



UNIVERSITY OF LEEDS

**Investigating the mutation spectrum for Familial
Exudative Vitreoretinopathy (FEVR)**

DONG SUN

Submitted in accordance with the requirements for the degree of

Master by Research

The University of Leeds

School of Medicine

September 2020

Under the supervision of: Dr. Carmel Toomes, Prof. Chris Inglehearn, Dr. James Poulter

The candidate confirms that the work submitted is his/her own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

©2020 The University of Leeds and Dong Sun

Acknowledgments

First of all, I would like to express my gratitude to my supervisors Dr. Carmel Toomes, Professor Chris Inglehearn and Dr. James Poulter for their kind support throughout the whole year. I was given a lot of valuable advice, guidance and feedback, which inspired me a lot during the entire research. Without their kind help the result of the research project would not be the same. Furthermore, I would like to thank two medical undergraduate students, Jasmine Chen and Dimitra-Ilektra Lerou, who put a lot of time in uploading variants information of FEVR to Leiden Open Variation Database (LOVD) database. Finally, I want to thank Dr. Sandra Bell who gave me a great deal of kind help during the entire year, and Mohammed Derar who supported me as a friend and gave me a lot of encouragement.

I am really grateful for all their kind help and support.

Abstract

Familial exudative vitreoretinopathy (FEVR) is a rare inherited retinal eye disease, which can lead to vision loss or total blindness. It is characterized by abnormal development of blood vessels in the retina, which may trigger neovascularization leading to retinal folds and detachment. FEVR is genetically heterogeneous with eleven genes reported to date (*NDP*, *FZD4*, *LRP5*, *TSPAN12*, *CTNNB1*, *KIF11*, *ZNF408*, *ATOH7*, *RCBTB1*, *JAG1* and *ILK*) and an ever-increasing number of variants being reported in these. This project focused on generating a mutation database and spectrum for FEVR. Variant information was gathered either from unpublished data from the Toomes' lab or from peer-reviewed journal articles. All of the mutations were annotated so that they complied with the latest Human Genome Variation Society nomenclature and were uploaded onto the Leiden Open Variation Database. Pathogenicity classifications of variants were reassessed using the latest pathogenic prediction tools and frequency data and were classified according to the American College of Medical Genetics (ACMG) guidelines. In addition, missense mutant protein structures were predicted using Modeller9.24 software and compared with the correspondent wildtype protein model using UCSF chimera to attempt to highlight differences to aid in variant classification. The final dataset categorized 537 unique mutations identified in 902 FEVR cases. However, only 183 variants were confirmed by ACMG as pathogenic, and the majority of them (55.41%) were considered as variants of uncertain significance. This study demonstrates the difficulties in using the ACMG criteria for FEVR molecular diagnoses and highlights the urgent need for an accurate functional assay.

Contents

List of Abbreviations	iv
List of Tables	vi
List of Figures.....	vii
1. Introduction.....	1
1.1 Familial Exudative Vitreoretinopathy (FEVR).....	1
1.1.1 Clinical phenotype	1
1.1.2 Treatment methods	2
1.1.3 Genetic basis and molecular pathogenesis.....	3
1.2 Human Variation and Mutations.....	11
1.2.1 Mutation pathogenicity classification (ACMG)	13
1.3 Protein modelling.....	15
1.3.1 Protein 3D structure prediction and optimization.....	16
1.3.2 Protein structure quality evaluation	17
1.3.3 Structural comparison of protein variants.....	17
1.4 Aims.....	18
2. Methodology	18
2.1 Ethical approval	18
2.2 FEVR mutation database	19
2.2.1 Literature Search.....	19
2.2.2 Unpublished data	19
2.2.3 Variant annotation.....	19
2.2.4 <i>In silico</i> pathogenicity predictions	20
2.2.5 Pathogenicity classification	20
2.2.6 Mutation spectra	22
2.2.7 Statistical analysis.....	22
2.2.8 Phenotype-genotype correlation	22
2.2.9 LOVD submission	22
2.3 Three-dimensional protein modelling.....	22

2.3.1 Wildtype protein structure	22
2.3.2 Mutant protein structure.....	23
2.3.3 Comparison of wildtype and mutant protein structures.....	23
3 Results.....	24
3.1 FEVR Variant database	24
3.1.1 Variants annotation	24
3.1.2 Unpublished data overview	25
3.1.3 All variants overview.....	27
3.1.4 Pathogenicity classification	29
3.1.5 Mutation spectra	31
3.1.6 Frequently reported variants	39
3.1.7 Phenotype-Genotype correlation	40
3.1.8 LOVD	41
3.2 Protein modelling.....	41
3.2.1 Wildtype 3D protein structures analysis.....	41
3.2.2 Comparison of wildtype and mutant protein structures.....	44
4. Discussion.....	50
4.1 Pathogenicity classification	51
4.2 Protein modelling.....	55
4.3 Mutation spectra	56
5. Summary and future expectations.....	57
6. References.....	58
7. <i>Appendices</i>	69
<i>Appendix 1.</i>	69
<i>Appendix 2.</i>	79
<i>Appendix 3.</i>	80

List of Abbreviations

3D	Three-dimensional
ACMG	American College of Medical Genetics and Genomics
APC	Adenomatous polyposis coli
CADD	Combined Annotation Dependent Depletion
CK1	Casein kinase 1
CRD	Cysteine-rich domain
CTNNB1	Catenin- β -1
DOPE	Discrete Optimized Protein Energy
DSH	Dishevelled protein
Eg5	Kinesin-5
ExAC	Exome Aggregation Consortium
FEVR	Familial exudative vitreoretinopathy
FFA	Fundus Fluorescein angiography
GSK-3	Glycogen synthase kinase
HGMD	Human Gene Mutation Database
HGVS	Human Genome Variation Society
LDL	Low density lipoprotein
LDRP	low-density lipoprotein receptor
LEF	Lymphoid enhancing factor (LEF)
LOVD	Leiden Open Variation Database
LRP5	Low-density lipoprotein receptor-related protein 5
MCLMR	Microcephaly with or without chorioretinopathy, syndrome lymphedema, or mental retardation syndrome
NCRNA disease	Non-syndromic congenital retinal nonattachment
NGS	Next-generation sequencing
NMR	Nuclear magnetic resonance
OPPG	Osteoporosis-Pseudoglioma Syndrome
PolyPhen-2	Polymorphism Phenotyping v2

PROVEAN	Protein Variation Effect Analyzer
RMSD	Root-Mean-Square Deviation
ROP	Retinopathy of prematurity
RP	Retinis pigmentosa
PFV	Persistent Fetal Vasculature
SIFT	The sorting intolerant from tolerant
TCF	T-cell factor
VEGF	Vascular endothelial growth factor
XRC	X-ray crystallography

List of Tables

Table 1. Nomenclature errors identified for variants in previous publications	25
Table 2. Summary of unpublished variants in FEVR patients. ^{*(a)}	26
Table 3. Summary of the number of cases and unique variants in FEVR.....	28
Table 4. Distribution of variants types in FEVR.....	28
Table 5. Pathogenicity classification of variants in FEVR.....	30
Table 6. Percentage of VUS missense variants in FEVR.....	30
Table 7. Summary of reported variants of FEVR in potential genes and loci.....	31
Table 8. Wildtype Structure comparison between ITASSER and Phyre2	42
Table 9. Model comparison of missense mutant structure and wildtype FZD4 structure of FEVR patients	44
Table A 1. Model comparison of missense mutant structure and wildtype FZD4 structure of FEVR patients (Raw data)	69

List of Figures

<i>Figure 1.</i> Clinical feature of FEVR.	2
<i>Figure 2.</i> Overview of the Norrin- β -catenin pathway.	5
<i>Figure 3.</i> Distribution of unique variants types in FEVR patients.	29
<i>Figure 4.</i> Mutation spectrum of FZD4 in FEVR patients.	33
<i>Figure 5.</i> Distribution of variants of TSPAN12 in FEVR.	34
<i>Figure 6.</i> The distribution of FEVR variants in Norrin/NDP.	35
<i>Figure 7.</i> The distribution of variants in LRP5 in FEVR patients.	37
<i>Figure 8.</i> Mutation spectrum of the β -catenin in FEVR patients.	39
<i>Figure 9.</i> Predicted 3D wildtype protein structures of FZD4, NDP/Norrin, TSPAN12, LRP5 and β -catenin (CTNNB1) by I-TASSER.	43
<i>Figure 10.</i> Representative examples of structure deviations between wildtype and missense mutant FZD4 protein 3D structures.	48
<i>Figure 11.</i> Representative examples of structure deviation comparison between wildtype and missense mutant NDP/Norrin protein 3D structures.	49
<i>Figure 12.</i> Representative examples of structure deviations comparison between wildtype and missense mutant TSPAN12 protein 3D structures. ..	50
<i>Figure A 1.</i> All mutant structures of FZD4 superimposing with wildtype model in FEVR, and the Ramachandran plot of the mutant model of FZD4.	78
<i>Figure A 2.</i> Example of the screenshots of the scripts used in Modeller9.24 for modelling G22E by using wildtype FZD4 structure as the reference.	82

1. Introduction

1.1 Familial Exudative Vitreoretinopathy (FEVR)

1.1.1 Clinical phenotype

Familial exudative vitreoretinopathy (FEVR) is a rare hereditary eye disease, which was first reported by Criswick and Schepens in 1969 (Criswick and Schepens, 1969). The unifying feature of FEVR is incomplete retinal vasculature development. This abnormal blood vessel development can lead to ischemia in the retina which triggers secondary defects including neovascularization, exudation, retinal folds, tears and detachments. The vision loss (due to ischemia) can develop at any age and patients usually have long periods of stable disease, but it is a progressive disorder and patients need life-long monitoring. However, the signs and symptoms of FEVR patients can be widely variable, even within the same family or between two eyes of the same patient (Toomes and Downey, 2011). Some affected individuals are asymptomatic, whereas others could be totally blind (Kondo et al., 2003). The majority of FEVR cases only have an eye phenotype. However, there are some non-ocular manifestations associated with mutations in *LRP5*, *KIF11* and *CTNNA1* (discussed in section 1.1.3).

There are different FEVR staging systems that have been used in the classification of FEVR based on the disease severity. Gow and Oliver proposed the first FEVR staging system in 1971, and Laqua modified it to a three-tier staging system in 1980 (Gow and Oliver, 1971, Laqua, 1980). However, the most commonly used system is by Pendergast and Trese which presents a detailed five-tier staging method that incorporates fundus fluorescein angiography (FFA) findings (Pendergast and Trese, 1998). In this system, stage 1 FEVR is classified as peripheral avascularity in the retina, often only revealed by FFA. In stage 2 FEVR, neovascularization (new blood vessels) occurs in the retina with (2B) or without (2A) retinal exudation. Stage 3 FEVR is represented by partial retinal detachment in the extramacular area (3A: without exudate; 3B: with exudate). Stage 4 FEVR shows partial retinal detachment involving the macula area with (4B) or without (4A) exudate. Stage 5 associates with

a total retinal detachment with open funnel (5A) or with closed funnel (5B) (Gilmour, 2015). Kashani et al. updated this classification by adding more details (Kashani et al., 2014); Abnormal intraretinal neovascularization was added to stage 1 and the exudate phenotype in the B subcategories was expanded to encompass exudate or leakage of the blood vessels in the retina. Examples of the clinical features of different stages of FEVR are shown in the *Figure 1*.

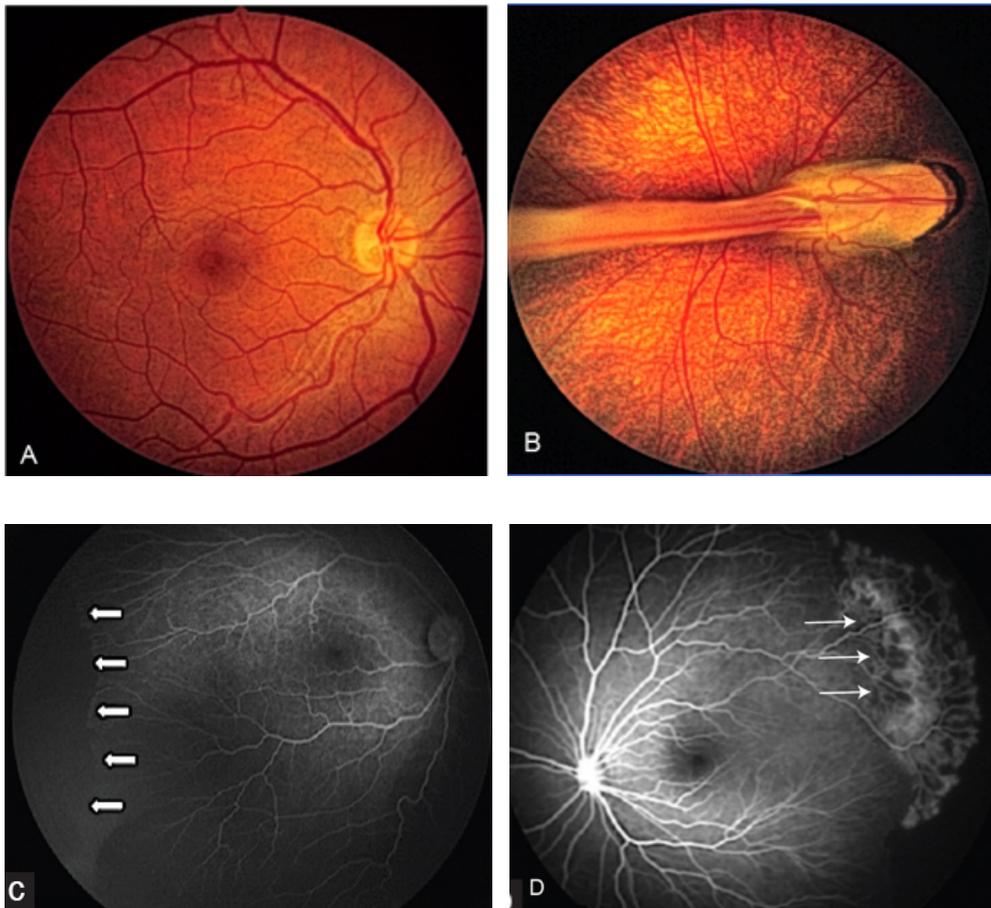


Figure 1. Clinical feature of FEVR. (A) Fundus image of a normal eye. **(B)** Fundus image of stage 4 FEVR showing marked vessel straightening and a retinal fold crossing the macula. **(C)** FFA image of stage 1 FEVR showing peripheral retinal avascularity (arrows). **(D)** FFA image of stage 2 FEVR showing neovascularization (arrows) around a region of peripheral retinal avascularity. Figures C and D reproduced with permission from Temkar et.al 2019 (Temkar et al., 2019).

1.1.2 Treatment methods

There is currently no cure for FEVR and treatments are focused on halting disease progression. The available treatment methods are based on the disease severity. In mild stages, for example, when the vision problems are not obvious, observation and follow-up are appropriate (Franco M. 2013). For stage 2 or higher FEVR, argon laser photocoagulation in the retina can effectively control the progression of FEVR into higher stages by triggering the regression of the new blood vessels (Kang et al., 2013). In the advanced stage cases (from stage3 to stage5), vitrectomy with or without scleral buckling surgery or prophylactic cryotherapy could be selected according to the severity of the retinal detachment (Katagiri et al., 2018, Shukla et al., 2003). These methods could contribute to release the traction and stabilize or improve the vision. Furthermore, FEVR is a lifelong disease and could lead to progressive vision loss, so follow-up is always recommended (Toomes and Downey, 2011).

There are currently new therapies being developed and trialled for FEVR. The most common involves injecting eyes with anti-VEGF therapeutics to halt neovascularisation and its tractional effects (Lu et al., 2018, Quiram et al., 2008, Henry et al., 2015). These therapies have been shown to be highly effective in treating other diseases with ocular neovascularisation, including age-related macular degeneration (Michalska-Małecka et al., 2016) and diabetic retinopathy (Zhao and Singh, 2018), but they are unproven in FEVR at the moment. A synthetic protein which mimics Norrin (Noregen™) has also been developed with the aim of developing treatments for FEVR and other disorders with abnormal retinal and choroidal blood vessel development (<https://www.retinalolutions.com/>).

1.1.3 Genetic basis and molecular pathogenesis

FEVR is genetically heterogeneous, the inheritance patterns include autosomal dominant, autosomal recessive and X-linked. To date, genetic mutations in five genes *FZD4* (EVR1, OMIM: 133780), *NDP* (EVR2, OMIM: 305390), *LRP5* (EVR4, OMIM: 601813), *TSPAN12* (EVR5, OMIM: 613310), *CTNNA1* (EVR7: 617572), and one locus on chromosome *11p13-p12* (EVR3, OMIM: 605750, exact gene unknown) have been confirmed to be involved in FEVR (Robitaille et al., 2002, Chen

et al., 1993a, Toomes et al., 2004, Jiao et al., 2004, Poulter et al., 2010, Nikopoulos et al., 2010, Collin et al., 2013, Panagiotou et al., 2017, Downey et al., 2001). In addition, *KIF11*, *ATOH7* and *ZNF408* (EVR6, OMIM: 616468) have been shown to be mutated in patients diagnosed with FEVR but it's unclear if these are misdiagnosed FEVR cases (Robitaille et al., 2014, Khan et al., 2011). Furthermore, *RCBTB1*, *ILK*, and *JAG1* have all been reported as potential new FEVR genes, but further evidence is needed (Wu et al., 2016, Park et al., 2019, Zhang et al., 2019).

It is well acknowledged that the impairment of the Norrin- β -catenin signalling pathway is one of the main molecular pathogenesis of FEVR and its related retinopathies Norrie disease (OMIM: 310600), retinopathy of prematurity (ROP, OMIM: 133780), Coats disease (OMIM: 300216), Persistent Fetal Vasculature (PFV) and osteoporosis pseudoglioma syndrome (OPPG, OMIM 259770) (Drenser, 2016). This signalling pathway is critical for the angiogenesis of the eye, studies showed that the aberrations of the pathway could result in vascular abnormalities in animal models as well as the human cases (Clevers, 2009). The Norrin- β -catenin signalling pathway is a derivative of the Wnt- β -catenin pathway, however there are important components of the signalling system which are different. In the Norrin- β -catenin signalling pathway, Norrin acts as the ligand, instead of a Wnt ligand, to binds with the unique receptor complex composed by FZD4, LRP5 and TSPAN12 to activate the pathway. Only FZD4 is used in Norrin- β -catenin signalling, whereas in Wnt signalling one of ten different frizzled receptors can be utilised (Xu et al., 2004). Similarly, TSPAN12 acting as an axillary receptor is only required for Norrin- β -catenin signalling (Junge et al., 2009).

Without Norrin binding, cytoplasmic β -catenin is degraded, which prevents its translocation into the nucleus to interact with T-cell factor (TCF) or lymphoid enhancing factor (LEF) and thus inhibits the expression of the target genes. However, upon Norrin binding, the degradation of β -catenin is decreased, which allows it to accumulate in the cytoplasm and then translocate into the nucleus and interact with TCF/LEF and activate the expression of the target genes (**Figure 2.**) (Ye et al., 2010).

Even though it is still unclear how Norrin binding activates the signalling pathway, Ke and co-workers found that Norrin works as a dimer, with each component separately binding to an FZD4 dimer or oligomer with high affinity (Ke et al. 2013). Similarly, Chang and colleagues found that FZD4 presents as a dimer regardless of whether it is binding with Norrin or not in the cell membrane, and that the combination of only FZD4 binding with Norrin is not sufficient for signal activation of the pathway (Chang et al. 2015). LRP5, or family member LRP6, play a pivotal role as a co-receptor to activate the pathway. Besides, TSPAN12 has been reported to act as a molecular chaperone to bring FZD4 to interact with the Norrin dimeric structure and facilitate the signalling system (Chang et al., 2015, Ke et al., 2013, Lai et al., 2017).

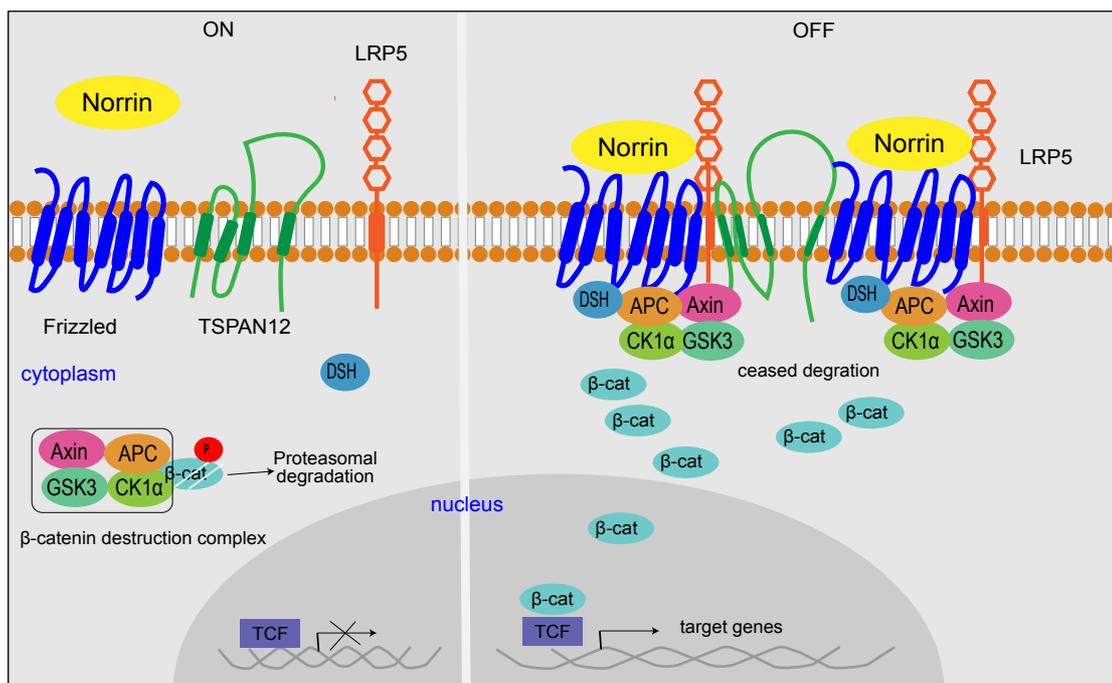


Figure 2. Overview of the Norrin- β -catenin pathway. Without Norrin binding, cytoplasmic β -catenin is recognized by the complex that composed by Adenomatous polyposis coli (APC), Glycogen synthase kinase (GSK-3), Casein kinase 1(CK1) and Axin), and followed by proteasomal degradation which then inhibits the translocation into the nucleus. Whereas with Norrin binding, the released signal leads the destruction protein complex to accumulate to the cell membrane through the activation of Dishevelled (DSH). This process inhibits the degradation of cytoplasmic

β -catenin, allowing it to move into the nucleus, β -catenin then interacts with TCF/LEF in the nucleus, and then activates the expression of the target genes.

1.1.3.1 Norrin

The cysteine-knot growth factor Norrin is encoded by the *NDP* gene which is located at chromosome *Xp11.3*. *NDP* is composed of three exons which encode the 133 amino acids ligand. Norrin is expressed and secreted in the Müller cells and acts upon the retinal endothelial cell (Ye et al., 2009). It contains the important cysteine knot motif, and three β loops (β 1- β 2, β 3- β 4, β 5- β 6) and a separate β 7 at the C-terminus (Chang et al., 2015). Mutations in the *NDP* gene have been found to cause Norrie disease, X-linked FEVR, Coats disease, Persistent Fetal Vasculature (PFV) and Retinopathy of Prematurity (ROP) (Sims, 2014). As described above, Norrin acts as the ligand and is capable of activating the Norrin- β -catenin pathway through two Norrin monomers forming a dimer structure allowing the two β -loops (β 1- β 2 and β 5- β 6) to bind with the cysteine-rich domain (CRD) of FZD4 (Chang et al., 2015). Norrin also binds to the Beta-propeller 1 or Beta-propeller 2 domains of LRP5 by a separate binding site, however, this remains unclear and need further evidence (Ke et al., 2013). There is also indirect evidence that Norrin has an interaction with the extracellular loops of TSPAN12 but further evidence is required to confirm this (Lai et al., 2017). The *Ndp* knockout mice replicate the human phenotypes and show delayed development of retinal blood vessels, incomplete development of the retinal vasculature and PFV (Richter et al., 1998, Berger et al., 1996). The mice also show defects in the vasculature of the inner ear and cerebellum (Luhmann et al., 2008, Rehm et al., 2002), and female mice have fertility problems (Luhmann et al., 2005) Progressive hearing loss is often a feature of Norrie disease but to date, no fertility or cerebellum defects have been reported in patients with *NDP* mutations.

1.1.3.2 FZD4

Frizzled class receptor 4 (FZD4) is encoded by the *FZD4* gene, which is located at chromosome *11q14.2*. *FZD4* contains 2 exons and encodes a protein of 537 amino

acids. FZD4 belongs to the Frizzled superfamily of receptors and is composed of seven-transmembrane regions, a highly-conserved CRD domain (~120 residues long) at the extracellular N-terminal and a cytoplasmic C-terminal containing a Lys-Thr-X-X-X-Trp (KTXXXW) PDZ-recognition motif (from residue number 497 to 504) (Bang et al., 2018). Mutation in *FZD4* lead to autosomal dominant or autosomal recessive inheritance pattern of FEVR (Musada et al., 2016, Robitaille et al., 2002). FZD4 acts as the receptor in the Norrin- β -catenin pathway, and the CRD domain is essential for Norrin-binding and initiating the signalling system. As well, the KTXXXW motif is important in the binding with the PDZ domain of the cytoplasmic downstream Dvl protein (Zhang et al., 2011). *Fzd4* null mice models also show incomplete development of capillary vessels in the retina, vascular defects in the inner ear and cerebellum along with infertility (Xu et al., 2004, Wang et al., 2001, Hsieh et al., 2005).

