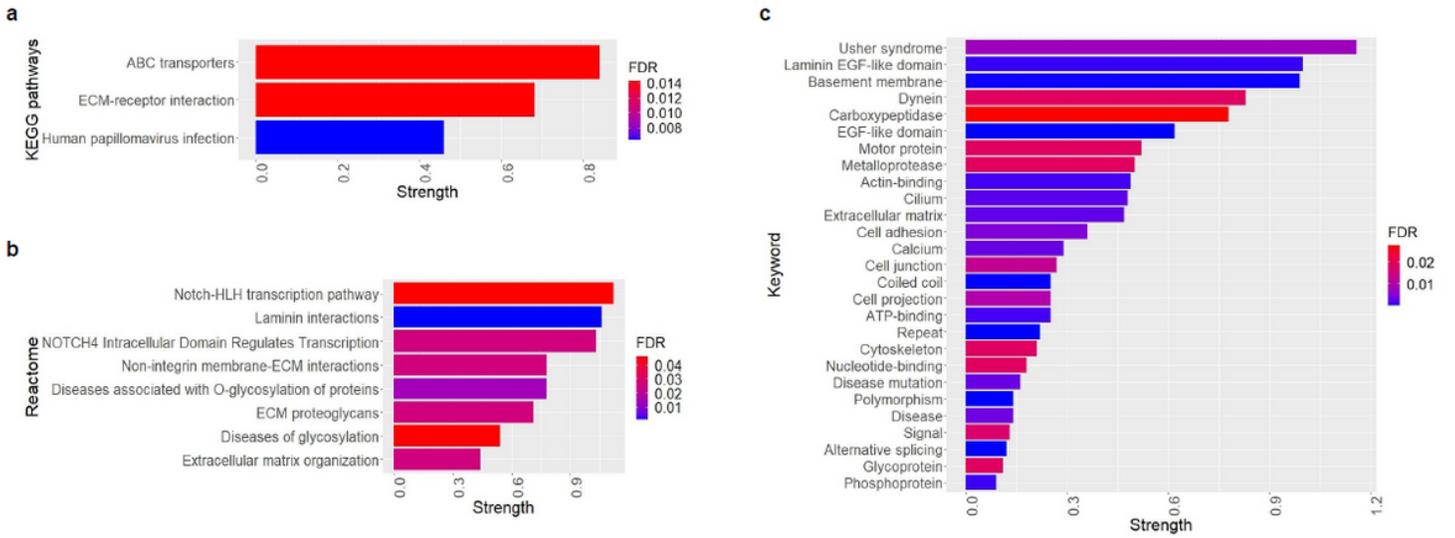


Figure 6

KEGG pathway, Reactome, and Keyword analysis with the most affected genes in our study. Pathway enrichment analysis performed with the 455 genes more commonly affected in our study group, according to KEGG (a), Reactome pathway databases (b), and a database of keywords (c) is shown in the figure. FDRs under 0.05 denote significantly enriched.