1.1.3.3 LRP5

Low-density lipoprotein receptor-related protein 5 (LRP5) is encoded by the *LRP5* gene, which is located on chromosome *11q13*. LRP5 is composed of 1615 amino acids (encoded by 23 exons) and is a member of low-density lipoprotein receptor (LDRP) superfamily. LRP5 contains four beta-propeller domains (composed of six YWTD repeats), one transmembrane domain and three LDL (low density lipoprotein) receptor domains (<https://www.uniprot.org/uniprot/O75197>). As described above, in the Norrin pathway, LRP5 acts as a coreceptor and interacts with FZD4 and Norrin to activate and induce the target gene transcription. LRP5 binds with Norrin through Beta-propeller 1 or Beta-propeller 2 domains. The C-terminus of LRP5 also contains five repetitive PPP(S/T)P motifs, which mediate the downstream signal of Norrin- β -catenin pathway by linking with Axin (Ke et al., 2013). Toomes et.al 2004 confirmed that mutations in *LRP5* could cause autosomal dominant FEVR, and Jiao et.al 2004 first reported mutations in *LRP5* linked with the recessive FEVR cases (Toomes et al., 2004) (Jiao et al., 2004). Recessive mutations in *LRP5* also cause OPPG which has the same eye phenotype as stage 5 FEVR but also has osteoporosis as a feature

(Gong et al., 2001). FEVR patients with mutations in *LRP5* also have reduced bone density phenotypes (osteopenia or osteoporosis) although these are often mild and require DEXA scanning to diagnose them (Toomes et al., 2004) (Downey et al., 2006). The *Lrp5* knockout mice models show defects in retinal vascular development abnormalities and the reduced bone density features (Charette et al., 2017, Kato et al., 2002).

1.1.3.4 TSPAN12

TSPAN12 is located on human chromosome *7q31* and contains 8 exons which encodes a 305 amino acids auxiliary membrane protein. TSPAN12 contains four transmembrane domains and plays an important role in cell adhesion, migration, and signalling (Bucher et al., 2018). Studies showed that mutations in *TSPAN12* cause autosomal dominant and recessive FEVR (Poulter et al., 2010) (Poulter et al., 2012, Nikopoulos et al., 2010). As described above, TSPAN12 interacts with FZD4 and Norrin through the extracellular loop to stabilize and facilitate the activation of Norrin signalling pathway reporter activity (Poulter et al., 2010). The *Tspan12* knockout mice have an identical eye phenotype to the *Fzd4* knockout mice and it was this similarity which identified TSPAN12 as a key component of Norrin- β -catenin pathway (Junge et al., 2009).

1.1.3.5 β -catenin

β -catenin is encoded by the *CTNNB1* gene, which is located on chromosome *3p22.1*. *CTNNB1* is composed of 16 exons which encode the 781 amino acids protein. β -catenin plays a vital role in cell adhesion and cell signalling. The important regions in β -catenin include the N-terminal domain (contain α -catenin binding site), the central 12 Armadillo repeats, Helix C and C-terminal domain. Somatic mutations in *CTNNB1* are associated with many cancers (Dubruc et al., 2014) (van Schie and van Amerongen, 2020). De novo heterozygous mutations in *CTNNB1* are associated with a severe intellectual disability syndrome and a few cases of autism have also been reported (O’Roak et al., 2012, De Ligt et al., 2012). However, recently, FEVR has

been shown to be a feature of this intellectual disability syndrome (Panagiotou et.al 2017). In addition, heterozygous mutations which just affect the C-terminal domain of β -catenin have also been shown to cause non-syndromic FEVR and these mutations can be inherited (Panagiotou et al. 2017). β -catenin is the key downstream mediator of the Norrin– β -catenin pathway. Knockout *Ctnnb1* mice are embryonic lethal and heterozygotes have a wild-type phenotype showing that these mice don't replicate the human phenotypes (Panagiotou et al. 2017). However, targeted deletion of β -catenin in the vasculature show defects in the retinal vasculature development (Panagiotou et al. 2017).

1.1.3.6 ZNF408

ZNF408 is located at chromosome *11p11.2* and contains 5 exons which encodes a zinc finger protein (consisting of 720 amino acids). Heterozygous mutations in *ZNF408* are associated with FEVR (Collin et al., 2013), but recessive mutations are associated with retinitis pigmentosa (Avila-Fernandez et al., 2015). *ZNF408* is highly expressed in the retina and may play a role in the development and growth of retinal vascular development (Avila-Fernandez et al., 2015). In the *znf408*-null zebrafish model, the deletion of the gene led to abnormalities in the retinal vascularization, but the pathogenic process of how this proposed transcription factor causes FEVR or whether it is part of the β -catenin pathway is unknown (Collin et al., 2013).

1.1.3.7 EVR3

A locus on chromosome *11p13-p12* has been mapped and reported in a large dominant FEVR family but the gene is not yet published (Downey et al., 2001).

1.1.3.8 KIF11

KIF11 is located on chromosome *10q23.33* and has 22 exons which encodes a kinesin-5 (Eg5) protein with 1056 amino acids. Robitaille and colleagues first reported that heterozygous mutations in *KIF11* are associated with FEVR (Robitaille et al., 2014). Previously dominant mutations in *KIF11* had been reported as a cause of microcephaly with or without chorioretinopathy, lymphedema, or mental

retardation (MCLMR syndrome) which has an ocular phenotype which resembles FEVR (Ostergaard et al., 2012). However, it is now clear that the microcephaly caused by *KIF11* mutations can be mild and therefore go unreported which is why cases are diagnosed with FEVR (Robitaille et al., 2014). KIF11 is involved in tumorigenesis and angiogenesis and the lack of KIF11 has been shown to impaired endothelial cell proliferation (Exertier et al., 2013). However, whether KIF11 has a role in the β -catenin pathway is still unclear (Exertier et al., 2013). In *kif11*-knockout zebrafish experiment, defects in the glial cells were shown which proved Eg5 plays a role in retinal angiogenesis (Barresi et al., 2010).

1.1.3.9 Other genes and locus

There are also several genes and loci that been reported in FEVR, but the evidence and understanding of the molecular mechanism of these genes in FEVR are not clear or unconfirmed. Park et.al 2019 recently reported three heterozygous missense mutations in the *ILK* gene on chromosome *11p15.4* (encodes Integrin-linked kinase protein) that may cause FEVR. The *Ilk*-null mice also showed similar FEVR-like vascular defects, which may indicate it plays an important role in the angiogenesis of the retina (Park et al., 2019).

The Notch ligand JAG1 is encoded by the *JAG1* gene, which is located on chromosome *20p12.2*. JAG1 is involved in the evolutionarily conserved Notch signalling pathway. Heterozygous *JAG1* mutations usually cause Alagille syndrome (OMIM: 118450) which is an autosomal dominant disease with multiple defects involving the heart, liver and other systems (Fischetto et al., 2019). Recently, Zhang and colleagues identified three heterozygous missense mutations in *JAG1* in FEVR patients and proposed it as a candidate FEVR gene. Knock-out Notch mice show abnormalities and disorganized of blood vessels indicating a defect in normal angiogenesis. In addition, cell based assays using *Jag1*-null mice endothelial cells also retarded vessel growth which is similar to FEVR (Zhang et al., 2019).

Wu et. al 2016 reported *RCBTB1* (located on chromosome *13q14.3*) as a candidate causative gene of FEVR as they identified heterozygous predicted null mutations in FEVR patients. They showed knockdown of *rcbtb1* in a zebrafish model leads to vascular defects. Similarly, in a cell based assay, they showed that the *RCBTB1* mutations reduced the activation of a β -catenin transcriptional reporter construct indicated that RCBTB1 may act in the Norrin- β -catenin pathway (Wu et al., 2016). More recently, recessive mutations in *RCBTB1* have also been reported in retinitis pigmentosa (RP) (Coppieters et al., 2016). This is a similar situation to *ZNF408* where dominant mutations lead to FEVR and recessive mutations cause RP.

Atonal homolog 7 is encoded by the *ATOH7* gene, which is located at chromosome *10q21.3*. Khan et.al 2011 first reported mutations in *ATOH7* cause recessive FEVR (Khan et al., 2011). Mutations in *ATOH7* have also been reported in other related eye disorders which all have a phenotype overlapping FEVR, non-syndromic congenital retinal nonattachment (NCRNA disease) (Ghiasvand et al., 2011) and PFV (Prasov et al., 2012). Mice studies have shown that ATOH7 is a transcription factor which plays an essential role in the development and formation of the retinal ganglion cell and optic nerve (Brown et al., 2001).

Finally, studies have shown that loci on chromosomes *22q11.2* and *7q22* may harbour genes that cause FEVR through the identification of FEVR patients with cytogenetic mutations at these regions (Gilmour et al., 2009, Gandhi et al., 2014) (Miller et al., 2015).

1.2 Human Variation and Mutations

Genetic mutations refer to the permanent alteration in the genomic DNA molecules. Common types of mutation at the DNA level include insertion, deletion, substitution, translocation and inversion. Mutations are often characterised by their effect on protein products. The substitutions of the base pairs in the DNA could lead to amino acid substitution (nonsynonymous/missense variant) or have no effect on the protein product (synonymous). In some instances, the DNA substitution could result in a

nonsense variant if the amino acid codon turns into stop codons (TAA, TAG or TGA) and result in the premature termination of the protein. DNA substitutions located in non-coding regions of a gene, such as a splice site or a promotor region, could also cause a dramatic change to the protein by either causing splicing defects in the mRNA or altering the transcription of the mRNA. A small deletion or insertion in the DNA can cause a shift of reading frame in the mRNA leading to a frameshift variant. Alternatively, if the deletion or insertion is the integer multiple of three, it is known as indel deletion and indel insertion variants (Mark O Johnston, 2006). Large scale mutations are often called structural variants or mutations, and these can disrupt the protein by altering the gene sequence.

Unfortunately, while all of these types of mutations are known to cause Mendelian disease, the majority of them are part of the natural variation seen among the human population. This can make it a significant challenge to be able to identify the causative mutation in many rare diseases with small numbers of patients and often no family history (Gilissen et al., 2012). This is now one of the biggest issues in diagnostic and research laboratories as the majority of testing is based on whole exome or whole genome sequencing methodologies. As an example of the size of this issue, an average genome is expected to contain 3-4 million variants. Just focusing on those affecting the coding regions reduces this figure to ~25,000 and removal of synonymous variants reduced this to 5000 variants. However, removing synonymous variants is a bit risky, as they can still cause disease by the change of codon usage bias especially in the exonic splice enhancers regions (ESEs) (Brandis and Hughes, 2016). For example, Savissar and Hurst found that the synonymous mutations in the ESEs could lead to large fitness or the clinical effect due to the change of translational speed or accuracy (Savisaar and Hurst, 2018). If common variants are removed this leave ~100-500 candidate variants (Gilissen 2012). Even when only looking at known causative genes this means there is often a number of plausible candidates.

It is therefore very important that databases like gnomAD (Karczewski et al., 2020) and dbVAR (Lappalainen et al., 2013) and resources like the 100,000 genomes are

available to catalogue the variation observed in large populations. Similarly, databases with patient variation data are becoming more and more important to allow quick and easy interrogation of whether variations have been found previously in patients with the same cohort, for example the Human Gene Mutation Database (Stenson et al., 2020) and ClinVar (Landrum et al., 2018). However, all databases are only as reliable as the data they contain and in many historical papers describing mutations the evidence that mutations are causative, especially missense variants, can be limited.

1.2.1 Mutation pathogenicity classification (ACMG)

The accurate interpretation of the variant effect on human diseases are essential to the molecular diagnosis and the development of gene therapy methods. Several guidelines have been published in preventing confusion of variant calling and interpretation, for example, a unified variant naming method was established by the Human Genome Variation Society (HGVS), which is of great significance for the consistent sharing and using of variant data (den Dunnen et al., 2016). As well as this, in 2015, the American College of Medical Genetics and Genomics (ACMG) announced a variant pathogenicity classification guideline, which provides a five-tier terminology to describe a variants clinical significance (pathogenic, likely pathogenic, uncertain significance (VUS), likely benign and benign) (Richards et al., 2015a). To determine the clinical significance of the variants, the collection of evidence from various sources is needed. In the ACMG guidelines, a total number of 28 criteria are provided, which include pathogenic evidences and benign evidences (Full ACMG criteria descriptions can be accessed through the online ACMG calculator tool: https://www.medschool.umaryland.edu/Genetic_Variant_Interpretation_Tool1.htm) (Kleinberger et al., 2016).

A range of comprehensive criteria are considered in the ACMG guidelines including variant types information, population frequency data, *in silico* prediction data, functional data, segregation data and *de novo* status of the variant. For example, the

type of variant could be an important factor and null variants (nonsense, frameshift etc) are classified under very strong pathogenic evidence criterion (PVS1).

For *in silico* prediction data, the consensus prediction results of variants are obtained from multiple computational algorithms and could be used as supportive evidence of pathogenic (PP3) or benign (BP4) status in the ACMG). Several reliable and commonly used predictors recommended in the ACMG include Sorting Intolerant from Tolerant (SIFT) (Vaser et al., 2016), Polymorphism Phenotyping v2 (PolyPhen-2) (Adzhubei et al., 2010), Protein Variation Effect Analyzer (PROVEAN) (Choi and Chan, 2015), Combined Annotation Dependent Depletion (CADD) (Rentzsch et al., 2019), and Mutation Taster (Schwarz et al., 2014) but there are many more available.

In addition, incorporating evidence from reliable functional assays or well-established functional evidence could be used as pathogenic (PS3) or benign (BS3) supporting evidence. Moreover, if a reputable source reported the variant as pathogenic (PP5) or benign (BP6), it could also be used as a supportive evidence in the ACMG. Such as the reliable expert judgement or reputable variant datasets which include human annotation of variant classification (like ClinVar) (Landrum et al., 2020).

The use of different reliable population databases (like gnomAD), which provides the frequency of the variants in human beings is also an important factor in variant interpretation, such as when the allele frequency is greater than 5% in human cohorts, the strong benign variant (BS1) should be selected in the ACMG. As well as, if the variant is absent from these databases or large control cohorts, the pathogenic moderate evidence (PM2) should be applied.

However, there are also limitations in using ACMG classification guideline, it could be only used to classify germline variant for Mendelian diseases. There are also several challenges on how to ensure the greatest consistency when interpreting variants. The quality and consistency of variant classification could be different even after human annotations, as the probability of the variant pathogenicity could be

weighted differently (Variant classification and interpretation-workshop report, 2017). For example, segregation criterion could be weighted for different levels of evidence, as it subdivided into supporting, moderate or strong evidence according to the number of individuals segregated with the disease. However, no quantitative methods were provided by the ACMG guideline. In addition, intercalating the 28 criteria for each variant to be interpreted can be time consuming, especially for the reclassification of large numbers of previously reported variants. In order to solve the above challenges, several strategies have been reported in recent years. For example, a quantitative approach underlying the cosegregation evidence has been developed in 2016 (Jarvik and Browning, 2016). Similarly, several online calculation tools based on the supplied evidence of ACMG are also available, like the ACMG calculator prevented by University of Maryland (https://www.medschool.umaryland.edu/genetic_variant_interpretation_tool1.html/) (Kleinberger et al., 2016). These calculators often need the entire human annotation on all evidence criteria. However, there are also several web tools that provides automatically annotation of the variants, like InterVar, which provides automatically annotation for 18 criteria in the ACMG (Li and Wang, 2017).

1.3 Protein modelling

Understanding protein function is vital in biological, medical and pharmaceutical fields, and plays a key role in understanding molecular pathogenesis in the human diseases. The determination of a protein's structure is an important step in understanding its function. Several experimental techniques, like nuclear magnetic resonance (NMR) and X-ray crystallography (XRC), are available in determining the structure. However, due to the limit of some proteins could not be crystalized (which is an essential step in the XRC) and the time-consuming properties of these experimental techniques, the computational modelling methods becomes valuable in predicting the three-dimensional (3D) structure of the protein (Soni and Madhusudhan, 2017).

1.3.1 Protein 3D structure prediction and optimization

There are currently three methods that are commonly used in protein modelling: homology modelling, threading method and *de novo* (ab initio) method. As similar amino acid sequences could potentially share similar structures, homology-based methods use known structures as templates (with amino acid sequence homology higher than 30%) to predict the structure of uncharacterised proteins. The commonly used online homology modelling tools are SWISS-MODEL, Phyre2 and Modeller, although others are available (Waterhouse et al., 2018, Kelley et al., 2015, Webb and Sali, 2016).

The threading approach (fold recognition) also predicts protein structure based on known homologous protein structures, but this method relies on proteins which have similar folds or structure rather than the actual amino acid sequence. This method is viable as there are only a limited number of protein folds seen in nature so models can be created from evolutionarily distant proteins. Online tools are available, like I-TASSER (Yang and Zhang, 2015).

The *de novo* (*Ab initio*) method directly predicts the protein structure through its amino acid sequence without using any templates. It focuses on using folding energetics and statistical methods and is the least accurate among these three modelling approaches (Schwede, 2013).

Even though computational protein modelling has helped to bridge the gap between amino acid sequence and protein structure, computational protein modelling is limited and it is possible that the predictions contain illegal bond lengths, bond angles or peptide dihedral torsion angles etc (Park et al., 2013). Thus, there are some procedures which can be applied to refine and optimize the protein prediction structures to increase their accuracy such as methods based on molecular dynamics (Park et al., 2018). These optimization approaches can be basically divided into localized (applied to the small part of the structure) and global optimization methods (applied to the whole structure) (Gront et al., 2012).

1.3.2 Protein structure quality evaluation

Protein structure quality evaluation is also an important step in protein modelling, which could present the basic view of the modelling reliability. There are many methods which can be applied, including ERRAT, Ramachandran plot, Verify 3D among others (Colovos and Yeates, 1993, Ramachandran et al., 1963, Eisenberg et al., 1997). ERRAT is mainly based on the statistical analysis of non-bonded interactions of atoms. It relies on the observation that certain atoms are distributed nonrandomly in proteins, due to energetic and geometric reasons. Thus, errors in a predicted protein structure are more likely to cause random distribution of these non-bonded atom types. It uses an overall quality factor to present the statistical error in each residue in the structure (Colovos and Yeates, 1993).

Ramachandran plots use a scatter plot of two peptide torsion angles (Phi: ϕ and Psi: ψ) between residues to reveal the allowed and non-allowed backbone conformations of the protein. Many web-tools are available for plotting the Ramachandran plot of a protein structure including RAMPAGE, PROCHECK and other (Lovell et al., 2003) (Laskowski et al., 1996).

1.3.3 Structural comparison of protein variants

Comparative structural data could reveal the similarity and difference between protein structures, which could contribute to highlight the evolutionary relationship, or help to get the knowledge of unknown protein function as previously discussed. It is well known that an amino acid substitution (point mutation) could cause biophysical-related changes to a protein structure. Thus, it is always worthwhile to compare the mutant protein structures with the wildtype protein structure to help highlight the effect of the point mutation on the protein structure. For example, in order to understand the effect a point mutation has on protein structures, MuPro uses machine learning methods (support vector and neural networks) to predict the protein stability changes of a point mutation by using sequence and structural information with the prediction accuracy of 84%. It generates a statistical score (folding energy

change: $\Delta\Delta G$) to describe relative stability changes, if the score is positive (or negative), it means the point mutation predicted to be increased (or decreased) the protein stability, the higher the score the higher the predicted confidence (Cheng et al., 2005) (Cheng et al., 2006). There are other approaches available to evaluate point mutations effects based on superimposing mutant and wildtype protein models. One model deviation method uses the Root-Mean-Square Deviation (RMSD) score to quantitate the differences, with lower values associated with higher similarity (Soni and Alok, 2018).

1.4 Aims

The development of NGS-sequencing has led to the rapid identification of novel genes and variants in the patients. However, it also brings new challenges to determine the clinical function of the variants. Over the years, many variants and genes were reported in FEVR patients, however, the majority of them were reported using different classification systems. Thus, in this study, the pathogenicity of all reported variants and available unpublished data of FEVR patients were reassessed through ACMG guideline. The mutation spectrum and database of FEVR were generated in order to give an overview of genetic basis of FEVR.

Thus, the aim of the project was threefold:

- (1) to reassess the pathogenicity of all published FEVR-causing variants.
- (2) to create a mutation spectrum and online database for FEVR.
- (3) to compare *in silico* models of mutant proteins with wildtype models in FEVR causative genes.

2. Methodology

2.1 Ethical approval

Ethical approval for this work was provided by the Leeds East Teaching Hospitals NHS Trust Research Ethics Committee (Project numbers 03/362 and 17/YH/003).

2.2 FEVR mutation database

2.2.1 Literature Search

All peer-reviewed publications which reported variant information in FEVR patients before Feb 2020 were collected through a PubMed search. For each variant the following information was extracted where available and annotated in an Excel file: the variant combinations, phenotype details of patients (including clinical features, age of onset or examination, gender, population and treatment records), screening methods, frequency in patients' cohorts and controls, cosegregation information, functional assays result and pathogenicity classification of the variants reported by the authors. Obvious duplications, like the patients reported from the same lab or different publications from the same author group, were removed for further analysis.

2.2.2 Unpublished data

Unpublished variant data previously identified in FEVR cases from the Toomes' Lab were collected. As described above, the variant and phenotype information were all summarized into the same Excel file when available. These variants were identified by the following researchers: Dr James Poulter, Dr Helen Bottomley, Dr Evangelia Panagiotou, Dr Denisa Dzulova, Dr David Gilmour and Dr Carmel Toomes.

2.2.3 Variant annotation

Variants were annotated according to the latest Human Genome Variation Society nomenclature guidelines (<http://varnomen.hgvs.org/>) (den Dunnen et al., 2016). For numbering the cDNA, the A of the translation start codon ATG was numbered as position number 1 (+1) and the initiation codon as codon 1 in the reference sequence. HGVS protein nomenclature and HGVS cDNA level change were converted using VarSome tool (<https://varsome.com/>) (Kopanos et al., 2019). LUMC Mutalyzer (version 2.0.32) was used to check gene transcript symbols and variant HGVS nomenclature (<https://mutalyzer.nl/>). Mutalyzer Position Converter tool (<https://mutalyzer.nl/position-converter>) was specifically used to convert cDNA position into genomic locations in two human reference genomes, CRCh37 (hg19,

assembled in 2009) and CRCh38 (hg38, assembled in 2013) (Wildeman et al., 2008). Variant allele frequency from different populations, like Finnish European, non-Finnish European populations and total allele frequency, were taken from gnomAD v2.1.1 (<https://gnomad.broadinstitute.org/>) (Karczewski et al., 2020). In addition, ClinVar ID and dbSNP ID (rs number) of the variants were also searched through HGVS nomenclature in dbSNP database (<https://www.ncbi.nlm.nih.gov/snp/>) and ClinVar database (<https://www.ncbi.nlm.nih.gov/clinvar/>) (Landrum et al., 2018). Ambiguous variant nomenclatures reported in several previous publications were either resolved manually or removed from further analysis.

2.2.4 *In silico* pathogenicity predictions

Several computational tools were used to predict the effect of the variants on protein structure and function. Pathogenicity of all non-synonymous variants in FEVR patients were predicted using SIFT and PolyPhen-2 (HumVar dataset) (<http://genetics.bwh.harvard.edu/pph2/>) (Vaser et al., 2016) (Adzhubei et al., 2010). Missense and indels mutations were also predicted using PROVEAN (<http://provean.jcvi.org/index.php>) (Choi and Chan, 2015). In addition, the effect of all of the variants, including indels, missense, frameshift and nonsense variants, were predicted using the CADD scoring system (<https://cadd.gs.washington.edu/>). Scaled C-scores were used in CADD, and the single fixed cutoff value was set at 15 in this study (a score prediction ≥ 15 , the variant was considered as deleterious in the CADD predictor) (Rentzsch et al., 2019).

2.2.5 Pathogenicity classification

The pathogenicity of all of the variants reported in FEVR cases were assessed according to the ACMG in a five-tier classification system (Pathogenic, Likely pathogenic, uncertain significance, likely benign and benign) (Richards et al., 2015b). The criteria of evidence were selected by combining all the variant information reported in previous publications of FEVR cases and updating frequency and pathogenic predictions as described. For example, when classifying pathogenic

variants, there are very strong (PVS1), strong (PS1-PS4), moderate (PM1-PM6) and supporting evidence criteria (PP1-PP5) options in the ACMG guidelines. After categorizing the variant evidence criteria, all of the criteria were then incorporated together in the online ACMG scoring algorithm (Genetic Variant Interpretation Tool) (https://www.medschool.umaryland.edu/Genetic_Variant_Interpretation_Tool1.html) to reach for the final classification. If the evidence of the variant was insufficient, the variant was classified as a variant of uncertain significance (VUS). More detailed methods used the study of several ACMG criteria were described below.

In the ACMG guideline, classification of the variants in a reputable source can be used as a single piece of evidence. In this study, the ClinVar database (<https://www.ncbi.nlm.nih.gov/clinvar/>) was used as reputable source. If the variant was annotated as pathogenic or benign in ClinVar, the PP5 supporting evidence or BP6 criterion were selected respectively.

Cosegregation analysis of the variant can be taken as either PP1-supporting evidence, PP1-Moderate evidence, or PP1-strong evidence based on the extent of the segregation. In this study, if the variant found to be segregated with FEVR in at least three families from diverse ethnic backgrounds, the evidence of segregation was selected as strong PP1, whereas if segregation analysis found in two different families (or a large one family pedigree as more than 10 affected family members), the PP1 moderate evidence was selected. Similarly, if the variant was only found to be segregated in one small family (less than 10 affected people), the supporting evidence PP1 was selected. In addition, as reduced penetrance is common in FEVR, and many patients do not undergo FFA, only clinically diagnosed cases were included in the cosegregation analysis.

Multiple *in silico* predictions results were only considered as a single piece of evidence in the variant interpretation. PolyPhen-2, SIFT, PROVEAN and CADD scores were used as computational evidence in this study. Only when all four predictors suggested the same consistent result, either damage to the gene or gene product (PP3) or reported as no impact (BP4), were the criteria selected respectively.

2.2.6 Mutation spectra

Mutation spectra were generated on protein structures using Adobe Illustrator software. The ACMG results were used for all of the variants reported in known Norrin- β -catenin pathway genes in FEVR. The gene sequence, transmembrane region, domain structure and disulphide bond information of these genes/proteins were gathered from UniProt database (<https://www.uniprot.org/>).

2.2.7 Statistical analysis

Statistical analysis was performed using Excel.

2.2.8 Phenotype-genotype correlation

The severity of the FEVR patients, when phenotype data were available, were classified based on the five-tier staging method presented by Pendegast and Trese, 1998 (Pendegast and Trese, 1998). The phenotype-genotype correlation was then performed in known Norrin- β -catenin pathway FEVR genes based on the severity of the phenotype in accordance with different variant types and genes.

2.2.9 LOVD submission

Detailed information concerning variants, clinical phenotype, screening methods of FEVR were all uploaded to the Leiden Open Variation Database (LOVD) database by Jasmine Chen and Dimitra-Ilektra Lerou under the supervision of the author, Dr Carmel Toomes and Professor Johan den Dunnen (Leiden University, Netherlands).

2.3 Three-dimensional protein modelling

2.3.1 Wildtype protein structure

Three-dimensional protein modelling and comparison methods used in the study were based on the methods performed by Seemab et al. 2019 with only minor changes (Seemab et al., 2019).

The amino acid sequence of known β -catenin pathway FEVR genes were collected in FASTA format from UniProt (<https://www.uniprot.org/>). The corresponding three-

dimensional (3D) protein structures of the genes were then predicted through I-TASSER software (<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>) and Phyre2 web tool (<http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>). The best structure was selected using the largest C-score in I-TASSER. Structures were then compared using its coverage percentage, Ramachandran plot and ERRAT score. Ramachandran plots were generated through Rampage software (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>). ERRAT scores were generated through its official website (<https://servicesn.mbi.ucla.edu/ERRAT/>). The best predicted structures were then selected between I-TASSER and Phyre2 based on the above quality evaluation tool, and then visualized through UCSF chimera (Pettersen et al., 2004).

2.3.2 Mutant protein structure

Missense mutant structures were modelled using Modeller (version 9.24) software using wildtype 3D structures as the templates. A basic modelling script was obtained from Modeller 9.24 Tutorial (<https://salilab.org/modeller/>), and were provided in the *Appendix 3*. Mutant sequence was annotated arbitrary using the wildtype sequence, and the best mutant modelling structure was selected based on the smallest Discrete Optimized Protein Energy (DOPE) score. Overall quality of the mutant structures was then analysed using ERRAT and Ramachandran plots as described in section 2.3.1.

2.3.3 Comparison of wildtype and mutant protein structures

The impact of missense variants on the conformation of the protein structures were evaluated through superimposing the mutant structure and the wildtype structure together in the UCSF chimera software (<https://www.cgl.ucsf.edu/chimera/>) (Yang et al., 2012). Root Mean Square Deviation (RMSD) scores were calculated after superimposing using UCSF chimera software (Reply log function) (Kufareva and Abagyan, 2012). The major deviated region between missense mutant structure and wildtype structure was evaluated using the backbone RMSD. The impact of the

missense variants on protein stability was also evaluated using the MuPro online server (<http://mupro.proteomics.ics.uci.edu/>).

3 Results

3.1 FEVR Variant database

The first aim of this project was to perform a PubMed search to identify all of the papers describing FEVR mutations since the first report describing *NDP* mutations in 1993 (Chen et al., 1993b). In total, 112 papers were identified, describing mutations in 11 different genes with 537 unique variants. The full dataset is available in the accompanying Excel File (All FEVR variants and cases) in **Appendix 2**.

3.1.1 Variants annotation

The data in the papers was not all annotated in a uniform way with different reference transcripts being used by different authors and a variety of different mutation annotation styles. To overcome these issues, all of the variants reported in the FEVR patients were standardised by updating the annotations to match the latest HGVS nomenclature guidelines and mapping all variants onto the following transcripts reference sequences: *FZD4* (NM_012193.3), *NDP* (NM_000266.3), *CTNNB1* (NM_001904.3), *KIF11* (NM_004523.3), *LRP5* (NM_002335.2), *TSPAN12* (NM_012338.3), *ATOH7* (NM_145178.3), *ZNF408* (NM_024741.2 and NM_001184751.1), *RCBTB1* (NM_018191.3), *JAG1* (NM_000214.2) and *ILK* (NM_004517.2). Unfortunately, several reported variants contained mistakes so that they were not understandable or contained conflicting annotations (**Table 1**). For example, for the variant c.316T>C (p.C106G) in *FZD4* reported by Robitaille and colleagues, the cDNA level nomenclature did not match the given protein change, if the cDNA level nomenclature is assumed to be right, the corresponding protein level nomenclature should be p.C106R (Robitaille et al., 2011). Thus, all of these variants were removed from further analysis in the study. In cases where the nomenclature was changed from that published, the original reported variant and the updated annotations are detailed in **Appendix 2** in the accompanying raw data Excel File.

Table 1. Nomenclature errors identified for variants in previous publications

References	Gene	cDNA change	Protein change	Problem
Wang et al. 2019	<i>FZD4</i>	c.1043G>A	p.Arg348Gln	The reference nucleotide at the position is not 'G', it is 'A'. The reference amino acid at position 348 of transcript is 'His', not 'Arg'
Dailey et al. 2015	<i>FZD4</i>	c.40Del/ins	NA	The inserted sequence was not given
Robitaille et al. 2011	<i>FZD4</i>	c.316T>C	p.C106G	cDNA combination did not match the given protein nomenclature.
Keser et al. 2017	<i>NDP</i>	c.336_402IVS4del	NA	Nomenclature not understandable
Wu et al. 2007	<i>NDP</i>	c.*717T>C	3'UTR	The reference nucleotide at the position is not 'T', it is 'A'.
Li et al. 2018	<i>KIF11</i>	c.1030dupT	p.S348Efs*8	The reference in c.1030 is not T, it is 'G'.
Tang et al. 2017	<i>TSPAN12</i>	c.916_918+3delTAAAAA	p.*306Eext*35	Nomenclature was for an intronic position, but the coding position 918 is not at the 3' end of the exon
Keser et al. 2017	<i>TSPAN12</i>	IVS4-2A>G	p.Arg42Pro	The cDNA nomenclature was not understandable and the reference amino acid at position 42 of is 'Glu', not 'Arg'.

3.1.2 Unpublished data overview

In addition to the previously published mutations, a further 37 unpublished unique variants (summarized in **Table 2**) found in *FZD4*, *NDP*, *LRP5*, *KIF11*, *TSPAN12* and *ATOH7* in FEVR patients were included in the mutation dataset. Among all these variants, 17 were novel variants (marked in bold in **Table 2**). Among these 17 novel variants, nine were classified as pathogenic using the ACMG criteria, which include p.E431Dfs*2 and p.V17Afs*116 in *FZD4*; exon3 del in *NDP*; p.R281Sfs*34 in *KIF11*; p.G479Efs*50 and p.E1270Rfs*169 in *LRP5*; exon5_7del and exon7_8del in *TSPAN12* and p.K75* in *ATOH7*. The details of the patient phenotype, allele frequency information, segregation data and *in silico* prediction scores were all calculated and are provided in **Appendix 2** in an Excel file under the name as unpublished FEVR variants.

Six of these variants were found in gnomAD but were not reported in previously in FEVR patients. In this unpublished dataset, 10 protein truncating variants (including nonsense and frameshift variants), 2 splicing variants (c.4488+2T>G and c.2318+1G>C in *LRP5*) and 3 large deletions (exon3 del in *NDP* and exons 5_7 del, exons 7_8 del in *TSPAN12*) were all assessed as pathogenic using ACMG criteria. However, the majority of missense variants (21/22) were annotated as VUS in accordance with ACMG guideline.

Table 2. Summary of unpublished variants in FEVR patients. ^{*(a)}

Gene	cDNA change	Protein change	ACMG	Previous reported information
<i>FZD4</i>	c.1513C>T	p.Q505*	Pathogenic	Dailey et al. 2015, Yang et al. 2012, Drenser et al. 2009, Toomes et al. 2004
<i>FZD4</i>	c.1463G>C	p.G488V	VUS	Novel
<i>FZD4</i>	c.1293_1296delAAGA	p.E431Dfs*2	Pathogenic	Novel
<i>FZD4</i>	c.1282_1285delGACA	p.D428Sfs*2	Pathogenic	Wang et al. 2019, Li et al. 2018, Rao et al. 2017, Tang et al. 2016, Musada et al. 2016, Salvo et al. 2015, Yang et al. 2012, Robitaille et al. 2011, Nikopoulos et al. 2009, Seo et al. 2015
<i>FZD4</i>	c.476T>G	p.M159G	VUS	Novel
<i>FZD4</i>	c.355G>A	p.G119S	VUS	Found in gnomAD, first report for FEVR
<i>FZD4</i>	c.313A>G	p.M105V	VUS	Tang et al. 2016, Chen et al. 2019, Iwata et al. 2019, Wang et al. 2019, Li et al. 2018, Huang et al. 2017, Musada et al. 2016, Seo et al. 2015, Salvo et al. 2015, Yang et al. 2012, Robitaille et al. 2011, Jia et al. 2010, Qin et al. 2005, Kondo et al. 2003
<i>FZD4</i>	c.169G>C	p.G57R	VUS	Novel
<i>FZD4</i>	c.40_49delCCCCGGGGGC G	p.P14Sfs*44	Pathogenic	Li et al. 2018, Khan et al. 2017
<i>FZD4</i>	c.49insCCCCGGGGGCG	p.V17Afs*116	Pathogenic	Novel
<i>NDP</i>	c.125A>G	p.H42R	VUS	Shastry, 1997, Wu et al. 2007, Drenser et al. 2006
<i>NDP</i>	exon3 del	large deletion	Pathogenic	Novel
<i>KIF11</i>	c.839_840delAG	p.R281Sfs*34	Pathogenic	Novel
<i>KIF11</i>	c.907G>A	p.V303I	VUS	Novel
<i>LRP5</i>	c.118C>T	p.R40C	VUS	Found in gnomAD, but first report in FEVR
<i>LRP5</i>	c.487G>A	p.G163R	VUS	Found in gnomAD, but first report in FEVR
<i>LRP5</i>	c.1279G>A	p.A427T	VUS	Novel
<i>LRP5</i>	c.1378G>A	p.E460K	VUS	Li et al. 2018
<i>LRP5</i>	c.1436delG	p.G479Efs*50	Pathogenic	Novel

<i>LRP5</i>	c.1741A>G	p.I581V	VUS	Found in gnomAD, but first report in FEVR
<i>LRP5</i>	c.1828G>C	p.G610R	VUS	Keser et al 2017, Qin et al. 2005
<i>LRP5</i>	c.2116G>A	p.G706R	VUS	Found in gnomAD, but first report in FEVR
<i>LRP5</i>	c.2254C>T	p.R752W	VUS	Hull et al. 2019, Jiao et al. 2004 (c.2302C>G in the article, wrong nomenclature) and Ai et al 2005
<i>LRP5</i>	c.2318+1G>C	splicing/intronic	Pathogenic	Wang et al. 2019
<i>LRP5</i>	c.2918T>C	p.L973P	VUS	Novel
<i>LRP5</i>	c.3107G>A	p.R1036Q	VUS	Stiegel et al. 2012, Stiegel et al. 2013
<i>LRP5</i>	c.3245A>G	p.Y1082C	VUS	Found in gnomAD, but first report in FEVR
<i>LRP5</i>	c.3379T>C	p.F1127L	VUS	Chen et al. 2019, Wang et.al 2019
<i>LRP5</i>	c.3561delG	p.K1187Nfs*65	Pathogenic	Novel
<i>LRP5</i>	c.3683G>A	p.C1228Y	VUS	Novel
<i>LRP5</i>	c.3804delA	p.E1270Rfs*169	Pathogenic	Novel
<i>LRP5</i>	c.3914G>A	p.C1305Y	VUS	Hull et al. 2019
<i>LRP5</i>	c.4099G>A	p.E1367K	Likely pathogenic	Ai et al. 2005, Jiao et al. 2004
<i>LRP5</i>	c.4488+2T>G	splicing/intronic	Pathogenic	Chen et al. 2019, Wang et.al 2019, Tang et al. 2017, Toomes et al. 2004
<i>TSPAN12</i>	exon5-7del	large deletion	Pathogenic	Novel
<i>TSPAN12</i>	exon7-8del	large deletion	Pathogenic	Novel
<i>ATOH7</i>	c.223A>T	p.K75*	Pathogenic	Novel

^(a). Novel variants are in bold typeface.

3.1.3 All variants overview

The complete mutation dataset consisted of all the published and unpublished mutations minus the ones removed due to errors. In total, 537 unique variants were identified in 902 FEVR cases in 11 genes (*FZD4*, *NDP*, *CTNNB1*, *KIF11*, *LRP5*, *TSPAN12*, *ATOH7*, *ZNF408*, *RCBTB1*, *ILK* and *JAG1*) and three loci (*11p13-p12* (EVR3), *22q11.2* and *7q22*). The number of unique variants and cases with mutations in each of the different FEVR genes were summarized in **Table 3**. From this data, it can be seen that 339 FEVR cases were associated with variants in *FZD4*, accounting for 37.58% of the total 902 FEVR cohort. The next most commonly mutated genes in FEVR were *LRP5* (26.71%), *TSPAN12* (12.97%) and *NDP* (11.31%). However, the highest number of unique variants were found in *LRP5*, which contributed 33.33% of the total number of unique variants identified in FEVR patients, followed by *FZD4* (25.14%), *TSPAN12* (15.08%) and *NDP* (10.80%) respectively. For *CTNNB1*, *KIF11*,

ATOH7, and *ZNF408* genes, a relatively small number of cases harboured mutations in one of these genes (92 in total) and a total of only 74 unique variants were reported. Furthermore, only 11 FEVR cases with 10 unique variants were reported in the new candidate FEVR genes: *RCBTB1*, *ILK* and *JAG1*.

Table 3. Summary of the number of cases and unique variants in FEVR

	Cases	Unique variants
<i>FZD4</i>	339 (37.58%)	135 (25.14%)
<i>NDP</i>	102 (11.31%)	58 (10.80%)
<i>CTNNB1</i>	12 (1.33%)	9 (1.68%)
<i>KIF11</i>	44 (4.88%)	40 (7.45%)
<i>LRP5</i>	241 (26.72%)	179 (33.33%)
<i>TSPAN12</i>	117 (12.97%)	81 (15.08%)
<i>ATOH7</i>	14 (1.55%)	7 (1.30%)
<i>ZNF408</i>	22 (2.44%)	18 (3.35%)
<i>RCBTB1</i>	2 (0.22%)	1 (0.19%)
<i>ILK</i>	3 (0.33%)	3 (0.56%)
<i>JAG1</i>	3 (0.33%)	3 (0.56%)
<i>Loci</i>	3 (0.33%)	3 (0.56%)
Total	902	537

The distribution of the different predicted protein consequences observed in the *FZD4*, *LRP5*, *NDP* and *TSPAN12* were investigated (**Figure 3**). In all of these four proteins, missense variants accounted for the highest percentage followed by truncating variants. For example, in *LRP5*, 80% of variants were predicted to cause missense alterations, 12% were predicted to result in premature termination codons, 7% were splicing variants, and 1% caused in-frame deletions.

Similarly, when the total dataset was analysed, the majority of variants were found to result in missense variants (61%), followed by premature termination codons (31%). The splicing variants, stop loss or start loss, and in-frame variants were also reported in these genes, but with a lower percentage. In addition, raw data and distribution of variants types in other genes were also provided in **Table 4**.

Table 4. Distribution of variants types in FEVR

<i>FZD4</i>	<i>NDP</i>	<i>CTNNB1</i>	<i>KIF11</i>	<i>LRP5</i>	<i>TSPAN12</i>	<i>ATOH7</i>	<i>ZNF408</i>	Total

Missense	74	42	1	7	143	38	2	15	322
Protein truncating^a	56	15	7	27	21	29	4	3	162
In frame indels	4	0	0	1	2	0	1	0	8
Stop loss & Start loss	1	0	0	0	0	2	0	0	3
5'UTR & 3'UTR	0	1	0	0	0	0	0	0	1
Splicing	0	0	1	5	13	12	0	0	31
Total	135	58	9	40	179	81	7	18	527

a: including nonsense, frameshift and large deletions.

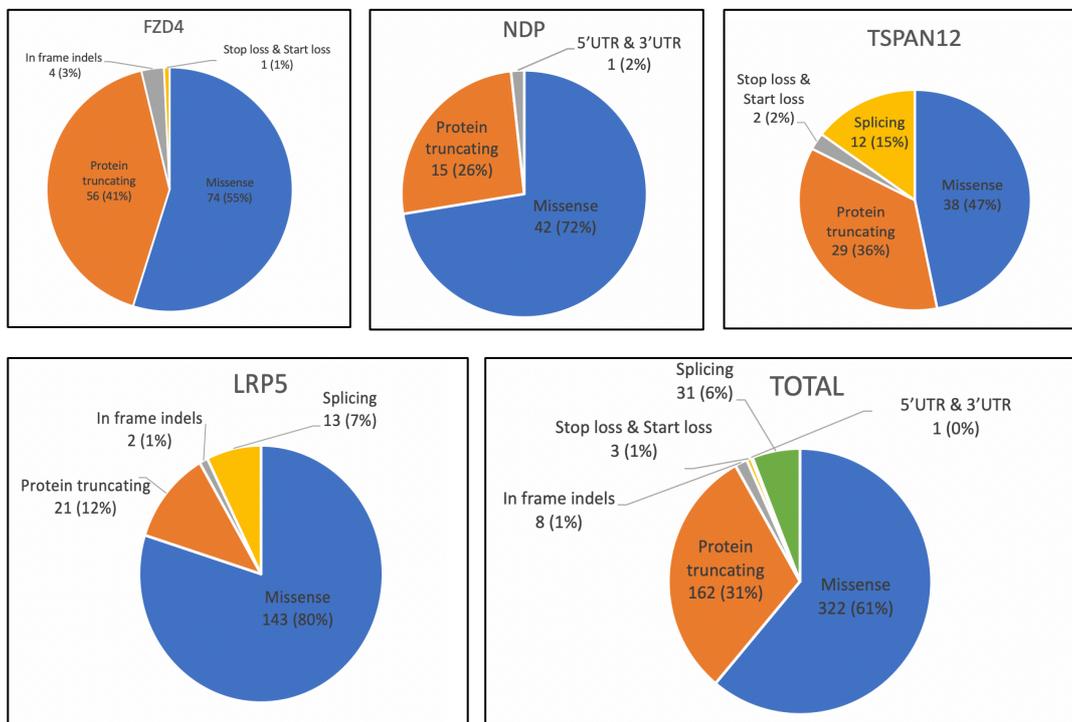


Figure 3. Distribution of unique variants types in FEVR patients. Protein truncating variants showed in the Pie Chart include nonsense, frameshift and large deletions variants.

3.1.4 Pathogenicity classification

As detailed in **Table 3**, the majority of variants and cases were reported in eight genes (*FZD4*, *NDP*, *CTNNB1*, *KIF11*, *LRP5*, *TSPAN12*, *ATOH7* and *ZNF408*). Analysis was therefore undertaken to classify the pathogenicity of unique variants found in these eight genes according to ACMG guidelines. In total, analysis of 527 variants

classified 183 (34.7%) as pathogenic and 42 (7.97%) as likely pathogenic (**Table 5**). However, VUS accounted for the majority of variants with 55.41% (292 in 527) receiving this score after reclassifying by the ACMG guideline. As expected, the missense variants reported in FEVR were problematic to classify, with 86.34% of missense variants lacking sufficient evidence to classify as (likely) pathogenic or (likely) benign and were categorised as VUS in the study (**Table 6**). There were also several variants reclassified as benign (3) or likely benign (6) in FEVR (details were provided in Mutation Spectra section **3.1.5**).

Table 5. Pathogenicity classification of variants in FEVR

ACMG	<i>FZD4</i>	<i>NDP</i>	<i>CTNNB1</i>	<i>KIF11</i>	<i>LRP5</i>	<i>TSPAN12</i>	<i>ATOH7</i>	<i>ZNF408</i>	Total
Pathogenic	56	14	6	31	29	43	4	0	183
Likely pathogenic	16	9	2	3	8	2	0	2	42
VUS	59	35	1	6	137	36	3	15	292
Likely benign	1	0	0	0	4	0	0	1	6
Benign	2	0	0	0	1	0	0	0	3
Total	135	58	9	40	179	81	7	18	527

Table 6. Percentage of VUS missense variants in FEVR

	<i>FZD4</i>	<i>NDP</i>	<i>CTNNB1</i>	<i>KIF11</i>	<i>LRP5</i>	<i>TSPAN12</i>	<i>ATOH7</i>	<i>ZNF408</i>	Total
VUS missense	57	34	1	5	131	36	2	12	278
Total missense	74	42	1	7	143	38 ^(a)	2	15	322
Percentage	77.03%	80.95%	100.00%	71.43%	91.61%	94.74%	100.00%	80.00%	86.34%

*^(a). include 2 missense start loss variants.

Apart from these eight genes, mutations in *RCBTB1*, *ILK* and *JAG1* were reported as potential candidate genes in FEVR and three cases of FEVR were reported with cytogenetic abnormalities on chromosome *22q11.2* and *7q22*, but no genes have been identified at these loci yet. However, for these genes and loci, only a limited number of cases and variants have been reported (**Table 7**). Only three missense variants were reported in *JAG1* in FEVR, and the pathogenicity evidence for all of them were not sufficient, with all of the variants classified as VUS by the ACMG criteria. Similarly,

three missense variants were reported in *ILK* but only one of these variants was classified as likely pathogenic. Furthermore, two studies found a large microduplication and a large microdeletion in the same locus *22q11.2* in FEVR by Gilmour et al. 2009 and Gandhi et al. 2014 respectively.

Table 7. Summary of reported variants of FEVR in potential genes and loci

Reference	Gene	cDNA change	Protein change	ACMG ^(a)	AF (Total)
Wu et al. 2016	<i>RCBTB1</i>	c.1172+1G>A	splicing/intronic	Pathogenic	NA
	<i>ILK</i>	c.157T>A	p.L53M	VUS	0.0004313
Park et al. 2019	<i>ILK</i>	c.631C>T	p.R211C	VUS	0.0009512
	<i>ILK</i>	c.950G>A	p.R317Q	Likely pathogenic	0.00001193
Zhang et al. 2019	<i>JAG1</i>	c.413C>T	p.A138V	VUS	NA
	<i>JAG1</i>	c.1415G>A	p.R472H	VUS	0.00008749
	<i>JAG1</i>	c.2884A>G	p.T962A	VUS	0.00008354
Gandhi et al. 2014	<i>locus 22q11.2</i>	microduplication of chr22q11.2 (2.21 Mb in size)	large deletion	pathogenic	NA
Gilmour DF et.al 2009	<i>locus 22q11.2</i>	microdeletion of chr22q11.2	large deletion	pathogenic	NA
Miller et al. 2015	<i>locus 7q22</i>	genetic deletion at chr7q22	large deletion	pathogenic	NA

* 'AF' indicates allele frequency

*(a) The ACMG classification of the variants in these candidate genes were made by assuming the genes are causative for FEVR.

3.1.5 Mutation spectra

To look at the types and location of the reported mutations on the protein domain structures, mutation spectra were generated for the genes encoding the Norrin- β -catenin pathway components: *FZD4*, *NDP*, *TSPAN12*, *LRP5* and *CTNNB1*.

In **Figure 4**, the schematic distribution of all of the variants at the amino acid level of *FZD4* is provided. From the mutation spectrum it can be seen that nearly all of the truncating variants were classified as pathogenic in *FZD4*, only one variant p.F82Afs*45 was considered as likely pathogenic due to the lack of sufficient evidence (cosegregation data was not provided and *in silico* prediction result was not available through CADD). Several new insights of previously reported variants in

FZD4 were revealed by their reanalysis. For example, p.G530E was considered as likely benign, p.P33S and p.P168S were considered as benign according to ACMG criteria. The two benign results were due to their relatively high allele frequency in gnomAD (p.P33S: 0.01618 and p.P168S: 0.01775 in total population respectively), whereas the likely benign variant was due to p.G530E was reported to be observed in healthy individuals and the Clinvar database also reported it as a benign variant.