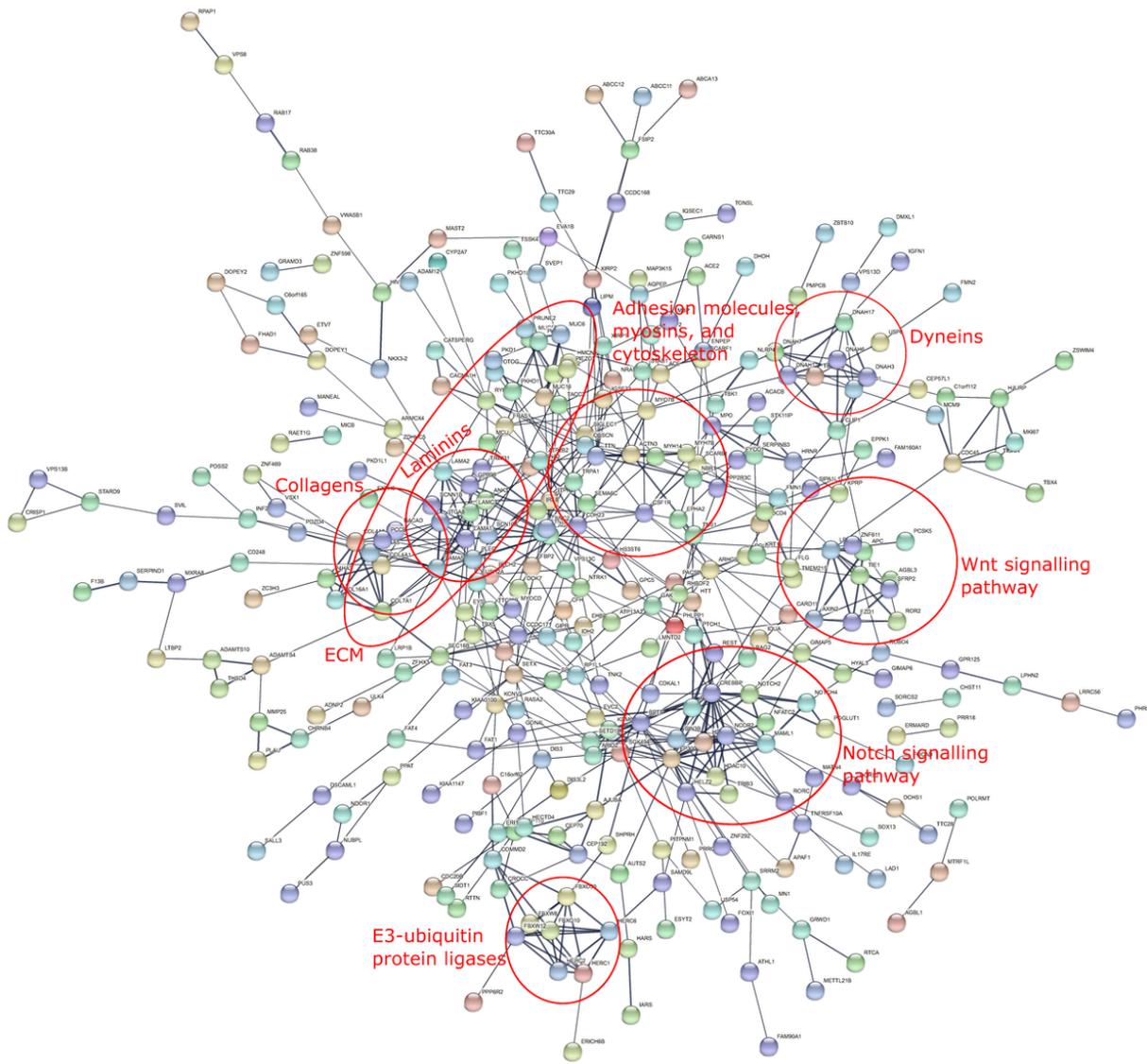


Figure 7

Protein interaction network of the most frequently affected genes In the figure is shown a protein interaction network calculated using the STRING software. The spherical nodes in the network are single proteins; loose nodes with no connection to other proteins are not visualized in this representation. The lines represent functional and physical protein associations, while the thickness of the lines represents the confidence level, with the thickest representing maximum confidence. Some of the clusters visible as an aggregation of high confidence connections between nodes are highlighted in red. ECM: extracellular matrix.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFig.1.pdf](#)
- [SupplementaryFig.2.pdf](#)
- [SupplementaryTable1.xlsx](#)
- [SupplementaryTable2.xlsx](#)
- [SupplementaryTable3.xlsx](#)
- [SupplementaryTable4.xlsx](#)
- [SupplementaryTable5.xlsx](#)
- [SupplementaryTable6.xlsx](#)

- [SupplementaryTable7.xlsx](#)
- [SupplementaryTable8.xlsx](#)