In addition, FZD4 contains a number of the disulphide bonds located between residues 45-106, 53-99, 90-128, 117-158, 121-145, 181-200, 204-282 and 302-377. From the mutation spectrum it can be seen that many of these were mutated in FEVR cases with a total number of 14 missense variants located in the disulphide formation points of *FZD4*: p.C45S, p.C45R, p.C45Y, p.C106S, p.C53S, p.C90R, p.C128R, p.C117R, p.C181Y, p.C181R, p.C204Y, C204R, p.C204F and p.C302Y. Among them, several variants (p.C45R, p.C45Y, p.C181Y and p.C204Y, p.C204R) were classified as likely pathogenic whereas the remaining were considered as VUS due to the lack of the evidence, such as cosegregation data etc.

The number of variants reported in different domains of *FZD4* in FEVR could also be observed from the mutation spectrum. The majority of the variants were located in the CRD domain (amino acids 45-204), with 52 variants reported in total including 36 missense variants (removed one benign variant p.P168S which confirmed by the ACMG). Another hotspot for mutations was the Lys-Thr-X-X-X-Trp (KTXXXW) PDZ-binding motif (499-504), in the cytoplasmic tail of FZD4, which contained six variants including one missense variant (p.K499E).

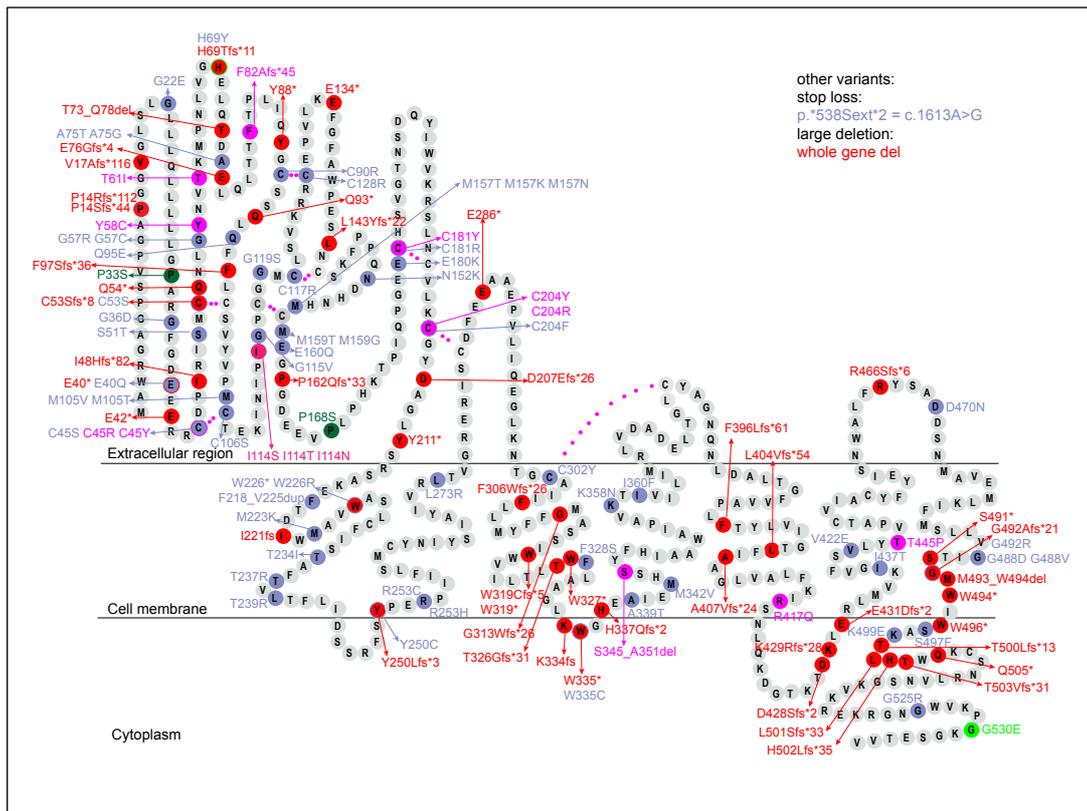


Figure 4. Mutation spectrum of FZD4 in FEVR patients. A schematic representation of the FZD4 protein with pathogenic variants marked in red and likely pathogenic, uncertain significance, likely benign and benign variants coloured in purple, blue, light green and dark green respectively. All of the variants were annotated with HGVS protein level nomenclature, but ‘p.’ symbols were all omitted in order to have a clearer view. “*” was used as describing stop codon in the study. “...” symbol in the figure indicates a disulphide bond.

The mutation spectrum of TSPAN12 in FEVR patients is shown in **Figure 5**. There were 13 pathogenic splice variants and 7 large pathogenic deletions (include a whole gene deletion and exon deletions) reported which are not annotated on the protein structure.

Six missense variants were located in amino acid residues which form disulphide bonds in the large extracellular loop of *TSPAN12* (159-206, 160-181 and 182-189), which were p.C159G, p.C159Y, p.C160Y, p.C181F, p.C189Y and p.C189R. The p.C189Y variant was classified as likely pathogenic in the ACMG.

likely benign and benign variants coloured in purple, blue, light green and dark green respectively.

The schematic distribution of variants in LRP5 in FEVR is shown in **Figure 7**. This figure shows several variants (p.I1362V, p.A97V, p.T173M and c.2092-4C>T) which were reclassified as likely benign by ACMG and a missense mutation (p.Q89R) classified as benign. The reason for the likely benign reclassifications includes a lack of segregation data, observed in healthy individual or *in silico* evidence suggest the variant as neutral. The benign variant p.Q89R is mainly due to the relatively higher allele frequency (0.01833 in total) found in the human population (details can be seen in the **Appendix 2**.)

Several disulphide bonds formation points, 913-926, 1240-1253, 1305-1323, and 1343-1361, in LRP5 are the location of 6 missense variants (p.C913F, p.C1253F, p.C1305R, p.C1305Y, p.C1343G and p.C1361G). The highest reported hotspot region of variants found in LRP5 in FEVR was Beta-propeller 2 domain (341-602), which includes 43 missense variants (48 variants in total). The second hotspot variant region in LRP5 was Beta-propeller 3 (644-903), which reported 28 missense variants in FEVR (31 variants in total). Moreover, in the Beta-propeller 1 (32-288) and Beta-propeller 4 (945-1212) domains, 16 missense (18 in total, removed three likely benign and benign variants classified in the ACMG in the area) and 16 missense (18 in total) variants were reported respectively. In LDL-receptor region (1257-1370), 14 missense variants (18 in total) were found in FEVR (removed one missense likely benign variant (p.I1362V) in this region). Only five variants were reported in the PPPSP motifs.

draw the LRP5 protein structure due to the large size of LRP5 (1615aa). The sequence and domain information were gathered through UniProt database. The exon and domain length of the protein were drawn approximately to scale. The area marked by light blue in the figure represent PPPSP motifs.

The FEVR mutation spectrum of β -catenin is shown in **Figure 8**. Only 9 variants were reported in this gene in FEVR, which mostly occurred in the 12 Armadillo repeats and C-terminal domain region. These are all the reported cases with FEVR but clinical information pertaining to the wider syndromic features seen in the β -catenin intellectual disability syndrome was not investigated in all of these cases.

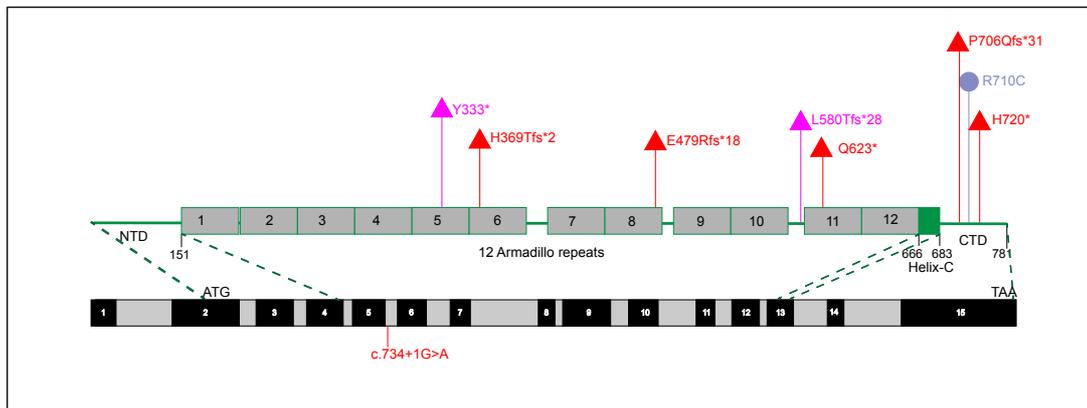


Figure 8. Mutation spectrum of the β -catenin in FEVR patients. Pathogenic variants marked in red and likely pathogenic, uncertain significance, likely benign and benign variants coloured in purple, blue, light green and dark green respectively. Due to β -catenin having a relatively long sequence length (781aa), distribution of the variants in the protein structure overview were given. Sequence of the domain information were gathered through the combination of the mostly frequently used in the peer-reviewed articles (Mo et al., 2009, Gottardi and Peifer, 2008, Xing et al., 2008, Panagiotou et al., 2017).

3.1.6 Frequently reported variants

The majority of mutations in FEVR patients are family specific. However, a few variants which were reported more than once were observed. The most frequently reported variant in *FZD4* was c.313A>G (p.M105V), which was reported in 32 FEVR patients with different ethnicities including Chinese, Japanese, Finnish and Caucasian. This was followed by c.1282_1285del (p.D428Sfs*2) and c.205C>T (p.H69Y), which reported in 24 and 16 FEVR patients respectively. The available ethnicity data

reported in p.D428Sfs*2 including Chinese, Caucasian and Indian; For p.H69Y, the reported ethnicity including Chinese and Japanese.

Similarly, in *NDP* the most frequently reported variants were c.362G>A (p.R121Q), c.361C>T (p.R121W) and c.125A>G (p.H42R), which have been reported in 9, 5 and 5 FEVR patients respectively. The available ethnicity data of p.R121Q mutation were reported in the Chinese, Italian, Polish, Indian, Mexican, Dutch and Spanish FEVR patients. However, for the other two variants, the ethnicity data were mostly unavailable or only reported in patients from the same ethnicity.

In *LRP5*, the c.4488+2T>G variant was reported in 6 FEVR patients in Chinese and Caucasian ethnicity, and c.4099G>A (p.E1367K), c.3361A>G (p.N1121D), c.1264G>A (p.A422T), and c.3379T>C (p.F1127L), which reported in 5, 5, 4 and 4 FEVR patients respectively and either with unavailable ethnicity data or were from the same ethnicity group.

The most frequently reported variants in *TSPAN12* were c.1A>G (p.M1V, also described as p.0?), c.149+3A>G (p.L23Gfs*88), c.469-1G>A and c.285+1G>A (p.R50Dfs*12), which were reported in 6, 5, 3 and 3 FEVR patients respectively. The ethnicity of c.285+1G>A were reported from Chinese and Nigeria patients, and c.149+3A>G were reported from Caucasian and Chinese FEVR patients. However, for other three variants, the ethnicity data were unavailable or only reported from the same ethnicity group. The raw data were available at *Appendix 2*.

3.1.7 Phenotype-Genotype correlation

In the total number of 902 FEVR cases, only 501 cases of them were presented with available phenotype data (228 in *FZD4*, 52 in *NDP*, 12 in *CTNNA1*, 29 in *KIF11*, 90 in *LRP5*, 61 in *TSPAN12*, 10 in *ATOH7*, 14 in *ZNF408*, 2 in *RCBTB1* and 3 in *JAG1* respectively). However, no new phenotype-genotype correlation was found in the study other than the association of reduced bone mineral density in cases with *LRP5* mutations, microcephaly in *KIF11* cases and syndromic features in some *CTNNA1* cases.

3.1.8 LOVD

All of the variants in this study have been uploaded to the LOVD database along with their ACMG classifications. This data will be accessible online after the manuscript containing the data from this study is published. In the meantime, the raw data of all the variants and ACMG classification details are provided in the accompanying excel files. Unpublished data information was also provided in a separate file in *Appendix 2* under the name of Unpublished FEVR variants.

3.2 Protein modelling

Protein modelling was performed to try and determine if this methodology could be used as an alternative method to accurately classify a mutation as pathogenic or benign.

3.2.1 Wildtype 3D protein structures analysis

The 3D protein structures of FZD4, NDP/Norrin, TSPAN12, β -catenin and LRP5 proteins were predicted through I-TASSER and Phyre2 software respectively. 100% coverage of FZD4, NDP/Norrin, TSPAN12 and β -catenin was achieved in I-TASSER protein modelling, whereas in Phyre2, relatively lower coverage (FZD4: 84.92%, NDP/Norrin: 78.20%, TSPAN12: 72.13% and β -catenin: 70.42%) were presented. In addition, for LRP5, the 92.88% coverage obtained through I-TASSER was not 100% because the software limits the protein length to 1500 so only residues pertaining to amino acids 1-1500 in LRP5 were predicted. Only 36.35% coverage of LRP5 was achieved by Phyre2.

The overall protein modelling quality was evaluated through Ramachandran plots and ERRAT. The ERRAT score for FZD4, NDP/Norrin, TSPAN12, and β -catenin protein structures from I-TASSER were all higher than the structures predicted by Phyre2. However, the Ramachandran plot data favoured the structures predicted by Phyre2 higher than compared to the structures given by I-TASSER. The details of the comparison results were provided in the *Table 8*.

Table 8. Wildtype Structure comparison between ITASSER and Phyre2

Protein	Software	COVERAGE	ERRAT	Ramachandran (number of residues)
FZD4	ITASSER	100%	89.981	In favoured region: 419 (78.3%) In allowed region: 65 (12.1%) In outlier region: 51 (9.5%)
	Phyre2	84.92% (456/537)	32.540	In favoured region: 421 (92.1%) In allowed region: 24 (5.3%) In outlier region: 12 (2.6%)
NDP	ITASSER	100%	73.729	In favoured region: 81 (61.8%) In allowed region: 34 (26.0%) In outlier region: 16 (12.2%)
	Phyre2	78.20% (104/133)	72.973	In favoured region: 98 (96.1%) In allowed region: 4 (3.9%) In outlier region: 0 (0.0%)
TSPAN12	ITASSER	100%	80.135	In favoured region: 232 (76.6%) In allowed region: 43 (14.2%) In outlier region: 28 (9.2%)
	Phyre2	72.13% (220/305)	65.537	In favoured region: 215 (91.9%) In allowed region: 13 (5.6%) In outlier region: 6 (2.6%)
CTNNB1	ITASSER	100%	95.520	In favoured region: 678 (87.0%) In allowed region: 74 (9.5%) In outlier region: 27 (3.5%)
	Phyre2	70.42% (550/781)	94.891	In favoured region: 495 (89.4%) In allowed region: 55 (9.9%) In outlier region: 4 (0.7%)
LRP5	ITASSER	92.88% (1500/1615)	NA ^(a)	In favoured region: 1073 (71.6%) In allowed region: 271 (18.1%) In outlier region: 154 (10.3%)
	Phyre2	36.35% (587/1615)	66.944	In favoured region: 533 (87.5%) In allowed region: 59 (9.7%) In outlier region: 17 (2.8%)

^(a): NA indicates not available

The I-TASSER modelling results of wildtype protein structures of all five proteins were chosen in this study due to the higher coverage and higher ERRAT values. The wildtype structures of FZD4, TSPAN12, NDP/Norrin, β -catenin, and LRP5 generated C-scores (confidence score for evaluating the models predicted through I-TASSER) of -1.16, -1.35, -1.24, -2.67 and 0.39 respectively (**Figure 9**). The 3D protein structure in these proteins were marked by different colours according to different domain sequences.

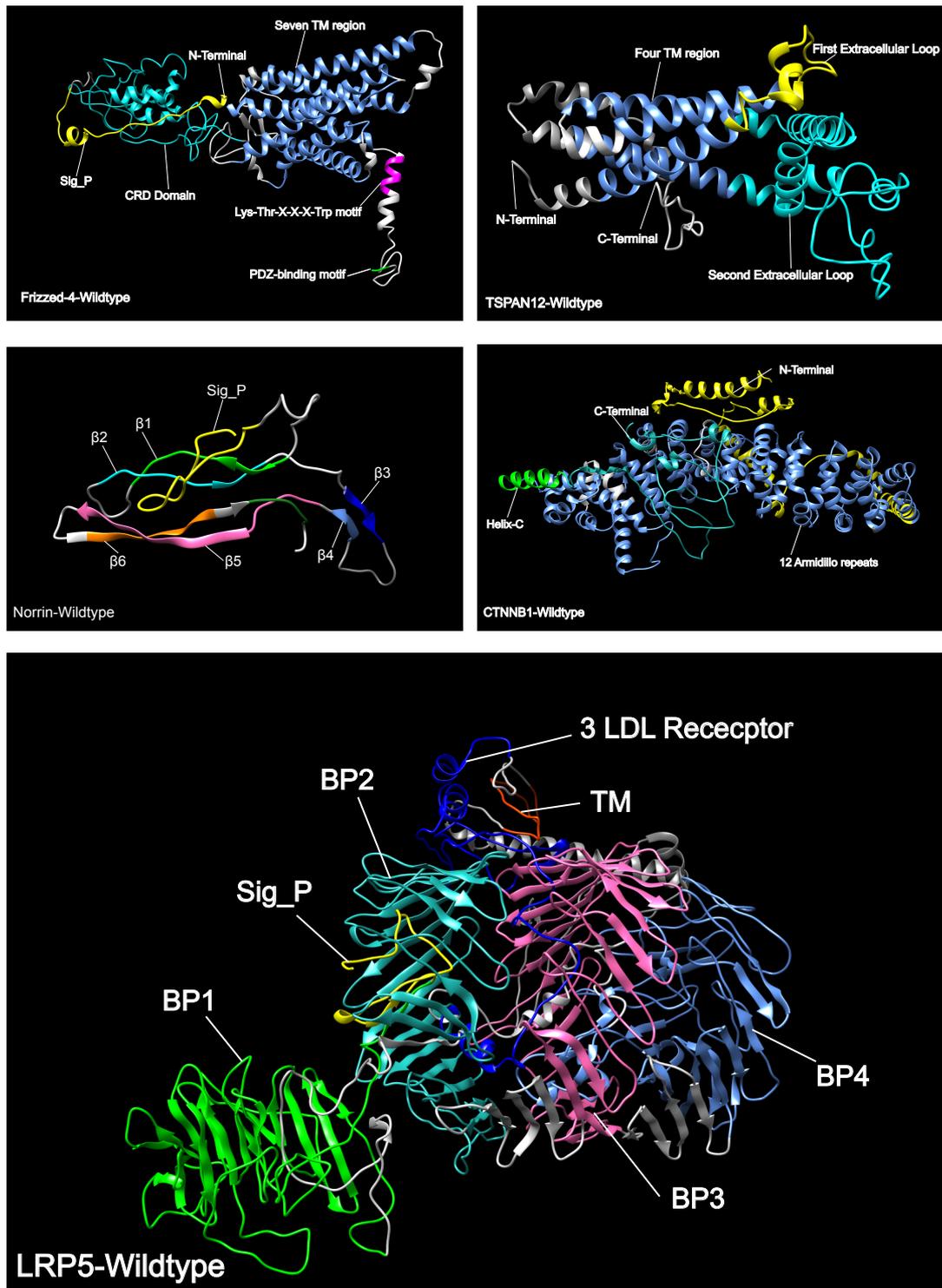


Figure 9. Predicted 3D wildtype protein structures of FZD4, NDP/Norrin, TSPAN12, LRP5 and β -catenin (CTNNB1) by I-TASSER. Domain sequences were obtained from UniProt database. All the structures were predicted through I-TASSER, different domains and motifs were annotated by different colours in UCSF Chimera. ‘TM’ symbol represents transmembrane domain and ‘Sig_P’ represents

signal peptide region. ‘BP’ in the LRP5 structure is the abbreviation of Beta-propeller domain.

3.2.2 Comparison of wildtype and mutant protein structures

After superimposing of the mutant protein structures with the corresponding wildtype structure of FZD4 in the UCSF chimera, RMSD values were calculated (**Table 9**). From **Table 9**, it can be seen that the overall RMSD value ranges from 0.419 Å (p.C45S) to 0.762 Å (p.C204R) between the missense mutant models and wildtype FZD4 model. In addition, except for the variants p.C45S (0.419 Å), p.G119S (0.462 Å) and p.G530E (0.499 Å), the overall RMSD value were all higher than 0.5 Angstrom (Å), and indicated the deviation occurred between the mutant protein structures and the wildtype structure.

The value of MUpro was also calculated (**Table 9**). It is noteworthy that the majority of missense variants in FZD4 decreased the protein stability ($\Delta\Delta G < 0$), and only two missense variants (p.S497F:0.035 and p.E40Q:0.039) increased the protein stability ($\Delta\Delta G > 0$). In addition, the overall model quality from the ERRAT score was acceptable (quality score higher than 70) with the exception of p.C117R (68.242) and p.C90R (68.242). As well as, full raw data of all the deviations of the RMSD backbone value in the amino acid level were provided in the **Appendix 1. Table A1**. The data showed that the deviated region in all of the variants affected the Signal-peptide region, CRD domain region in FZD4, and they all shared the common deviated region from residues 508-537. Interestingly, it is noteworthy that there was no significant difference found in the comparisons between the variants that have already been reassessed as likely benign or benign and those likely pathogenic variants (like the comparative data of p.P33S (benign) and p.C45Y (likely pathogenic)).

Table 9. Model comparison of missense mutant structure and wildtype FZD4 structure of FEVR patients

Mutant structures ^{*(a)}	Overall RMSD value ^{*(b)}	ERRAT ^{*(c)}	MUpro score ^{*(d)}
-----------------------------------	------------------------------------	-----------------------	-----------------------------

G22E	0.643	71.456	-0.815
P33S	0.696	75.992	-1.028
E40Q	0.578	78.639	0.039
C45S	0.419	75.425	-1.822
C45R	0.633	73.535	-1.399
C45Y	0.602	76.371	-1.336
S51T	0.571	73.346	-0.791
C53S	0.564	74.102	-1.308
G57R	0.582	76.182	-1.359
Y58C	0.621	76.281	-0.927
T61I	0.644	76.371	-0.463
H69Y	0.625	77.127	-0.275
A75T	0.704	78.828	-0.720
A75G	0.666	75.992	-0.845
C90R	0.614	68.242	-0.965
Q95E	0.574	77.316	-0.607
M105V	0.681	72.968	-1.017
M105T	0.740	71.834	-1.550
C106S	0.645	73.724	-1.319
I114N	0.603	79.206	-1.117
I114T	0.691	75.425	-1.266
I114S	0.632	70.321	-0.943
G115V	0.578	77.505	-0.746
C117R	0.630	68.242	-0.523
G119S	0.462	75.047	-0.559
C128R	0.646	73.157	-0.967
N152K	0.565	73.724	-1.091
M157V	0.736	79.017	-1.127
M157K	0.547	76.749	-1.774
M157T	0.564	78.072	-1.530
M159T	0.649	74.669	-1.071
M159G	0.694	74.291	-1.683
E160Q	0.509	74.858	-0.554
P168S	0.668	76.560	-0.847
E180K	0.525	75.992	-0.698
C181R	0.573	76.560	-1.150
C181Y	0.636	78.072	-0.741
C204R	0.762	74.291	-0.950
C204Y	0.582	73.724	-0.603
C204F	0.704	72.212	-0.427
M223K	0.689	75.047	-1.383
W226R	0.634	75.236	-0.108
T234I	0.563	77.505	-0.154
T237R	0.680	75.992	-0.635

L239P	0.653	78.639	-1.345
Y250C	0.642	73.913	-0.895
R253C	0.514	77.316	-0.738
R253H	0.667	77.883	-1.173
L273R	0.534	73.724	-2.157
C302Y	0.507	72.023	-0.864
F328S	0.686	77.883	-1.474
W335C	0.676	71.456	-1.098
A339T	0.689	74.102	-0.480
M342V	0.689	76.560	-0.986
K358N	0.700	76.938	-0.377
I360F	0.548	75.992	-0.640
R417Q	0.592	77.127	-0.689
I437T	0.666	73.535	-2.407
V442E	0.704	75.236	-1.276
T445P	0.635	71.456	-1.010
D470N	0.652	76.182	-1.590
G488D	0.616	77.316	-0.518
G488V	0.627	72.590	-0.310
G492R	0.706	75.614	-0.932
S497F	0.569	76.938	0.035
K499E	0.613	73.724	-0.829
G525R	0.587	77.316	-1.107
G530E	0.499	70.132	-0.433

^(a). Missense mutant variants were annotated with the latest protein change HGVS nomenclature with a omitted symbol “p.” for a clearer view.

^(b). The overall RMSD value indicated the deviation of mutant structure with its wildtype FZD4 structure

^(c). The overall model quality score of the missense mutant structures predicted through ERRAT software.

^(d). The $\Delta\Delta G$ predicted through MUpro software, which indicated the evaluation of the missense variants on the protein stability.

The quality of the *FZD4* missense mutant structure models were also assessed through Psi/Phi Ramachandran plot, and the residues favoured region were all with the acceptable model quality as over 85.0%. For example, in the representative figures provided in **Figure 10**, the Ramachandran plot of FZD4 p.G22E mutant structure revealed that, among 537 residues analysed, 462 (86.4%) were in the favoured region and a further 54 residues (10.1%) were in the allowed region whereas only 19 (3.6%) of residues were in the outlier region. The outlier residues included in the C-terminus region were only pertaining to residues 536, 531 and 523 in the p.G22E model.

However, similar results also showed in the reclassified benign p.P33S mutant model, 465 (86.9%) residues were in the favoured region, and 49 (9.2%) and 21 (3.9%) were in allowed and outlier region respectively. From the superimposing figures, the most deviated region between mutant protein structures and wildtype structure could also be seen (see all figures in *Appendix 1 Figure A1.*) and representative figures of p.G22E and p.P33S (*Figure 10*). The superimposing figures between mutant protein structures (marked yellow) with wildtype FZD4 structure (coloured in blue) of p.G22E and p.P33S shared the same deviation region in the C-terminus. However, in the representative ERRAT mutant p.G22E and p.P33S analysis models, the overall quality of p.G22E model was 71.456, which the majority of error occurred in the signal-peptide region and c-terminus (501-522) region (confidence level chosen by 99% in the study). Similarly, ERRAT analysis result also presented in the reclassified benign variant (p.P33S, overall quality value was 75.992).

The missense mutant structures of NDP/Norrin and TSPAN12 presented low ERRAT value (average ~ 40 to ~60 representatively), like the representative figures shown in *Figure 11*. and *Figure 12*. Thus, no further analysis was performed for these proteins.

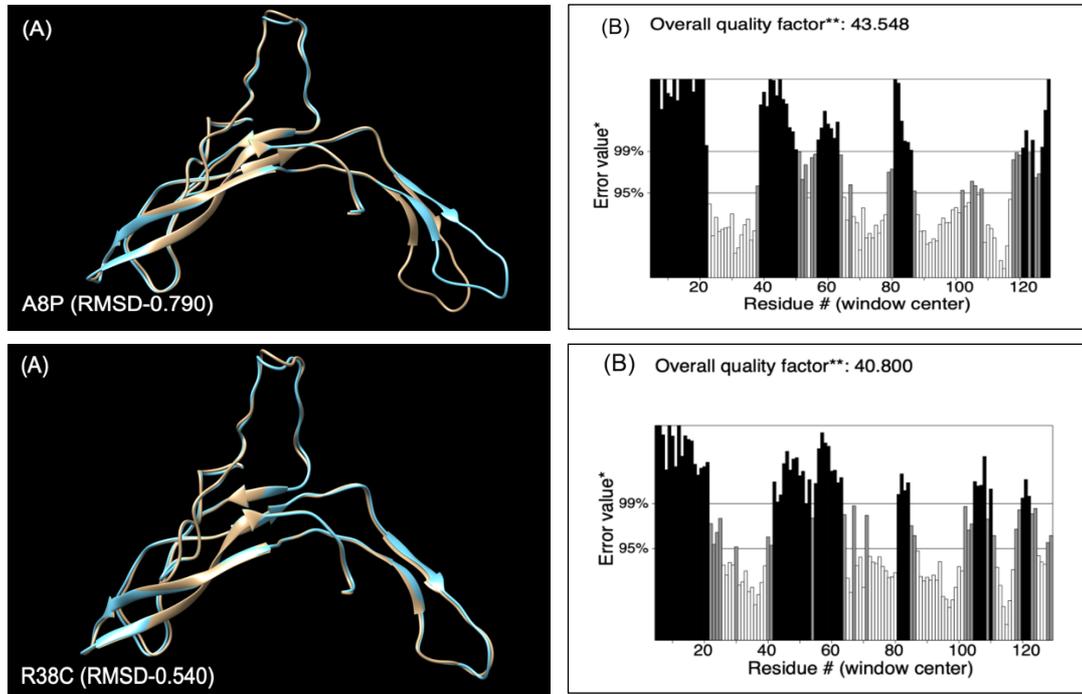


Figure 11. Representative examples of structure deviation comparison between wildtype and missense mutant NDP/Norrin protein 3D structures. (A). The mutant models (p.A8P and p.R38C) superimposing with wildtype model in Chimera. Wildtype and mutant structure coloured by blue and yellow respectively. (B). The mutant models ERRAT score.

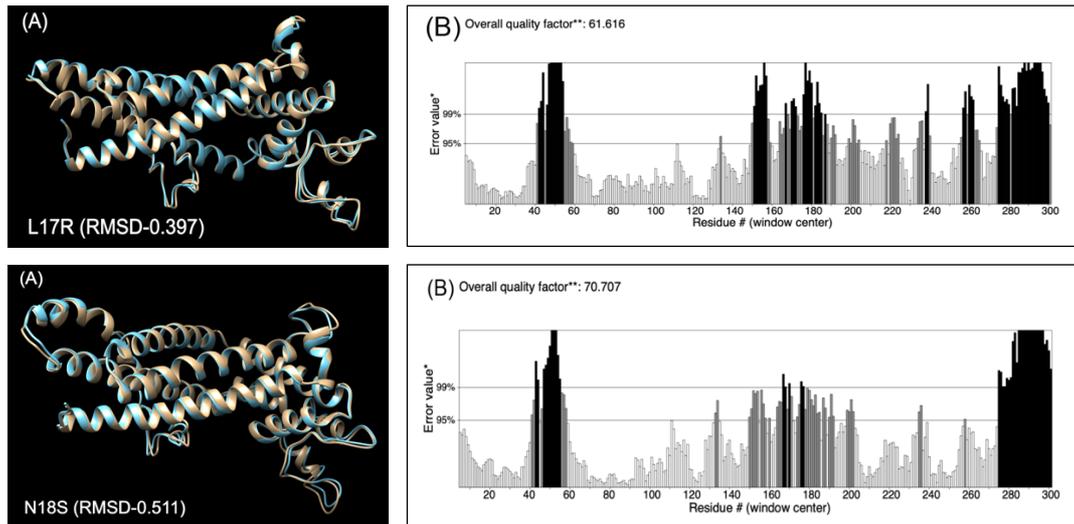


Figure 12. Representative examples of structure deviations comparison between wildtype and missense mutant TSPAN12 protein 3D structures. (A). The mutant models (p.L17R and p.N18S) superimposing with wildtype model in Chimera. Wildtype and mutant structure marked by blue and yellow respectively. **(B)** The mutant models ERRAT score.

4. Discussion

In this project an FEVR mutation database has been created using the LOVD platform. The database contains the variants identified in eight genes known to cause FEVR when mutated, but also includes the data from three candidate FEVR genes which are awaiting confirmation from second reports.

While the majority of the information contained in the database was already available in peer-reviewed journals, it is a time-consuming process to review the 100+ papers. Furthermore, the data presented in the articles was often annotated using different nomenclature or non-standard descriptions which made the data hard to interpret in a timely manner. Similarly, a number of mutation repositories are available such as ClinVar (Landrum et al., 2020) or HGMD (Human Gene Mutation Database) (Stenson et al., 2020) and there were a few FEVR mutations already uploaded onto LOVD. However, HGMD is not freely available to the research community and the other available datasets are often incomplete or are only limited to the variant information and miss off valuable information such as the associated phenotypic

information or cosegregation data. Therefore, the systematic collection, annotation and *in silico* re-assessment of the variants using the latest ACMG guidelines and frequency data is of great benefit to the scientific and diagnostic community.

4.1 Pathogenicity classification

The reclassification of the variants reported in FEVR has highlighted a number of errors in the published literature. A number of these are likely to be typographical errors or poor reporting methods and were removed from the study (**Table 1**). However, some of the variants were reported as pathogenic or potentially pathogenic but the historical context of the paper meant that the evidence supporting these classifications was often weak. For example, before the availability of population databases like gnomAD, researchers would screen control cohorts using Sanger sequencing to exclude the variant as a polymorphism. As the definition of a polymorphism is a variant with an allele frequency greater than 1%, this meant that the researchers often only excluded the variant in 175 individuals from the same ethnic background (Collins and Schwartz, 2002). This resulted in a number of variants which are now known to be too common to be mutations being in the literature as causative. Examples of this include the variants p.P33S and p.P168S in *FZD4* which have allele frequencies of 0.01618 and 0.01775 respectively in gnomAD. In total, nine variants were re-classified as benign or likely benign in this study (p.G530S, p.P33S and p.P168S in *FZD4*; p.Q89R, p.A97V, p.T173M, p.I1362V and c.2092-4C>T in *LRP5*; and p.G290S in *ZNF408*). As well as, for the variants interpretation in *ZNF408* gene, the evidence for the loss-of-function variants (like nonsense variants) in this gene were not applied with the PVS1 criterion (very strong criterion for null variant in a gene which loss of function could cause of the disease) in the ACMG due to the probability of loss of function intolerance (pLI) is zero in gnomAD v2.1.1, which means the gene or transcript is more likely to tolerate protein truncating variation. As well, among all 18 variants in *ZNF408* in FEVR patients, 83.33% of variants (15/18) were classified as VUS (**Table 5**), only two *de novo* missense variants which absent from controls were able to classify as likely

pathogenic. This result suggests that the evidence of *ZNF408* in FEVR patients may not that strong. Thus, cleaning up the data in this way is important as there are many examples of researchers reporting variants as pathogenic just because they have previously been published as such. A recent study has reassessed the frequency of a number of autosomal dominant inherited retinal disease genes and variants and shown that as many as 10% of reported mutations have allele frequencies greater than 1% and some reported disease gene may not actually be real (Hanany and Sharon, 2019).

The ACMG classification system was used to classify all of the variants in the database (Richards et al., 2015a). The ACMG system is an attempt to add some standardisation to the classification of a variant as pathogenic or benign. In this study, only 183 variants among 527 unique variants in *FZD4*, *NDP*, *CTNNB1*, *KIF11*, *LRP5*, *TSPAN12*, *ATOH7* and *ZNF408* were reclassified as pathogenic by ACMG (**Table 4**), and 55.41% (292/527) variants were reclassified as uncertain significance. These results all indicated the difficulty and limit of classifying variants pathogenicity in the ACMG. This is a particular issue for missense variants as the majority of these (86.34%) in the study were confirmed as VUS. While the data will still facilitate genetic counselling, and potentially pre-implantation technologies and prenatal diagnostics (Hoffman-Andrews, 2017), VUS are often a limiting factor for new upcoming treatments (as mentioned in section 1.2). For rare diseases like FEVR, getting enough patients to test therapies in clinical trials is often an issue as patient numbers are often small. However, many of the clinical trials will not risk including cases with a tentative mutation or VUS as if the variant is not the cause of the disease there is the potential that the whole clinical trial will fail. VUS therefore have the potential to slow down the development of therapies but even those which make it through to clinical services may not be available to patients who do not have a very convincing molecular diagnosis.

FEVR is a particularly problematic disease to classify using the ACMG guidelines. Non-penetrance is a major issue with FEVR and there are many reports of asymptomatic mutation carriers (Toomes and Downey, 2011). It is therefore expected

that a number of mutations will be present in population databases such as gnomAD. The high proportion of asymptomatic carriers also means that many cases are recruited as “sporadic” cases with no family history so this limits the opportunity for cosegregation studies which is a major feature of ACMG classification (Jarvik and Browning, 2016). These features are likely to contribute to the high number of VUS observed in this study.

Clearly the availability of a functional assay is very important to help classify VUS in FEVR. As the majority of FEVR mutations affect the Norrin- β -catenin signalling pathway, many research teams use the TOPflash reporter assay to investigate if variants alter the transcriptional response of β -catenin (Han et al., 2020, Xu et al., 2004, Panagiotou et al., 2017). However, in this study results using this assay were not included as a functional assay due to questions about the reliability of this assay for diagnostic purposes. For example, some studies which have used this assay on FEVR mutations and SNPs and have shown that it can't differentiate between them (Qin et al., 2008). Similarly, *LRP5* polymorphisms are known to contribute to variations in bone mineral density in the general population and many of these polymorphisms, including some present in 20% of the general population, also shown statistically significant reductions in the TOPflash assay at levels similar to FEVR mutations (Urano et al., 2009).

Another functional assay which is used to assess the functional nature of variants in FEVR are zebrafish models (Collin et al., 2013, Wu et al., 2016). In these assays the zebrafish orthologue of the gene being analysed is knocked down using morpholino's or CRISPR gene editing and mRNA corresponding to the wildtype or mutated human gene is injected into the mutant zebrafish embryos to determine if the phenotype can be rescued. However, unpublished data from the Toomes' lab has shown that SNPs and mutations can both give similar pathogenic readouts from this assay, so again it was not included as functional data in the ACMG classifications. Developing a high-throughput robust functional assay for FEVR is therefore a priority.

In the absence of a reliable functional assay, *in-silico* tools are often used to predict the pathogenic nature of a variant. In this study, four predictors (SIFT, PROVEAN, CADD and PolyPhen-2) were used to contribute to the PP3/BP4 criteria in the ACMG classification. However, there are more than 20 predictors that are available in recent years. Thus, the use of more available computational predictors may help to provide more robust prediction evidence. In addition, the CADD score used in this study to determine whether the variants were deleterious was arbitrary set to a cutoff of 15 (CADD c-score >15 as deleterious). This is very low and a single cutoff in CADD may lead to the decrease in sensitivity and specificity as the boundary to consider whether a variant is deleterious or neutral may differ in individual genes, as different genes may have divergent tolerance to variants (van der Velde et al., 2017). Thus, if gene-specific thresholds were set for the CADD score, more robust *in silico* prediction could be obtained for certain criterion (PP3/BP4) in the ACMG. A tool has been designed which assigns each disease gene an optimal cutoff based on published mutations, GAVIN software (<https://molgenis20.gcc.rug.nl/>) (van der Velde et al., 2017). This tool suggests the optional cutoff for CADD for *FZD4* is 28.45 (>28.45 as deleterious), *TSPAN12* (24.75), *CTNNB1* (28.78), *LRP5* (27.71) and *ZNF408* (32). Clearly all of these figures are much higher than the cutoff used in this study and if these were used it is likely that even more of the published mutations would be reclassified as VUS. While *in silico* tools are better than nothing, it is clear that they are not developed enough to be used confidently in a diagnostic setting. Maybe new tools developed with the aid of artificial intelligence and machine learning will lead to improvements in the future.

Moreover, only the total allele frequency, European (non-finish) and European (finish) population data from the gnomAD database were taken into consideration when addressing corresponding frequency-related criteria in the ACMG (like stand-alone strong benign evidence (BA1 criterion), applied when allele frequency > 5%). However, it could be generating more accurate results if race-matched control data were used for the variant interpretation as the consequence of the impact of genetic

diversity (different populations may have divergent tolerance to certain mutations in the gene) on the Mendelian disease (Sirugo et al., 2019). For example, the c.3361A>G (p.N1121D) variant in *LRP5*, it was classified as uncertain significance by the ACMG guideline in the study when using the total allele frequency of the variant as 0.0006110 in the gnomAD database. However, this variant was reported only in the Chinese FEVR patients, thus the using of East Asian (EA) population frequency (0.007922) will lead to the classification of the variant as likely benign in the ACMG for the EA population of FEVR patients. This also addressed the using of race-matched frequency control data is important for more accurate variant interpretation.

4.2 Protein modelling

In this study, in silico protein modelling was used to try and aid the classification of FEVR mutations. The data generated highlighted differences between the mutant and the wildtype protein structures. For example, the MUpro values (**Table 9.**) showed that the majority of missense variants in FZD4 decreased the protein stability ($\Delta\Delta G < 0$), and two missense variants (p.S497F:0.035 and p.E40Q:0.039) increased the protein stability ($\Delta\Delta G > 0$). However, further work will be required to determine if these values are robust enough to use in a diagnostic setting. For example, the analysis should be repeated with SNPs in FZD4 to see what sort of MUpro scores these generate.

Similarly, superimposing mutant and wildtype structures of FZD4 showed regions which were commonly deviated for many variants such as residue 508-537 and the CRD domain. However, at the moment it is not easy to explain why variants in the signal peptide would cause deviations in the cytoplasmic domain of FZD4 so caution must be taken in overinterpreting these data. One limitation is that the deviated regions were only based on a single backbone RMSD value in amino acid sequence level. Also, it is noteworthy that the method used in this study may not be the best way of comparing protein structures and more accurate ways exist like RamRMSD or logPr (Jung et al., 2013). Furthermore, no model optimization methods were performed in the study and the model quality evaluation score obtained from ERRAT

were not satisfactory. Moreover, the ERRAT results showed the errors were highly likely to have occurred in the C-terminal regions and this may explain why this region was highlighted for all the missense mutations in *FZD4*. Thus, the deviation region found in this comparison of the protein structures may not be robust enough to use for diagnostic classifications. For example, the benign variants reclassified in the study like p.P33S generated results which were similar to pathogenic or uncertain significant variants. New developments in this field such as dynamic modelling may generate more reliable results but at the moment the tools were not satisfactory.

In addition, the *FZD4* and NDP (Norrin) wildtype 3D structures were also predicted through I-TASSER software, whereas the experimental models for these two proteins were available in the Protein Data Bank with good coverages (*FZD4*: covered residues 42-517; NDP: covered residues: 25-133). Thus, it could contribute to generate more reliable results if wildtype experimental models were used for these two proteins as references to predict the missense mutant structures in the study.

4.3 Mutation spectra

Mutation spectra in the study were generated to help highlight the mutation hotspot in the causative genes of FEVR. The missense mutation affecting disulphide bonds were shown to be reported at a relatively high frequency in FEVR genes. For example, 14 missense variants were reported in eight disulphide bonds positions in *FZD4* and many missense variants occurred in the same positions, eg. p.C45S (likely pathogenic), p.C45Y (likely pathogenic) and p.C45R (uncertain significance). The difference between the pathogenicity classifications of these three missense variants in the same position was due to p.C45R lacking cosegregation data. However, it is likely that the p.C45R is also a pathogenic variant as disulphide bonds are known to play a key role in the stabilization of the protein structures (Karimi et al., 2016). This example highlights the limit of the ACMG classification in practical use.

In *FZD4*, the mutation hotspot was CRD domain (52 variants), which followed by KTXXXW PDZ-binding motif (6). These results further supported the importance of

these regions in the normal function of FZD4 protein. For example, the CRD domain of FZD4 is the Norrin-binding site (Zhang et al., 2011).

The mutation hotspot of *TSPAN12* found in this study was the second (large) extracellular loop (31 variants) in FEVR, which also indicated that this region may play a vital role in protein structure and function and this is consistent with reports which show that the large extracellular loop of TSPAN12 is important in enhancing the signalling of Norrin-FZD4 binding (Lai et al., 2017).

The mutation hotspot found in the *NDP* gene in the study correspond to the $\beta 1$ - $\beta 2$ hairpin, which includes 14 variants in total. The second largest number of variants were reported in the $\beta 5$ - $\beta 6$ region (11 in total). These results suggested that these two hairpins may be important in the function of Norrin and consistent with this is the work of Chang and colleagues who found that the $\beta 1$ - $\beta 2$ and $\beta 5$ - $\beta 6$ regions of Norrin contact the CRD domain of FZD4 (Chang et al., 2015).

The highest reported hotspot region of variants found in LRP5 in FEVR in the study was the Beta-propeller domains in the extracellular region of the protein; domain 1 (18 variants), domain 2 (48 variants), domain 3 (31 variants) and domain 4 (18 variants). These regions are known to bind ligands and are therefore crucial to the function of LRP5 (Ke et al., 2013).

5. Summary and future expectations

In this study, the collection and reclassification of reported variants in FEVR led to the building of a mutation spectrum and database for FEVR. This data provides a free, up to date resource for diagnostics labs and will contribute to the accurate genetic diagnosis of patients. The study also highlights the difficulty of classifying the missense variants in FEVR, and highlights the importance of creating new reliable functional assays to classify variants in FEVR.

6. References

- [1] Mutations and New Variation: Overview. *eLS*.
- [2] ADZHUBEI, I. A., SCHMIDT, S., PESHKIN, L., RAMENSKY, V. E., GERASIMOVA, A., BORK, P., KONDRASHOV, A. S. & SUNYAEV, S. R. 2010. A method and server for predicting damaging missense mutations. *Nature methods*, 7, 248-249.
- [3] AVILA-FERNANDEZ, A., PEREZ-CARRO, R., CORTON, M., LOPEZ-MOLINA, M. I., CAMPELLO, L., GARANTO, A., FERNANDEZ-SANCHEZ, L., DUIJKERS, L., LOPEZ-MARTINEZ, M. A. & RIVEIRO-ALVAREZ, R. 2015. Whole-exome sequencing reveals ZNF408 as a new gene associated with autosomal recessive retinitis pigmentosa with vitreal alterations. *Human molecular genetics*, 24, 4037-4048.
- [4] BANG, I., KIM, H. R., BEAVEN, A. H., KIM, J., KO, S.-B., LEE, G. R., KAN, W., LEE, H., IM, W. & SEOK, C. 2018. Biophysical and functional characterization of Norrin signaling through Frizzled4. *Proceedings of the National Academy of Sciences*, 115, 8787-8792.
- [5] BARRESI, M. J. F., BURTON, S., DIPIETRANTONIO, K., AMSTERDAM, A., HOPKINS, N. & KARLSTROM, R. O. 2010. Essential genes for astroglial development and axon pathfinding during zebrafish embryogenesis. *Developmental dynamics : an official publication of the American Association of Anatomists*, 239, 2603-2618.
- [6] BERGER, W., VAN DE POL, D., BÄCHNER, D., OERLEMANS, F., WINKENS, H., HAMEISTER, H., WIERINGA, B., HENDRIKS, W. & ROPERS, H. H. 1996. An animal model for Norrie disease (ND): gene targeting of the mouse ND gene. *Hum Mol Genet*, 5, 51-9.
- [7] BRANDIS, G. & HUGHES, D. 2016. The Selective Advantage of Synonymous Codon Usage Bias in Salmonella. *PLOS Genetics*, 12, e1005926.
- [8] BROWN, N. L., PATEL, S., BRZEZINSKI, J. & GLASER, T. 2001. β -catenin is required for retinal ganglion cell and optic nerve formation. *Development*, 128, 2497-2508.
- [9] BUCHER, F., FRIEDLANDER, M. & YEA, K. 2018. Antibody targeting TSPAN12/ β -catenin signaling in vasoproliferative retinopathy. *Oncotarget*, 9, 12538.
- [10] CHANG, T.-H., HSIEH, F.-L., ZEBISCH, M., HARLOS, K., ELEGHEERT, J. & JONES, E. Y. 2015. Structure and functional properties of Norrin mimic Wnt for signalling with Frizzled4, Lrp5/6, and proteoglycan. *Elife*, 4, e06554.
- [11] CHARETTE, J. R., EARP, S. E., BELL, B. A., ACKERT-BICKNELL, C. L., GODFREY, D. A., RAO, S., ANAND-APTE, B., NISHINA, P. M. & PEACHEY, N. S. 2017. A mutagenesis-derived Lrp5 mouse mutant with abnormal retinal vasculature and low bone mineral density. *Molecular Vision*, 23, 140.
- [12] CHEN, Z., BATTINELLI, E., FIELDER, A., BUNDEY, S., SIMS, K., BREAKFIELD, X. & CRAIG, I. 1993a. A mutation in the norrie disease gene (ndp) associated with x-linked familial exudative vitreoretinopathy. *Nature genetics*, 5, 180.
- [13] CHEN, Z. Y., BATTINELLI, E. M., FIELDER, A., BUNDEY, S., SIMS, K., BREAKFIELD, X. O. & CRAIG, I. W. 1993b. A mutation in the Norrie disease gene (NDP) associated with X-linked familial exudative vitreoretinopathy. *Nat Genet*, 5, 180-3.
- [14] CHENG, J., RANDALL, A. & BALDI, P. 2006. Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins*, 62, 1125-32.

- [15] CHENG, J., RANDALL, A. Z., SWEREDOSKI, M. J. & BALDI, P. 2005. SCRATCH: a protein structure and structural feature prediction server. *Nucleic acids research*, 33, W72-W76.
- [16] CHOI, Y. & CHAN, A. P. 2015. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, 31, 2745-2747.
- [17] CLEVERS, H. 2009. Eyeing up new Wnt pathway players. *Cell*, 139, 227-229.
- [18] COLLIN, R. W., NIKOPOULOS, K., DONA, M., GILISSEN, C., HOISCHEN, A., BOONSTRA, F. N., POULTER, J. A., KONDO, H., BERGER, W. & TOOMES, C. 2013. ZNF408 is mutated in familial exudative vitreoretinopathy and is crucial for the development of zebrafish retinal vasculature. *Proceedings of the National Academy of Sciences*, 110, 9856-9861.
- [19] COLLINS, J. S. & SCHWARTZ, C. E. 2002. Detecting polymorphisms and mutations in candidate genes. *Am J Hum Genet*, 71, 1251-2.
- [20] COLOVOS, C. & YEATES, T. O. 1993. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci*, 2, 1511-9.
- [21] COPPIETERS, F., ASCARI, G., DANNHAUSEN, K., NIKOPOULOS, K., PEELMAN, F., KARLSTETTER, M., XU, M., BRACHET, C., MEUNIER, I. & TSILIMBARIS, M. K. 2016. Isolated and syndromic retinal dystrophy caused by biallelic mutations in RCBTB1, a gene implicated in ubiquitination. *The American Journal of Human Genetics*, 99, 470-480.
- [22] CRISWICK, V. & SCHEPENS, C. 1969. Familial exudative vitreoretinopathy. *American journal of ophthalmology*, 68, 578-594.
- [23] DE LIGT, J., WILLEMSSEN, M. H., VAN BON, B. W., KLEEFSTRA, T., YNTEMA, H. G., KROES, T., VULTO-VAN SILFHOUT, A. T., KOOLEN, D. A., DE VRIES, P. & GILISSEN, C. 2012. Diagnostic exome sequencing in persons with severe intellectual disability. *New England Journal of Medicine*, 367, 1921-1929.
- [24] DEN DUNNEN, J. T., DALGLEISH, R., MAGLOTT, D. R., HART, R. K., GREENBLATT, M. S., MCGOWAN-JORDAN, J., ROUX, A.-F., SMITH, T., ANTONARAKIS, S. E. & TASCHNER, P. E. M. 2016. HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Human Mutation*, 37, 564-569.
- [25] DOWNEY, L., BOTTOMLEY, H., SHERIDAN, E., AHMED, M., GILMOUR, D., INGLEHEARN, C., REDDY, A., AGRAWAL, A., BRADBURY, J. & TOOMES, C. 2006. Reduced bone mineral density and hyaloid vasculature remnants in a consanguineous recessive FEVR family with a mutation in LRP5. *British journal of ophthalmology*, 90, 1163-1167.
- [26] DOWNEY, L., KEEN, T., ROBERTS, E., MANSFIELD, D., BAMASHMUS, M. & INGLEHEARN, C. 2001. A new locus for autosomal dominant familial exudative vitreoretinopathy maps to chromosome 11p12-13. *The American Journal of Human Genetics*, 68, 778-781.
- DRENSER, K. A. 2016. Wnt signaling pathway in retinal vascularization. *Eye and Brain*, 8, 141.
- [27] DUBRUC, E., PUTOUX, A., LABALME, A., ROUGEOT, C., SANLAVILLE, D. & EDERY, P. 2014. A new intellectual disability syndrome caused by CTNNB1 haploinsufficiency. *American Journal of Medical Genetics Part A*, 164, 1571-1575.
- [28] EISENBERG, D., LÜTHY, R. & BOWIE, J. U. 1997. [20] VERIFY3D: Assessment of protein models with three-dimensional profiles. *Methods in Enzymology*. Academic Press.
- [29] EXERTIER, P., JAVERZAT, S., WANG, B., FRANCO, M., HERBERT, J., PLATONOVA, N., WINANDY, M., PUJOL, N., NIVELLES, O., ORMENESE, S., GODARD, V., BECKER, J.,

- BICKNELL, R., PINEAU, R., WILTING, J., BIKFALVI, A. & HAGEDORN, M. 2013. Impaired angiogenesis and tumor development by inhibition of the mitotic kinesin Eg5. *Oncotarget*, 4, 2302-2316.
- [30] FISCHETTO, R., PALMIERI, V. V., TRIPALDI, M. E., GAETA, A., MICHELUCCI, A., DELVECCHIO, M., FRANCAVILLA, R. & GIORDANO, P. 2019. Alagille Syndrome: A Novel Mutation in JAG1 Gene. *Frontiers in pediatrics*, 7, 199-199.
- [31] GANDHI, J. K., TOLLEFSON, T. T. & TELANDER, D. G. 2014. Falciform macular folds and chromosome 22q11.2: evidence in support of a locus for familial exudative vitreoretinopathy (FEVR). *Ophthalmic Genet*, 35, 112-6.
- [32] GHIASVAND, N. M., RUDOLPH, D. D., MASHAYEKHI, M., BRZEZINSKI IV, J. A., GOLDMAN, D. & GLASER, T. 2011. Deletion of a remote enhancer near ATOH7 disrupts retinal neurogenesis, causing NCRNA disease. *Nature neuroscience*, 14, 578.
- [33] GILISSEN, C., HOISCHEN, A., BRUNNER, H. G. & VELTMAN, J. A. 2012. Disease gene identification strategies for exome sequencing. *Eur J Hum Genet*, 20, 490-7.
- [34] GILMOUR, D. 2015. Familial exudative vitreoretinopathy and related retinopathies. *Eye*, 29, 1-14.
- [35] GILMOUR, D. F., DOWNEY, L. M., SHERIDAN, E., LONG, V., BRADBURY, J., INGLEHEARN, C. F. & TOOMES, C. 2009. Familial exudative vitreoretinopathy and DiGeorge syndrome: a new locus for familial exudative vitreoretinopathy on chromosome 22q11. 2? *Ophthalmology*, 116, 1522-1524.
- [36] GONG, Y., SLEE, R. B., FUKAI, N., RAWADI, G., ROMAN-ROMAN, S., REGINATO, A. M., WANG, H., CUNDY, T., GLORIEUX, F. H., LEV, D., ZACHARIN, M., OEXLE, K., MARCELINO, J., SUWAIRI, W., HEEGER, S., SABATAKOS, G., APTE, S., ADKINS, W. N., ALLGROVE, J., ARSLAN-KIRCHNER, M., BATCH, J. A., BEIGHTON, P., BLACK, G. C., BOLES, R. G., BOON, L. M., BORRONE, C., BRUNNER, H. G., CARLE, G. F., DALLAPICCOLA, B., DE PAEPE, A., FLOEGE, B., HALFHIDE, M. L., HALL, B., HENNEKAM, R. C., HIROSE, T., JANS, A., JÜPPNER, H., KIM, C. A., KEPPLER-NOREUIL, K., KOHLSCHUETTER, A., LACOMBE, D., LAMBERT, M., LEMYRE, E., LETTEBOER, T., PELTONEN, L., RAMESAR, R. S., ROMANENGO, M., SOMER, H., STEICHEN-GERSDORF, E., STEINMANN, B., SULLIVAN, B., SUPERTI-FURGA, A., SWOBODA, W., VAN DEN BOOGAARD, M. J., VAN HUL, W., VIKKULA, M., VOTRUBA, M., ZABEL, B., GARCIA, T., BARON, R., OLSEN, B. R. & WARMAN, M. L. 2001. LDL receptor-related protein 5 (LRP5) affects bone accrual and eye development. *Cell*, 107, 513-23.
- [37] GOTTARDI, C. J. & PEIFER, M. 2008. Terminal regions of beta-catenin come into view. *Structure*, 16, 336-8.
- [38] GOW, J. & OLIVER, G. L. 1971. Familial exudative vitreoretinopathy: an expanded view. *Archives of Ophthalmology*, 86, 150-155.
- [39] GRONT, D., KMIECIK, S., BLASZCZYK, M., EKONOMIUK, D. & KOLIŃSKI, A. 2012. Optimization of protein models. *WIREs Computational Molecular Science*, 2, 479-493.
- [40] HAN, S., SUN, J., YANG, L. & QI, M. 2020. Role of *NDP*- and *FZD4*-Related Novel Mutations Identified in Patients with FEVR in Norrin/ β -Catenin Signaling Pathway. *BioMed Research International*, 2020, 7681926.

- [41] HANANY, M. & SHARON, D. 2019. Allele frequency analysis of variants reported to cause autosomal dominant inherited retinal diseases question the involvement of 19% of genes and 10% of reported pathogenic variants. *J Med Genet*, 56, 536-542.
- [42] HENRY, C. R., SISK, R. A., TZU, J. H., ALBINI, T. A., DAVIS, J. L., MURRAY, T. G. & BERROCAL, A. M. 2015. Long-term follow-up of intravitreal bevacizumab for the treatment of pediatric retinal and choroidal diseases. *Journal of American Association for Pediatric Ophthalmology and Strabismus {JAAPOS}*, 19, 541-548.
- [43] HOFFMAN-ANDREWS, L. 2017. The known unknown: the challenges of genetic variants of uncertain significance in clinical practice. *Journal of Law and the Biosciences*, 4, 648-657.
- [44] HSIEH, M., BOERBOOM, D., SHIMADA, M., LO, Y., PARLOW, A. F., LUHMANN, U. F., BERGER, W. & RICHARDS, J. S. 2005. Mice null for Frizzled4 (Fzd4^{-/-}) are infertile and exhibit impaired corpora lutea formation and function. *Biol Reprod*, 73, 1135-46.
- [45] JARVIK, G. P. & BROWNING, B. L. 2016. Consideration of Cosegregation in the Pathogenicity Classification of Genomic Variants. *American journal of human genetics*, 98, 1077-1081.
- [46] JIAO, X., VENTRUTO, V., TRESE, M. T., SHASTRY, B. S. & HEJTMANCIK, J. F. 2004. Autosomal recessive familial exudative vitreoretinopathy is associated with mutations in LRP5. *The American Journal of Human Genetics*, 75, 878-884.
- [47] JUNG, S., BAE, S.-E., AHN, I. & SON, H. S. 2013. Protein backbone torsion angle-based structure comparison and secondary structure database web server. *Genomics & informatics*, 11, 155-160.
- [48] JUNGE, H. J., YANG, S., BURTON, J. B., PAES, K., SHU, X., FRENCH, D. M., COSTA, M., RICE, D. S. & YE, W. 2009. TSPAN12 regulates retinal vascular development by promoting Norrin but not Wnt-induced FZD4/beta-catenin signaling. *Cell*, 139, 299-311.
- [49] KANG, K. B., WESSEL, M. M., TONG, J., D'AMICO, D. J. & CHAN, R. V. 2013. Ultra-widefield imaging for the management of pediatric retinal diseases. *J Pediatr Ophthalmol Strabismus*, 50, 282-8.
- [50] KARCZEWSKI, K. J., FRANCIOLI, L. C., TIAO, G., CUMMINGS, B. B., ALFÖLDI, J., WANG, Q., COLLINS, R. L., LARICCHIA, K. M., GANNA, A., BIRNBAUM, D. P., GAUTHIER, L. D., BRAND, H., SOLOMONSON, M., WATTS, N. A., RHODES, D., SINGER-BERK, M., ENGLAND, E. M., SEABY, E. G., KOSMICKI, J. A., WALTERS, R. K., TASHMAN, K., FARJOUN, Y., BANKS, E., POTERBA, T., WANG, A., SEED, C., WHIFFIN, N., CHONG, J. X., SAMOCHA, K. E., PIERCE-HOFFMAN, E., ZAPPALA, Z., O'DONNELL-LURIA, A. H., MINIKEL, E. V., WEISBURD, B., LEK, M., WARE, J. S., VITTAL, C., ARMEAN, I. M., BERGELSON, L., CIBULSKIS, K., CONNOLLY, K. M., COVARRUBIAS, M., DONNELLY, S., FERRIERA, S., GABRIEL, S., GENTRY, J., GUPTA, N., JEANDET, T., KAPLAN, D., LLANWARNE, C., MUNSHI, R., NOVOD, S., PETRILLO, N., ROAZEN, D., RUANO-RUBIO, V., SALTZMAN, A., SCHLEICHER, M., SOTO, J., TIBBETTS, K., TOLONEN, C., WADE, G., TALKOWSKI, M. E., AGUILAR SALINAS, C. A., AHMAD, T., ALBERT, C. M., ARDISSINO, D., ATZMON, G., BARNARD, J., BEAUGERIE, L., BENJAMIN, E. J., BOEHNKE, M., BONNYCASTLE, L. L., BOTTINGER, E. P., BOWDEN, D. W., BOWN, M. J., CHAMBERS, J. C., CHAN, J. C., CHASMAN, D., CHO, J., CHUNG, M. K., COHEN, B., CORREA, A., DABELEA, D., DALY, M. J., DARBAR, D., DUGGIRALA, R., DUPUIS, J., ELLINOR, P. T., ELOSUA, R., ERDMANN, J., ESKO, T., FÄRKKILÄ, M., FLOREZ, J., FRANKE, A., GETZ, G., GLASER, B.,

- GLATT, S. J., GOLDSTEIN, D., GONZALEZ, C., GROOP, L., et al. 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581, 434-443.
- [51] KARIMI, M., IGNASIAK, M. T., CHAN, B., CROFT, A. K., RADOM, L., SCHIESSER, C. H., PATTISON, D. I. & DAVIES, M. J. 2016. Reactivity of disulfide bonds is markedly affected by structure and environment: implications for protein modification and stability. *Scientific Reports*, 6, 38572.
- [52] KASHANI, A. H., BROWN, K. T., CHANG, E., DRENSER, K. A., CAPONE, A. & TRESE, M. T. 2014. Diversity of retinal vascular anomalies in patients with familial exudative vitreoretinopathy. *Ophthalmology*, 121, 2220-2227.
- [53] KATAGIRI, S., YOKOI, T., YOSHIDA-UEMURA, T., NISHINA, S. & AZUMA, N. 2018. Characteristics of Retinal Breaks and Surgical Outcomes in Rhegmatogenous Retinal Detachment in Familial Exudative Vitreoretinopathy. *Ophthalmology Retina*, 2, 720-725.
- [54] KATO, M., PATEL, M. S., LEVASSEUR, R., LOBOV, I., CHANG, B. H.-J., GLASS, D. A., II, HARTMANN, C., LI, L., HWANG, T.-H., BRAYTON, C. F., LANG, R. A., KARSENTY, G. & CHAN, L. 2002. Cbfa1-independent decrease in osteoblast proliferation, osteopenia, and persistent embryonic eye vascularization in mice deficient in Lrp5, a Wnt coreceptor. *Journal of Cell Biology*, 157, 303-314.
- [55] KE, J., HARIKUMAR, K. G., ERICE, C., CHEN, C., GU, X., WANG, L., PARKER, N., CHENG, Z., XU, W. & WILLIAMS, B. O. 2013. Structure and function of Norrin in assembly and activation of a Frizzled 4–Lrp5/6 complex. *Genes & development*, 27, 2305-2319.
- [56] KELLEY, L. A., MEZULIS, S., YATES, C. M., WASS, M. N. & STERNBERG, M. J. E. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10, 845-858.
- [57] KHAN, K., LOGAN, C. V., MCKIBBIN, M., SHERIDAN, E., ELCIOGLU, N. H., YENICE, O., PARRY, D. A., FERNANDEZ-FUENTES, N., ABDELHAMED, Z. I. & AL-MASKARI, A. 2011. Next generation sequencing identifies mutations in Atonal homolog 7 (ATOH7) in families with global eye developmental defects. *Human molecular genetics*, 21, 776-783.
- [58] KLEINBERGER, J., MALONEY, K. A., POLLIN, T. I. & JENG, L. J. 2016. An openly available online tool for implementing the ACMG/AMP standards and guidelines for the interpretation of sequence variants. *Genet Med*, 18, 1165.
- [59] KONDO, H., HAYASHI, H., OSHIMA, K., TAHIRA, T. & HAYASHI, K. 2003. Frizzled 4 gene (FZD4) mutations in patients with familial exudative vitreoretinopathy with variable expressivity. *British journal of ophthalmology*, 87, 1291-1295.
- [60] KOPANOS, C., TSIOLKAS, V., KOURIS, A., CHAPPLE, C. E., AGUILERA, M. A., MEYER, R. & MASSOURAS, A. 2019. VarSome: the human genomic variant search engine. *Bioinformatics*, 35, 1978.
- [61] KUFAREVA, I. & ABAGYAN, R. 2012. Methods of protein structure comparison. *Methods Mol Biol*, 857, 231-57.
- [62] LAI, M. B., ZHANG, C., SHI, J., JOHNSON, V., KHANDAN, L., MCVEY, J., KLYMKOWSKY, M. W., CHEN, Z. & JUNGE, H. J. 2017. TSPAN12 is a Norrin co-receptor that amplifies Frizzled4 ligand selectivity and signaling. *Cell reports*, 19, 2809-2822.
- [63] LANDRUM, M. J., CHITIPIRALLA, S., BROWN, G. R., CHEN, C., GU, B., HART, J., HOFFMAN, D., JANG, W., KAUR, K., LIU, C., LYOSHIN, V., MADDIPATLA, Z., MAITI, R.,

- MITCHELL, J., O'LEARY, N., RILEY, G. R., SHI, W., ZHOU, G., SCHNEIDER, V., MAGLOTT, D., HOLMES, J. B. & KATTMAN, B. L. 2020. ClinVar: improvements to accessing data. *Nucleic Acids Res*, 48, D835-d844.
- [64] LANDRUM, M. J., LEE, J. M., BENSON, M., BROWN, G. R., CHAO, C., CHITIPIRALLA, S., GU, B., HART, J., HOFFMAN, D., JANG, W., KARAPETYAN, K., KATZ, K., LIU, C., MADDIPATLA, Z., MALHEIRO, A., MCDANIEL, K., OVETSKY, M., RILEY, G., ZHOU, G., HOLMES, J. B., KATTMAN, B. L. & MAGLOTT, D. R. 2018. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*, 46, D1062-d1067.
- [65] LAPPALAINEN, I., LOPEZ, J., SKIPPER, L., HEFFERON, T., SPALDING, J. D., GARNER, J., CHEN, C., MAGUIRE, M., CORBETT, M., ZHOU, G., PASCHALL, J., ANANIEV, V., FLICEK, P. & CHURCH, D. M. 2013. DbVar and DGVa: public archives for genomic structural variation. *Nucleic Acids Res*, 41, D936-41.
- [66] LAQUA, H. 1980. Familial exudative vitreoretinopathy. *Albrecht von Graefes Archiv für klinische und experimentelle Ophthalmologie*, 213, 121-133.
- [67] LASKOWSKI, R. A., RULLMANN, J. A., MACARTHUR, M. W., KAPTEIN, R. & THORNTON, J. M. 1996. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR*, 8, 477-86.
- [68] LI, Q. & WANG, K. 2017. InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. *The American Journal of Human Genetics*, 100, 267-280.
- [69] LOVELL, S. C., DAVIS, I. W., ARENDALL, W. B., 3RD, DE BAKKER, P. I., WORD, J. M., PRISANT, M. G., RICHARDSON, J. S. & RICHARDSON, D. C. 2003. Structure validation by C α geometry: phi,psi and C β deviation. *Proteins*, 50, 437-50.
- [70] LU, Y.-Z., DENG, G.-D., LIU, J.-H. & YAN, H. 2018. The role of intravitreal ranibizumab in the treatment of familial exudative vitreoretinopathy of stage 2 or greater. *International journal of ophthalmology*, 11, 976-980.
- [71] LUHMANN, U. F., MEUNIER, D., SHI, W., LÜTTGES, A., PFARRER, C., FUNDELE, R. & BERGER, W. 2005. Fetal loss in homozygous mutant Norrie disease mice: a new role of Norrin in reproduction. *Genesis*, 42, 253-62.
- [72] LUHMANN, U. F., NEIDHARDT, J., KLOECKENER-GRUISSEM, B., SCHÄFER, N. F., GLAUS, E., FEIL, S. & BERGER, W. 2008. Vascular changes in the cerebellum of Norrin /Ndp knockout mice correlate with high expression of Norrin and Frizzled-4. *Eur J Neurosci*, 27, 2619-28.
- [73] MICHALSKA-MAŁECKA, K., KABIESZ, A., KIMSA, M. W., STRZAŁKA-MROZIK, B., FORMIŃSKA-KAPUŚCIK, M., NITA, M. & MAZUREK, U. 2016. Effects of intravitreal ranibizumab on the untreated eye and systemic gene expression profile in age-related macular degeneration. *Clin Interv Aging*, 11, 357-65.
- [74] MILLER, K. E., WILLIS, M. J. & MCCLATCHEY, S. K. 2015. A case of familial exudative vitreoretinopathy identified after genetic testing. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, 19, 178-180.
- [75] MO, R., CHEW, T.-L., MAHER, M. T., BELLIPANNI, G., WEINBERG, E. S. & GOTTARDI, C. J. 2009. The terminal region of beta-catenin promotes stability by shielding the Armadillo repeats from the axin-scaffold destruction complex. *The Journal of biological chemistry*, 284, 28222-28231.

- [76] MUSADA, G. R., JALALI, S., HUSSAIN, A., CHURURU, A. R., GADDAM, P. R., CHAKRABARTI, S. & KAUR, I. 2016. Mutation spectrum of the Norrie disease pseudoglioma (NDP) gene in Indian patients with FEVR. *Molecular vision*, 22, 491.
- [77] NIKOPOULOS, K., GILISSEN, C., HOISCHEN, A., VAN NOUHUYS, C. E., BOONSTRA, F. N., BLOKLAND, E. A., ARTS, P., WIESKAMP, N., STROM, T. M. & AYUSO, C. 2010. Next-generation sequencing of a 40 Mb linkage interval reveals TSPAN12 mutations in patients with familial exudative vitreoretinopathy. *The American Journal of Human Genetics*, 86, 240-247.
- [78] O'ROAK, B. J., VIVES, L., FU, W., EGERTSON, J. D., STANAWAY, I. B., PHELPS, I. G., CARVILL, G., KUMAR, A., LEE, C. & ANKENMAN, K. 2012. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science*, 338, 1619-1622.
- [79] OSTERGAARD, P., SIMPSON, M. A., MENDOLA, A., VASUDEVAN, P., CONNELL, F. C., VAN IMPEL, A., MOORE, A. T., LOEYS, B. L., GHALAMKARPOUR, A., ONOUFRIADIS, A., MARTINEZ-CORRAL, I., DEVERY, S., LEROY, J. G., VAN LAER, L., SINGER, A., BIALER, M. G., MCENTAGART, M., QUARRELL, O., BRICE, G., TREMBATH, R. C., SCHULTE-MERKER, S., MAKINEN, T., VIKKULA, M., MORTIMER, P. S., MANSOUR, S. & JEFFERY, S. 2012. Mutations in KIF11 cause autosomal-dominant microcephaly variably associated with congenital lymphedema and chorioretinopathy. *American journal of human genetics*, 90, 356-362.
- [80] PANAGIOTOU, E. S., SANJURJO SORIANO, C., POULTER, J. A., LORD, E. C., DZULOVA, D., KONDO, H., HIYOSHI, A., CHUNG, B. H.-Y., CHU, Y. W.-Y., LAI, C. H. Y., TAFOYA, M. E., KARJOSUKARSO, D., COLLIN, R. W. J., TOPPING, J., DOWNEY, L. M., ALI, M., INGLEHEARN, C. F. & TOOMES, C. 2017. Defects in the Cell Signaling Mediator β -Catenin Cause the Retinal Vascular Condition FEVR. *The American Journal of Human Genetics*, 100, 960-968.
- [81] PARK, H., OVCHINNIKOV, S., KIM, D. E., DIMAIO, F. & BAKER, D. 2018. Protein homology model refinement by large-scale energy optimization. *Proceedings of the National Academy of Sciences*, 115, 3054-3059.
- [82] PARK, H., PARK, S. Y. & RYU, S. E. 2013. Homology modeling and virtual screening approaches to identify potent inhibitors of slingshot phosphatase 1. *Journal of Molecular Graphics and Modelling*, 39, 65-70.
- [83] PARK, H., YAMAMOTO, H., MOHN, L., AMBÜHL, L., KANAI, K., SCHMIDT, I., KIM, K.-P., FRACCAROLI, A., FEIL, S. & JUNGE, H. J. 2019. Integrin-linked kinase controls retinal angiogenesis and is linked to Wnt signaling and exudative vitreoretinopathy. *Nature communications*, 10, 1-14.
- [84] PENDERGAST, S. D. & TRESE, M. T. 1998. Familial exudative vitreoretinopathy: results of surgical management. *Ophthalmology*, 105, 1015-1023.
- [85] PETTERSEN, E. F., GODDARD, T. D., HUANG, C. C., COUCH, G. S., GREENBLATT, D. M., MENG, E. C. & FERRIN, T. E. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*, 25, 1605-12.
- [86] POULTER, J. A., ALI, M., GILMOUR, D. F., RICE, A., KONDO, H., HAYASHI, K., MACKEY, D. A., KEARNS, L. S., RUDDLE, J. B. & CRAIG, J. E. 2010. Mutations in TSPAN12 cause autosomal-dominant familial exudative vitreoretinopathy. *The American Journal of Human Genetics*, 86, 248-253.
- [87] POULTER, J. A., DAVIDSON, A. E., ALI, M., GILMOUR, D. F., PARRY, D. A., MINTZ-HITNER, H. A., CARR, I. M., BOTTOMLEY, H. M., LONG, V. W. & DOWNEY, L. M. 2012.

Recessive mutations in TSPAN12 cause retinal dysplasia and severe familial exudative vitreoretinopathy (FEVR). *Investigative ophthalmology & visual science*, 53, 2873-2879.

[88] PRASOV, L., MASUD, T., KHALIQ, S., MEHDI, S. Q., ABID, A., OLIVER, E. R., SILVA, E. D., LEWANDA, A., BRODSKY, M. C. & BORCHERT, M. 2012. ATOH7 mutations cause autosomal recessive persistent hyperplasia of the primary vitreous. *Human molecular genetics*, 21, 3681-3694.

[89] QIN, M., KONDO, H., TAHIRA, T. & HAYASHI, K. 2008. Moderate reduction of Norrin signaling activity associated with the causative missense mutations identified in patients with familial exudative vitreoretinopathy. *Human genetics*, 122, 615-623.

[90] QUIRAM, P. A., DRENSER, K. A., LAI, M. M., CAPONE, A. J. & TRESE, M. T. 2008. TREATMENT OF VASCULARLY ACTIVE FAMILIAL EXUDATIVE VITREORETINOPATHY WITH PEGAPTANIB SODIUM (MACUGEN). *RETINA*, 28.

[91] RAMACHANDRAN, G. N., RAMAKRISHNAN, C. & SASISEKHARAN, V. 1963. Stereochemistry of polypeptide chain configurations. *J Mol Biol*, 7, 95-9.

[92] REHM, H. L., ZHANG, D.-S., BROWN, M. C., BURGESS, B., HALPIN, C., BERGER, W., MORTON, C. C., COREY, D. P. & CHEN, Z.-Y. 2002. Vascular Defects and Sensorineural Deafness in a Mouse Model of Norrie Disease. *The Journal of Neuroscience*, 22, 4286.

[93] RENTZSCH, P., WITTEN, D., COOPER, G. M., SHENDURE, J. & KIRCHER, M. 2019. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic acids research*, 47, D886-D894.

[94] RICHARDS, S., AZIZ, N., BALE, S., BICK, D., DAS, S., GASTIER-FOSTER, J., GRODY, W. W., HEGDE, M., LYON, E. & SPECTOR, E. 2015a. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in medicine*, 17, 405-423.

[95] RICHARDS, S., AZIZ, N., BALE, S., BICK, D., DAS, S., GASTIER-FOSTER, J., GRODY, W. W., HEGDE, M., LYON, E., SPECTOR, E., VOELKERDING, K. & REHM, H. L. 2015b. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*, 17, 405-24.

[96] RICHTER, M., GOTTANKA, J., MAY, C. A., WELGE-LÜSSEN, U., BERGER, W. & LÜTJEN-DRECOLL, E. 1998. Retinal vasculature changes in Norrie disease mice. *Invest Ophthalmol Vis Sci*, 39, 2450-7.

[97] ROBITAILLE, J., MACDONALD, M. L., KAYKAS, A., SHELDAHL, L. C., ZEISLER, J., DUBÉ, M.-P., ZHANG, L.-H., SINGARAJA, R. R., GUERNSEY, D. L. & ZHENG, B. 2002. Mutant frizzled-4 disrupts retinal angiogenesis in familial exudative vitreoretinopathy. *Nature genetics*, 32, 326.

[98] ROBITAILLE, J. M., GILLETT, R. M., LEBLANC, M. A., GASTON, D., NIGHTINGALE, M., MACKLEY, M. P., PARKASH, S., HATHAWAY, J., THOMAS, A. & ELLS, A. 2014. Phenotypic overlap between familial exudative vitreoretinopathy and microcephaly, lymphedema, and chorioretinal dysplasia caused by KIF11 mutations. *JAMA ophthalmology*, 132, 1393-1399.

[99] ROBITAILLE, J. M., ZHENG, B., WALLACE, K., BEIS, M. J., TATLIDIL, C., YANG, J., SHEIDOW, T. G., SIEBERT, L., LEVIN, A. V. & LAM, W.-C. 2011. The role of Frizzled-4 mutations

- in familial exudative vitreoretinopathy and Coats disease. *British Journal of Ophthalmology*, 95, 574-579.
- [100] SAVISAAR, R. & HURST, L. D. 2018. Exonic splice regulation imposes strong selection at synonymous sites. *Genome research*, 28, 1442-1454.
- [101] SCHWARZ, J. M., COOPER, D. N., SCHUELKE, M. & SEELOW, D. 2014. MutationTaster2: mutation prediction for the deep-sequencing age. *Nature methods*, 11, 361-362.
- [102] SCHWEDE, T. 2013. Protein modeling: what happened to the "protein structure gap"? *Structure (London, England : 1993)*, 21, 1531-1540.
- [103] SEEMAB, S., PERVAIZ, N., ZEHRA, R., ANWAR, S., BAO, Y. & ABBASI, A. A. 2019. Molecular evolutionary and structural analysis of familial exudative vitreoretinopathy associated FZD4 gene. *BMC Evolutionary Biology*, 19, 72.
- [104] SHUKLA, D., SINGH, J., SUDHEER, G., SOMAN, M., JOHN, R. K., RAMASAMY, K. & PERUMALSAMY, N. 2003. Familial exudative vitreoretinopathy (FEVR). Clinical profile and management. *Indian J Ophthalmol*, 51, 323-8.
- [105] SIMS, K. B. 2014. NDP-related retinopathies. *GeneReviews®[Internet]*. University of Washington, Seattle.
- [106] SIRUGO, G., WILLIAMS, S. M. & TISHKOFF, S. A. 2019. The missing diversity in human genetic studies. *Cell*, 177, 26-31.
- [107] SONI, N. & MADHUSUDHAN, M. S. 2017. Computational modeling of protein assemblies. *Current Opinion in Structural Biology*, 44, 179-189.
- [108] SONI, S. & ALOK, J. 2018. STRUCTURAL CHARACTERISATION OF 5-HYDROXYTRYPTAMINE2A RECEPTOR IN HOMO SAPIENS BY IN - SILICO METHOD. *Asian Journal of Pharmaceutical and Clinical Research*, 11.
- [109] STENSON, P. D., MORT, M., BALL, E. V., CHAPMAN, M., EVANS, K., AZEVEDO, L., HAYDEN, M., HEYWOOD, S., MILLAR, D. S., PHILLIPS, A. D. & COOPER, D. N. 2020. The Human Gene Mutation Database (HGMD®): optimizing its use in a clinical diagnostic or research setting. *Human Genetics*, 139, 1197-1207.
- [110] TEMKAR, S., AZAD, S. V., CHAWLA, R., DAMODARAN, S., GARG, G., REGANI, H., NAWAZISH, S., RAJ, N. & VENKATRAMAN, V. 2019. Ultra-widefield fundus fluorescein angiography in pediatric retinal vascular diseases. *Indian journal of ophthalmology*, 67, 788.
- [111] TOOMES, C., BOTTOMLEY, H. M., JACKSON, R. M., TOWNS, K. V., SCOTT, S., MACKEY, D. A., CRAIG, J. E., JIANG, L., YANG, Z. & TREMBATH, R. 2004. Mutations in LRP5 or FZD4 underlie the common familial exudative vitreoretinopathy locus on chromosome 11q. *The American Journal of Human Genetics*, 74, 721-730.
- [112] TOOMES, C. & DOWNEY, L. 2011. Familial exudative vitreoretinopathy, autosomal dominant. *GeneReviews®[Internet]*. University of Washington, Seattle.
- [113] URANO, T., SHIRAKI, M., USUI, T., SASAKI, N., OUCHI, Y. & INOUE, S. 2009. A1330V Variant of the Low-density Lipoprotein Receptorrelated Protein 5 (LRP5) Gene Decreases Wnt Signaling and Affects the Total Body Bone Mineral Density in Japanese Women. *Endocrine Journal*, 56, 625-631.
- [114] VAN DER VELDE, K. J., DE BOER, E. N., VAN DIEMEN, C. C., SIKKEMA-RADDATZ, B., ABBOTT, K. M., KNOPPERS, A., FRANKE, L., SIJMONS, R. H., DE KONING, T. J.,

- WIJMENGA, C., SINKE, R. J. & SWERTZ, M. A. 2017. GAVIN: Gene-Aware Variant Interpretation for medical sequencing. *Genome Biology*, 18, 6.
- [115] VAN SCHIE, E. H. & VAN AMERONGEN, R. 2020. Aberrant WNT/CTNNB1 Signaling as a Therapeutic Target in Human Breast Cancer: Weighing the Evidence. *Frontiers in cell and developmental biology*, 8, 25-25.
- [116] VASER, R., ADUSUMALLI, S., LENG, S. N., SIKIC, M. & NG, P. C. 2016. SIFT missense predictions for genomes. *Nature protocols*, 11, 1.
- [117] WANG, Y., HUSO, D., CAHILL, H., RYUGO, D. & NATHANS, J. 2001. Progressive cerebellar, auditory, and esophageal dysfunction caused by targeted disruption of the frizzled-4 gene. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 21, 4761-4771.
- [118] WATERHOUSE, A., BERTONI, M., BIENERT, S., STUDER, G., TAURIELLO, G., GUMIENNY, R., HEER, F. T., DE BEER, T. A. P., REMPFER, C., BORDOLI, L., LEPORE, R. & SCHWEDE, T. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, 46, W296-W303.
- [119] WEBB, B. & SALI, A. 2016. Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Bioinformatics*, 54, 5.6.1-5.6.37.
- [120] WILDEMAN, M., VAN OPHUIZEN, E., DEN DUNNEN, J. T. & TASCHNER, P. E. 2008. Improving sequence variant descriptions in mutation databases and literature using the Mutalyzer sequence variation nomenclature checker. *Hum Mutat*, 29, 6-13.
- [121] WU, J.-H., LIU, J.-H., KO, Y.-C., WANG, C.-T., CHUNG, Y.-C., CHU, K.-C., LIU, T.-T., CHAO, H.-M., JIANG, Y.-J. & CHEN, S.-J. 2016. Haploinsufficiency of RCBTB1 is associated with Coats disease and familial exudative vitreoretinopathy. *Human molecular genetics*, 25, 1637-1647.
- [122] XING, Y., TAKEMARU, K.-I., LIU, J., BERNDT, J. D., ZHENG, J. J., MOON, R. T. & XU, W. 2008. Crystal structure of a full-length β -catenin. *Structure*, 16, 478-487.
- [123] XU, Q., WANG, Y., DABDOUB, A., SMALLWOOD, P. M., WILLIAMS, J., WOODS, C., KELLEY, M. W., JIANG, L., TASMAN, W., ZHANG, K. & NATHANS, J. 2004. Vascular Development in the Retina and Inner Ear: Control by Norrin and Frizzled-4, a High-Affinity Ligand-Receptor Pair. *Cell*, 116, 883-895.
- [124] YANG, J. & ZHANG, Y. 2015. I-TASSER server: new development for protein structure and function predictions. *Nucleic acids research*, 43, W174-W181.
- [125] YANG, Z., LASKER, K., SCHNEIDMAN-DUHOVNY, D., WEBB, B., HUANG, C. C., PETERSEN, E. F., GODDARD, T. D., MENG, E. C., SALI, A. & FERRIN, T. E. 2012. UCSF Chimera, MODELLER, and IMP: an integrated modeling system. *Journal of structural biology*, 179, 269-278.
- [126] YE, X., WANG, Y., CAHILL, H., YU, M., BADEA, T. C., SMALLWOOD, P. M., PEACHEY, N. S. & NATHANS, J. 2009. Norrin, Frizzled-4, and Lrp5 Signaling in Endothelial Cells Controls a Genetic Program for Retinal Vascularization. *Cell*, 139, 285-298.
- [127] YE, X., WANG, Y. & NATHANS, J. 2010. The Norrin/Frizzled4 signaling pathway in retinal vascular development and disease. *Trends Mol Med*, 16, 417-25.
- [128] ZHANG, K., HARADA, Y., WEI, X., SHUKLA, D., RAJENDRAN, A., TAWANSY, K., BEDELL, M., LIM, S., SHAW, P. X. & HE, X. 2011. An essential role of the cysteine-rich domain of

FZD4 in Norrin/Wnt signaling and familial exudative vitreoretinopathy. *Journal of Biological Chemistry*, 286, 10210-10215.

[129] ZHANG, L., ZHANG, X., XU, H., HUANG, L., ZHANG, S., LIU, W., YANG, Y., FEI, P., LI, S. & YANG, M. 2019. Exome sequencing revealed Notch ligand JAG1 as a novel candidate gene for familial exudative vitreoretinopathy. *Genetics in Medicine*, 1-8.

[130] ZHAO, Y. & SINGH, R. P. 2018. The role of anti-vascular endothelial growth factor (anti-VEGF) in the management of proliferative diabetic retinopathy. *Drugs in context*, 7, 212532-212532.

7. Appendices

Appendix 1.

Table A 1. Model comparison of missense mutant structure and wildtype FZD4 structure of FEVR patients (Raw data)

Mutant structures	Deviation in average distance of the atoms (RMSD backbone score) (residue number). *(a)
G22E	5-6, 10-12, 14, 16, 18-21, 25-28, 30-34, 51, 70, 109-111, 133-136, 141, 160-161, 163-164, 195, 208-209, 244, 246, 273-276, 335, 418-423, 466, 505-537
P33S	4, 6, 8, 10-12, 14-38, 40-43, 47-52, 55-59, 63-64, 66-78, 80, 109-112, 118, 133-161, 163-166, 208-209, 240, 243-244, 249, 275-276, 295-297, 336, 390, 404, 407-424, 432, 466, 496-537
E40Q	4-6, 10-13, 15-17, 20-21, 24-37, 41-51, 57-64, 66-77, 102, 106-112, 125, 133-142, 145-149, 163-164, 176, 183, 186, 195, 201, 240, 242-247, 276, 297, 330, 332-336, 339-341, 408, 411-412, 415-419, 421-425, 428, 466-467, 496, 498, 500-537
C45S	2-10, 14, 16-21, 23, 26-27, 30, 32-49, 51, 56-71, 86, 88, 104-115, 132-133, 135-145, 147-153, 158, 163-164, 168-170, 178, 182, 185-192, 203, 205-209, 211-214, 217, 244, 246, 273-277, 279, 283-285, 289, 295-299, 332, 334-338, 342, 372-373, 382, 390, 407-427, 430, 465-466, 468-470, 497-499, 502-537
C45R	5-6, 11-12, 14, 16, 18-21, 24-33, 36, 47-48, 51, 59, 109-110, 133-137, 141, 163-164, 195, 208-209, 244, 273-275, 335, 418, 421, 423, 466, 502-537
C45Y	6, 8, 10, 12-24, 26-27, 30-32, 36-52, 56-74, 86, 103-114, 133, 137, 141-145, 147-149, 151-153, 163-164, 168, 171-173, 176-178, 185-189, 205-209, 211-214, 216-217, 244, 246, 273-276, 279, 297, 332, 334-336, 342, 345-346, 373, 390, 411, 413-427, 429-430, 465-470, 473-474, 504-537
S51T	2, 14, 16-22, 24-28, 30-32, 35-37, 47, 51, 57, 59, 70, 86, 109-110, 133-135, 141, 163-164, 169-170, 244, 246, 276, 335-336, 418-421, 423-425, 429, 466, 505-537
C53S	5-6, 10-17, 19-34, 36, 51, 57, 64, 67-68, 70-71, 73, 110, 133-137, 141, 155, 163-164, 195, 243-244, 276, 332-333, 335-339, 411-424, 466, 496, 504-537
G57R	8-9, 11-12, 14, 18, 20, 26-27, 30, 32-38, 40-43, 45, 51, 59, 106-113, 133, 135-137, 141, 145, 158, 163, 166, 169-170, 174, 178, 206, 244, 273-276, 279, 297, 336, 408, 411-428, 430, 466, 469, 505-537
Y58C	2, 11-12, 14, 16, 18-21, 23-24, 26-37, 45-48, 50-51, 58-60, 64, 70-72, 81, 86-87, 107-110, 134, 136-137, 141, 153, 163-164, 168, 170-173, 175-180, 182, 186, 194, 239, 242-244, 246, 274-276, 336, 417-424, 464, 466-468, 470, 497-537
T61I	2, 14-21, 25-28, 30, 32-33, 35-37, 42-43, 46-48, 51, 57-60, 86, 107-112, 133-135, 141, 163-164, 168-170, 206, 244, 246, 273-274, 276, 335, 414-425, 466, 502, 504-537
H69Y	5-6, 10-12, 14-21, 23, 27, 33, 36-38, 40-52, 56-74, 86, 88, 104-114, 133, 141-145, 147-148, 151-153, 163-164, 166, 168-170, 178, 182, 185, 205-207, 211-213, 217, 244, 246, 273-276, 279, 297, 332, 334-337, 390, 407-408, 410-430, 466, 469-470, 499, 502, 505-537
A75T	2, 5-6, 10-18, 20-42, 86, 108-111, 133-136, 141, 145, 151-156, 158, 160, 163-164, 168-170, 176, 206, 242-248, 250, 276, 332, 335, 338, 342, 344-346, 362, 392, 415-425, 465-470, 497, 499, 502-537
A75G	3-8, 10, 14-38, 40-43, 49-52, 57, 63-64, 66-74, 76-77, 108-113, 118, 133-137, 140-142, 145-149, 151-152, 154-158, 160-161, 163, 166, 176, 195-196, 208-209, 240, 243-244, 275-276, 279, 294-297, 299, 363, 390, 411, 414-424, 466, 496, 498, 501-502, 504-537

C90R	4-6, 11-12, 14-21, 23-34, 36-38, 41-44, 47-52, 57, 59, 63-64, 66-77, 102, 109-110, 125, 133-142, 144-149, 155, 158, 163, 166, 174, 240-245, 247, 276, 297, 357, 385, 388, 390, 394, 400-401, 404, 407, 417-424, 432, 452, 455-457, 459-464, 466-467, 471-473, 476, 496, 498-537
Q95E	4-6, 11-17, 20-21, 26-27, 31-34, 36, 41-52, 57-63, 65-77, 102, 107-110, 112, 125, 134, 136-137, 140-142, 146-148, 163-164, 176, 186, 195, 240, 242-247, 276, 296-297, 330, 332-333, 335-337, 339, 341, 362, 364, 408, 411-412, 414-419, 421-425, 466-467, 498-537
M105V	3-8, 10, 14, 16-38, 40-44, 47, 49-52, 57, 63-77, 80, 109-113, 118, 133-138, 140-142, 145-158, 160-161, 163, 166, 169, 186, 188, 208-209, 240-244, 295, 297, 390, 408, 411-424, 496, 498-537
M105T	3-6, 8, 10-44, 47-58, 63-81, 102, 109-113, 118-119, 122, 125, 133-142, 144-161, 163-166, 168, 186, 208-209, 239-244, 249, 275-276, 279, 295-299, 349, 351, 360, 363, 390, 400, 408, 411-424, 432, 460, 466, 496-537
C106S	2-10, 12-20, 26-27, 30, 32-51, 57-70, 86, 88, 104-114, 132-139, 141-145, 147-149, 151-153, 158, 163-164, 168-170, 178, 182, 185-192, 195, 205-207, 244, 246, 273-276, 279, 284, 289, 295-297, 332, 334-338, 342, 345-346, 390, 407-408, 410-430, 497, 505-537
I114N	4-6, 11-17, 19-21, 26-27, 30-37, 39, 41-51, 57-59, 61-64, 66-77, 106-110, 112, 125, 133-137, 139-142, 146, 163-164, 174, 176, 183, 186, 195, 201, 204, 242-244, 247, 276, 333, 335-337, 339, 383, 417-418, 420-425, 466-468, 500-537
I114T	4-8, 10-12, 14-44, 47, 49-58, 63-80, 109-113, 118, 133-137, 139-142, 144-161, 153-164, 166, 168, 186, 208-209, 240-245, 276, 295, 297, 339, 351, 390, 400, 408, 411-424, 496-537
I114S	2, 4-6, 8, 10-17, 20, 26-27, 30-51, 57-76, 102-115, 118, 125-126, 132-134, 136-151, 160-161, 163, 169-170, 174, 176, 186-188, 195, 201, 204, 208-209, 240-249, 277, 284-285, 296-297, 330-336, 339-345, 351, 362, 365-366, 411-412, 414-426, 428, 466, 495-537
G115V	2, 5-6, 15-33, 36-37, 45-48, 50-52, 57-60, 67-69, 70, 86, 108-111, 133-137, 141, 153, 155, 158-161, 163, 176-177, 206-207, 209, 211-214, 217, 244, 273-274, 276, 332-333, 337-337, 339, 390, 411-430, 457, 460-475, 478, 504-537
C117R	1, 3-6, 10-12, 15-28, 30, 33-34, 36-38, 41-51, 57, 63-77, 102, 109-110, 112-113, 118-119, 125, 134-142, 144-155, 158, 163, 166, 174, 239-245, 276, 297, 335, 357, 360, 363-364, 382-383, 385, 390, 417, 412-424, 463-464, 466-468, 496-537
G119S	5-6, 10-21, 24-33, 36-38, 41-43, 45-48, 51, 57, 59-60, 86, 107-113, 133-135, 141, 163-164, 168-170, 178, 206, 244, 273-276, 279, 297, 336, 407-408, 411-412, 414-415, 417-419, 421-424, 496-497, 501-502, 505-537
C128R	1-2, 8-12, 14, 16-21, 24-36, 38-39, 43-51, 57-61, 64, 70-74, 76-77, 81, 86-87, 107-111, 125, 132, 134, 136-138, 141, 153, 160, 163-164, 168-170, 172-173, 176-178, 180, 182, 186, 194-195, 210, 244, 246, 252-253, 273-274, 276, 335, 345-346, 361-362, 365, 415, 418-424, 464-470, 473, 497-499, 501-537
N152K	2, 14-21, 25-28, 30-33, 35-37, 48, 51, 59, 86, 108-111, 141, 163-164, 169-170, 178, 206, 211-212, 244, 273-276, 335, 411, 414-423, 425, 464-466, 468-470, 504-537
M157V	4-8, 10-59, 63-80, 109-113, 177-179, 133-142, 144-166, 186, 188, 240-244, 273, 276, 297, 335, 339, 390, 408, 411-412, 414-423, 466, 496, 498, 500-537
M157K	14-33, 36-37, 51, 57, 59, 70, 86, 108-111, 133-135, 141, 163-164, 169-170, 206, 243-244, 273-277, 332, 335-337, 411, 415-416, 418-421, 423-425, 466, 501-537
M157T	4, 6, 8, 10-12, 15, 17, 20-37, 42, 47, 50-52, 57-59, 68, 70-71, 73, 77, 108-111, 118, 133-141, 144-145, 147, 151-155, 158-159, 163-164, 186, 244, 273-276, 279, 295, 297, 336, 411, 414-424, 466, 505-537
M159T	4-5, 10-12, 15, 17, 20-21, 25-27, 30, 32-34, 36-37, 42, 109-112, 133-135, 141, 158, 160-161, 163, 169-170, 174, 195-196, 240, 243-244, 276, 297, 336, 415, 417-426, 466, 496, 498-537

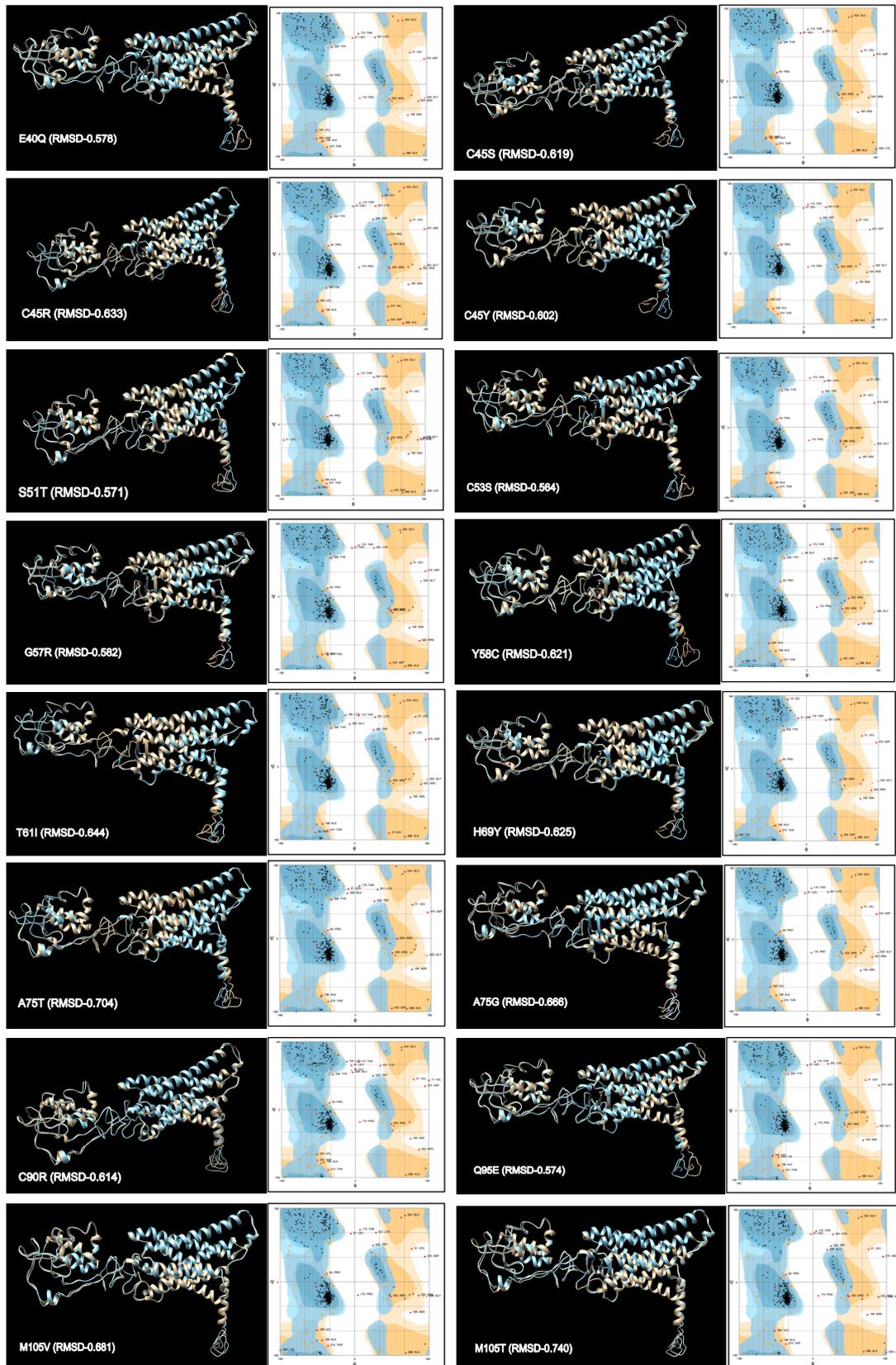
M159G	2, 5-6, 8-9, 14-44, 48, 50-52, 56-60, 68, 80, 105-114, 133, 137, 141, 145, 147, 149, 151-155, 157-159, 161, 163-164, 166, 178, 186, 188-189, 205-209, 211-214, 217, 240, 243-244, 273-276, 279, 284, 289, 295, 297, 331-339, 372, 390, 404, 407-408, 411-430, 466, 469, 496-537
E160Q	5-7, 14-27, 29-34, 36, 59, 70, 109-110, 112-113, 132-133, 137, 141, 163-164, 169-170, 176, 244, 273-275, 279, 284, 297, 335-336, 417, 421-423, 466, 496, 498, 502-537
P168S	4, 10-38, 40-58, 63-80, 109-113, 118-119, 133-142, 144-159, 161, 163-166, 168-169, 186, 240, 242-244, 276, 297, 335, 339, 408, 411-412, 414-424, 466, 496, 498-537
E180K	4-6, 11-16, 20, 27-28, 30, 32-34, 36, 46-51, 57-59, 63-64, 66-74, 76-77, 109-110, 112, 125, 134, 104-142, 146, 163-164, 177-178, 243-245, 247, 276, 297, 333, 335-336, 339, 415, 418-419, 421-425, 428, 466-467, 496, 498, 500-537
C181R	6, 8-14, 16-21, 26-38, 40-51, 57-62, 70, 105-114, 133-136, 141, 145, 158, 163-164, 186, 206, 211-213, 244, 273-276, 279, 297, 411, 414-428, 430, 465-466, 468-470, 499, 502, 504-537
C181Y	4, 11-12, 15-16, 18-21, 26-27, 30-36, 48-49, 59, 70, 109-110, 112, 133, 135-136, 141, 163-164, 169-170, 243-244, 273-275, 297, 335-336, 417-424, 432, 466, 496, 502, 505-537
C204R	1-2, 4-8, 10-11, 15-51, 59-65, 67-82, 84, 86-87, 104-114, 132, 134-137, 140-143, 146-148, 150-156, 163-164, 167-170, 172-173, 176-183, 185-187, 190, 193-197, 200, 202, 205-206, 209-212, 214, 242, 244-248, 250, 276, 286, 291-292, 295, 297, 335-336, 342, 345, 361-361, 364-366, 372, 381-389, 392, 417, 419-426, 428-429, 453, 456-471, 473-475, 478, 497-537
C204Y	13-21, 26-27, 31-34, 108-111, 133, 135-136, 141, 160-161, 163-164, 243-244, 273, 276, 336, 418-419, 421-423, 466, 501-502, 505-537
C204F	1-2, 6-7, 10, 14-34, 36-39, 41-52, 57-62, 64, 67-77, 80-81, 84, 86-87, 105-112, 134-135, 141, 153, 163-164, 167-170, 172-173, 176-178, 182, 185-186, 195, 244, 273-276, 288, 291, 297, 335-336, 345, 362, 416-425, 428-429, 460-461, 463-471, 473, 475, 504-537
M223K	5-6, 11-12, 14, 16-17, 20-21, 23-36, 45-51, 57-60, 64, 68, 70-77, 81, 86, 107, 110, 125, 134-137, 141, 160, 163-164, 172-173, 176-177, 186, 244, 252, 274-276, 336, 418, 421, 423-424, 466, 504-537
W226R	2, 5-6, 15, 18-39, 81, 86, 109-110, 125, 132, 134-137, 141, 153, 163-164, 172-173, 176-178, 195, 244, 246, 273, 335, 362, 365, 418, 421-423, 466, 505-537
T234I	4, 6, 8, 10, 15-30, 32, 34, 36-38, 42, 47-52, 57-59, 63-64, 66-77, 109-111, 118, 133-142, 145-146, 149, 152, 154-155, 158-159, 163, 166, 169-170, 243-244, 276, 279, 295, 297, 360, 364, 417-419, 421-423, 466-467, 496, 498, 501, 503-504, 507-537
T237R	2, 4, 10-16, 18-21, 23-24, 26-37, 39, 41-51, 57-59, 61, 63-64, 67-73, 86-87, 106-112, 125, 132-134, 136-137, 140-142, 146-147, 163, 169-170, 174, 176-177, 186-187, 194-195, 243-244, 276, 297, 333, 335-336, 415-430, 466-468, 498-505, 508-537
L239P	4, 6, 8, 10-12, 14-38, 40-42, 49-52, 57, 63, 66-74, 76-77, 80, 109-112, 118, 133-142, 145-149, 151-159, 163, 166, 169-170, 239-244, 276, 297, 390, 404, 407-408, 411-412, 414-424, 466, 496-537
Y250C	4, 8, 10, 13-16, 18-21, 26-28, 30, 32-37, 51, 55, 70, 86, 109-110, 125, 132-133, 137, 141, 163-164, 169-170, 176-177, 195, 243-244, 333-336, 417-432, 466, 496, 501-502, 505-506, 509, 511-537
R253C	4-6, 10, 13-14, 16, 18, 20-38, 42, 51, 64, 68, 70-71, 73, 109-111, 133, 141, 146, 154-155, 158, 163, 169-170, 274, 243-244, 252-253, 274-276, 297, 411-424, 466, 504-537
R253H	3-8, 10-44, 47, 49-52, 56-57, 63-64, 66-77, 80, 108-113, 118, 133, 135-137, 140-142, 145-159, 163, 166, 169-170, 239-244, 249, 274-276, 295, 297, 335, 363-364, 390, 408, 411, 414-424, 466, 496, 498, 500-537
L273R	3-6, 10-17, 20-30, 32-33, 36-37, 42, 50-51, 57, 59, 66-68, 70-71, 73, 76-77, 80, 86, 108-112, 133-141, 145, 151, 158, 163, 166, 169-170, 178, 243-244, 252, 275-276, 297, 336, 390, 411-412, 414-415, 417-424, 466, 496, 505-537

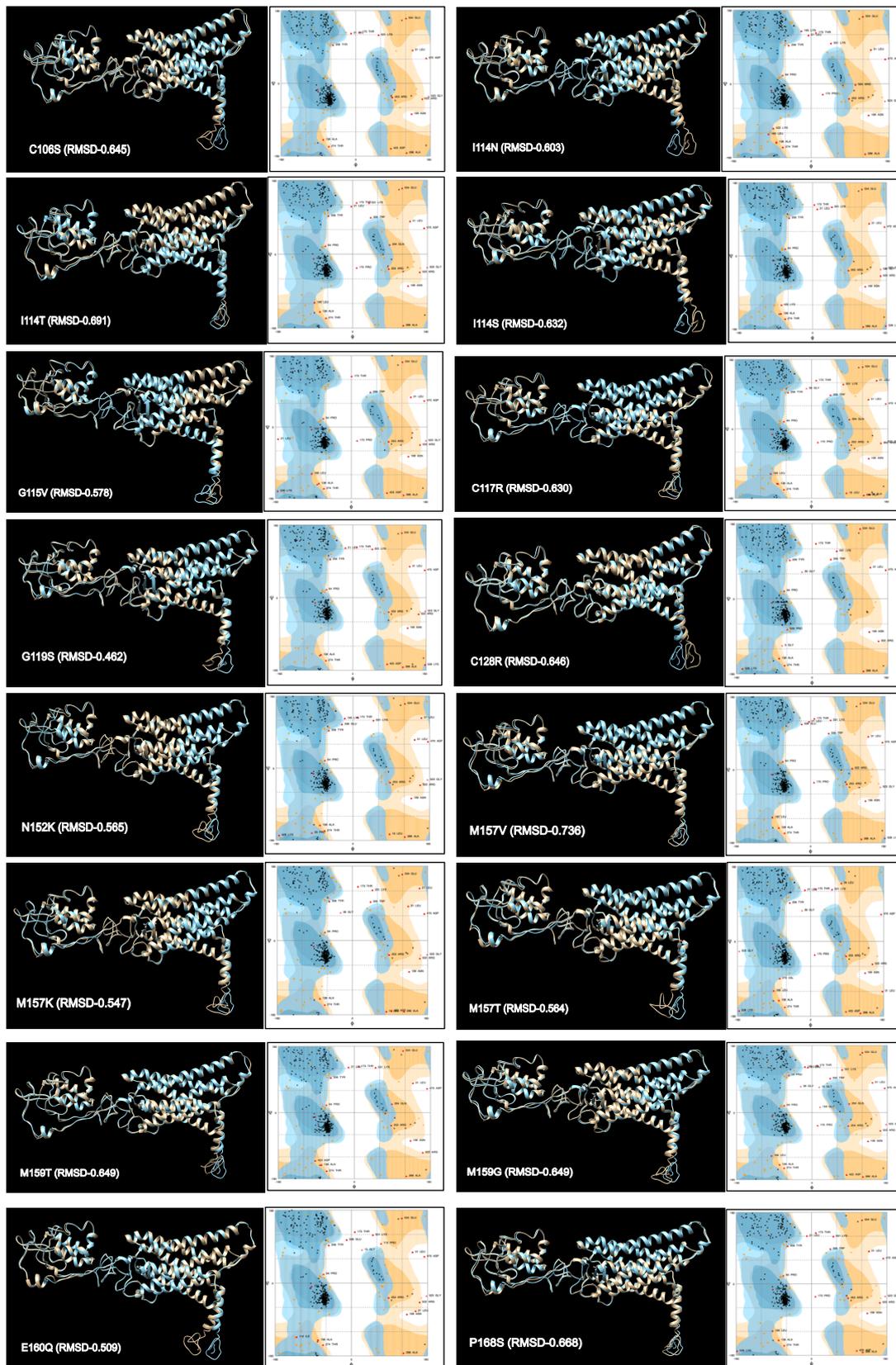
C302Y	14, 16, 18, 20, 24-33, 43-53, 55-64, 66-80, 86, 102-103, 105-110, 134-137, 139-142, 146, 160-161, 163-166, 169-170, 186, 195, 243-244, 276, 418, 421-423, 466-467, 498-537
F328S	2, 5-6, 10-12, 15-18, 20-37, 44, 46-52, 57-61, 64, 70-77, 81, 86, 106-111, 130, 133-137, 141, 160, 163-164, 172-173, 176-177, 195, 208-209, 242-248, 250, 276, 291, 335, 337, 362, 365, 418, 422-424, 464-468, 470, 498-499, 502, 505, 508-537
W335C	1-2, 4-6, 11-14, 16, 20-21, 25-28, 30, 32-53, 57-77, 102, 106-113, 125, 134-142, 145-151, 163-164, 166, 169, 174, 176, 183, 186-187, 195, 201, 204, 231-247, 276, 297, 330, 332-337, 339, 341, 383, 411, 416, 418-419, 421-425, 466-467, 495-496, 498-537
A339T	4, 8-9, 11-21, 26-28, 30, 32-37, 39, 64, 109-112, 125, 132-133, 137, 141, 163-164, 172-173, 176-178, 180-183, 186, 195-196, 244, 246, 333-337, 416-432, 466-467, 506-537
M342V	4, 6-8, 10-38, 40-45, 47, 49-51, 57, 63-64, 66-74, 76-77, 109-113, 118, 133-138, 140-142, 145-147, 149, 151-158, 163-164, 166, 169-170, 239-245, 273, 276, 297, 363, 390, 411, 414-423, 466, 496-537
K358N	4, 6-8, 10, 14-38, 40-44, 47, 49-51, 57, 63-64, 66-74, 76-77, 108-113, 118, 125, 133, 135-137, 141, 145-147, 149, 151, 154-159, 163, 166, 169-170, 239-247, 276, 297, 360, 364, 390, 411, 414-424, 466, 496-537
I360F	4, 8, 10-17, 20-34, 36-43, 51, 57, 59, 66-70, 77, 86, 88, 107-113, 118, 133-141, 144-145, 147-149, 151-152, 154-155, 158-159, 163-164, 169-171, 174-178, 240, 243-244, 252-253, 273, 276, 279, 297, 332, 390, 400, 404, 407-408, 410-430, 466, 496, 501-502, 504-537
R417Q	4-6, 11-14, 16-22, 25-30, 32-52, 57-64, 66-77, 102, 106-113, 125, 133-137, 140-142, 145-151, 155, 163, 174, 176, 183, 186, 195, 239-249, 297, 330, 332-344, 408, 411-412, 414-416, 421-425, 466-467, 495-537
I437T	6, 8, 10-12, 15-17, 20-42, 46-53, 56-60, 67-68, 70-74, 76-78, 80, 107-113, 117-119, 122, 130, 133-141, 144-161, 163-164, 166, 177, 186, 205-206, 208-209, 214, 244, 273-276, 279, 284, 295, 297-299, 336, 390, 400-401, 403-424, 466, 505-506, 508-537
V442E	2, 5-6, 10-12, 14, 16-19, 20-36, 38-39, 46-48, 50-51, 57, 59, 67-68, 86, 107-111, 134-137, 141, 152-156, 160, 163-164, 169-170, 242-248, 273-277, 336, 345, 362, 415-415, 464-470, 473, 504, 506-537
T445P	2, 4-6, 10-16, 18-21, 26-27, 30-51, 57-59, 63, 65-73, 106-113, 125, 132-134, 140-142, 145-149, 151, 155, 163-164, 169-170, 174, 176, 183, 186-187, 195, 201, 204, 240, 242-245, 247, 273-275, 284-285, 296-297, 330, 332-339, 365, 411, 418, 411-424, 428, 432, 466-467, 496, 498, 500-537
D470N	3-10, 12-21, 26, 30, 32-52, 56-70, 86, 88, 101, 104-115, 133-139, 141-145, 147-153, 158-159, 163-164, 168-170, 182, 185-192, 205-209, 211-214, 216-218, 244, 273-277, 279, 284-285, 289, 295-299, 329, 332, 334-338, 342, 345-347, 372-373, 382, 390, 403-432, 435, 462-463, 465-466, 468-471, 474, 505-537
G488D	4, 10-17, 20-21, 23-24, 26-28, 30-51, 57-64, 66-74, 86-87, 106-110, 112-113, 125, 132-134, 136-138, 140-142, 145-149, 160-161, 163-164, 166, 169-170, 174, 176-178, 183, 186-189, 194-195, 201, 243-244, 276, 330-336, 339, 412, 415-432, 464, 466-468, 496, 498-502, 508-537
G488V	2, 5-6, 10, 15-34, 36-50, 59-62, 64, 67, 69-74, 78, 80-81, 86, 107-111, 134, 137, 141-142, 147-148, 151-155, 160-161, 163-164, 168-170, 172, 176-178, 182, 194, 206-207, 209, 211-213, 217, 244, 273, 276, 332, 335-336, 392, 417, 421-425, 428-429, 463-471, 473, 478, 507-537
G492R	4, 8-9, 11-12, 15-17, 19-43, 46-59, 63-81, 83, 86, 102, 109-113, 133-159, 161, 163-166, 169-170, 239-244, 247, 295, 297, 332, 336, 351, 356-357, 360, 363-364, 390, 400-401, 404, 407-408, 411, 417-418, 421-423, 463-464, 466-467, 496-537
S497F	2, 8-9, 11-12, 14, 20-21, 23-24, 26-34, 36, 39, 43-48, 50-51, 57-59, 64, 70-72, 81, 86-87, 107-111, 134, 137, 141, 153, 160, 163-164, 168-170, 172-173, 175-178, 180, 182, 185-186, 194, 242, 244, 246, 276, 291, 336, 345, 415-416, 418-419, 421, 423-425, 465-468, 470, 473, 499, 502-537

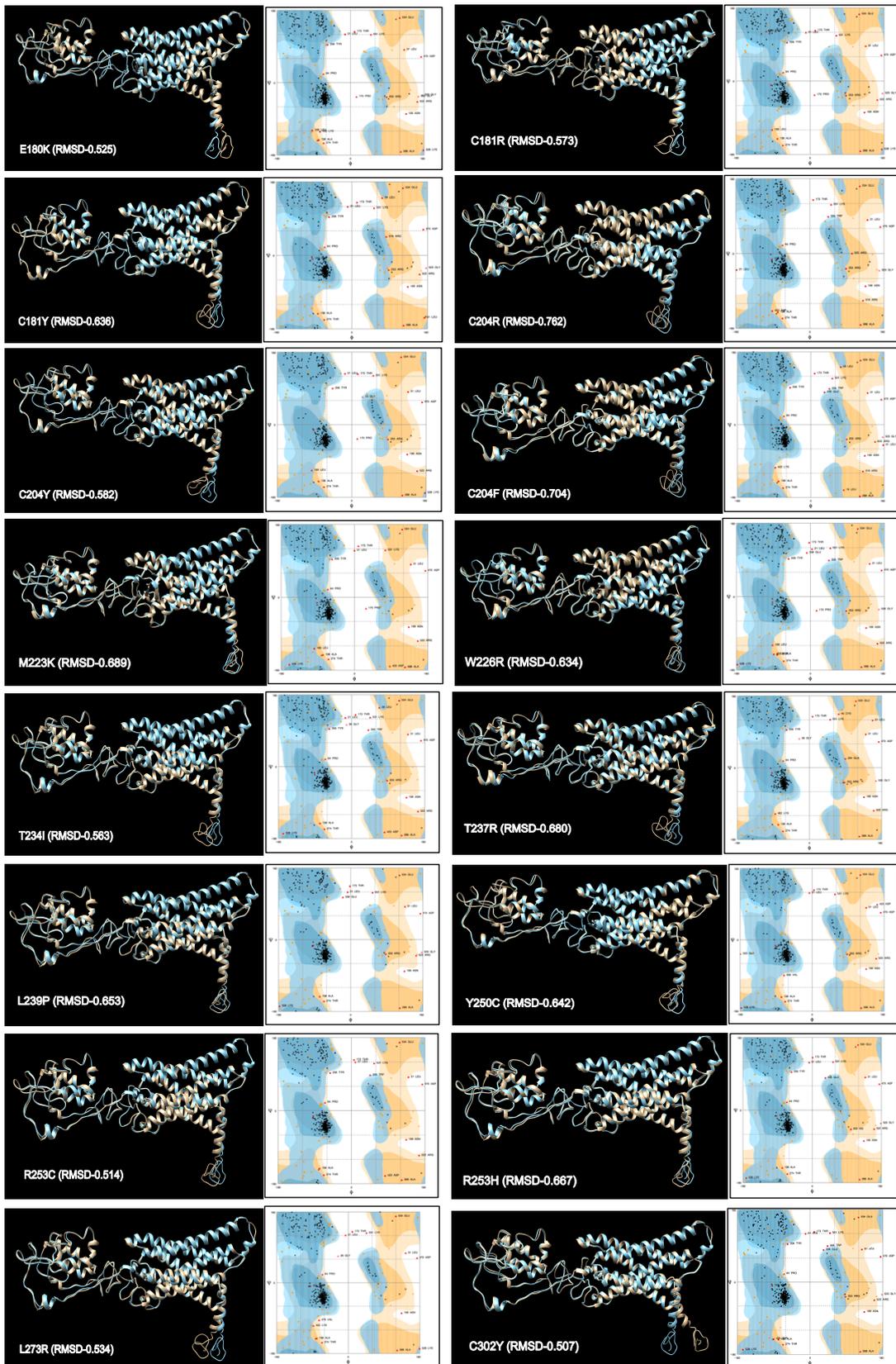
K499E	2, 4-10, 12-14, 16-20, 26, 30, 31-51, 57-70, 86, 88, 104-115, 133, 137-138, 141-145, 147-153, 158, 163-164, 168-170, 174, 178, 182, 185-191, 202, 205-207, 217, 244, 273-276, 279, 284, 286, 289, 295-297, 332, 334-339, 342, 346, 372, 390, 407-408, 410-430, 466, 469, 497-499, 502, 505-537
G525R	2, 8-21, 23, 26-27, 30, 36-38, 40-51, 57-62, 64-65, 67-68, 70, 86, 106-113, 133, 137, 141, 143-145, 147-148, 152-153, 160-161, 163-164, 168-170, 172, 176-178, 206-207, 211-214, 217, 244, 246, 273-276, 336, 342, 345-346, 411, 414-424, 464-470, 473, 505-537
G530E	4-6, 10-12, 14-30, 33-34, 36-38, 41-47, 49-51, 57, 63-64, 66-77, 80, 109-110, 112, 130-142, 145-149, 151, 154-155, 158, 163, 166, 169-170, 174, 176, 186, 209-210, 240, 242-246, 276, 297, 335-340, 390, 411, 417, 421-424, 460-471, 496, 498, 500-537

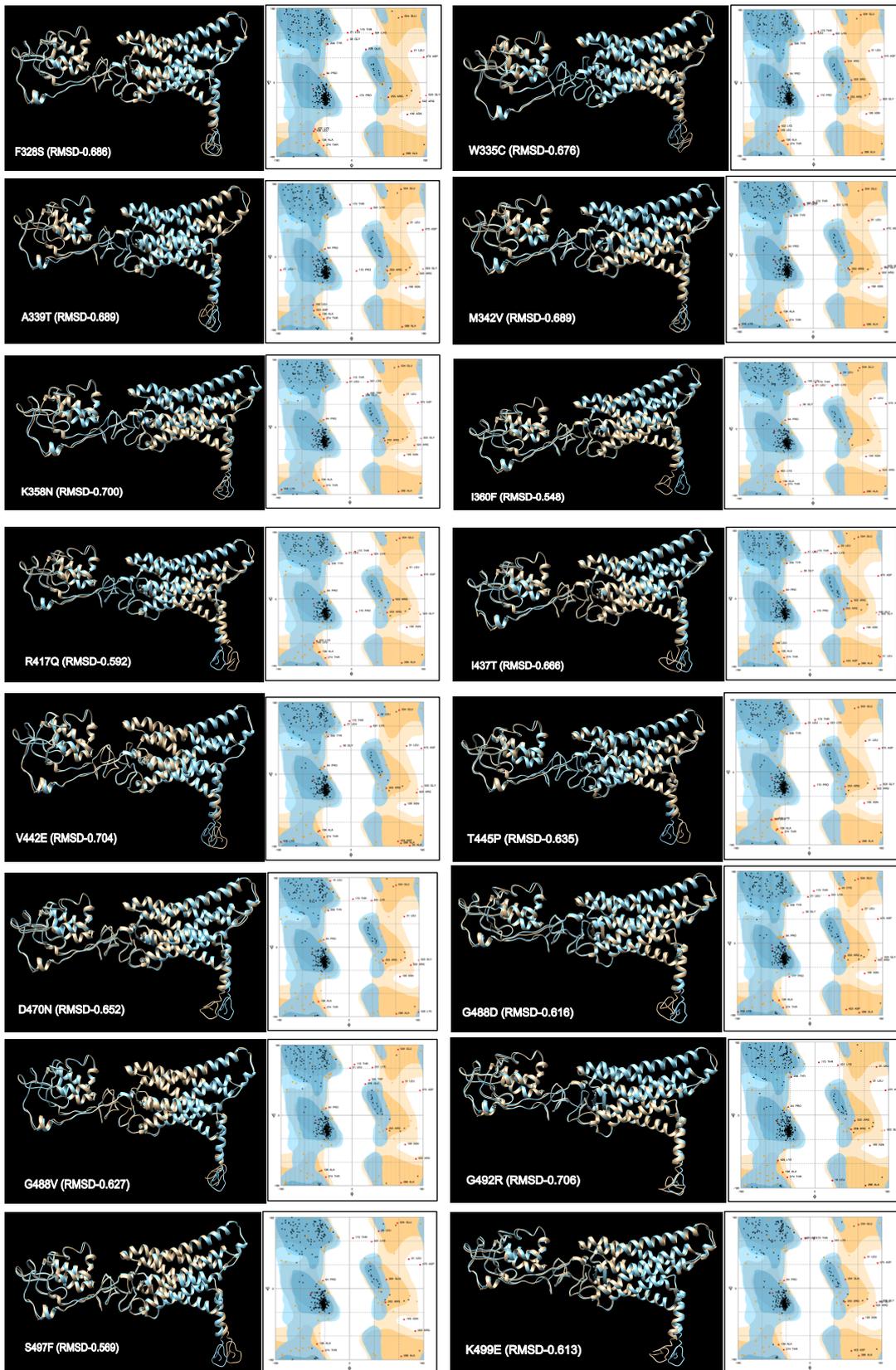
^(a) The column indicated the deviation of RMSD backbone score found in the amino acid position of the residue in the mutant protein structure with its wildtype FZD4 structure.

Appendix 1.









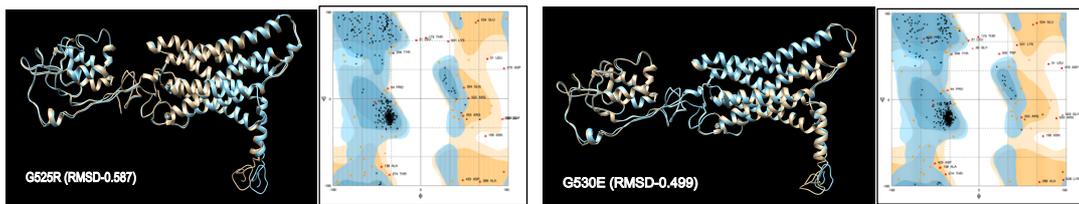


Figure A 1. All mutant structures of FZD4 superimposing with wildtype model in FEVR, and the Ramachandran plot of the mutant model of FZD4.

Appendix 2.

The following link 1 was the Excel file for all FEVR variants and cases.

[All FEVR cases \(cases reported from the same lab were removed\).xlsx](#)

The link 2 was the Excel file for all the unpublished FEVR variants data.

[unpublished FEVR variant overview.xlsx](#)

Appendix 3.

The script used for the mutant protein modelling in the Modeller9.24 software were based on the basic modelling script which could obtain from the Modeller website (<https://salilab.org/modeller/tutorial/basic.html>). The example of the script used of modelling the p.G22E mutant structure by using FZD4 wildtype protein structure as the reference in the Modeller9.24 was given below (**Figure A2**).

The target sequence of p.G22E of *FZD4* was prepared in the FASTA format, and named as qseq.ali (**Figure A2 (A)**.) Build_profile.py basic script was downloaded from the Modeller website. Changed the target sequence file name as qseq.ali (**Figure A2 (B)**.) Downloaded align.py script from Modeller website, and changed the template as the wildtype FZD4 protein structure (generated from I-TASSER in the study) and changed the name of the protein structure into the wildtype FZD4.pdb file name (**Figure A3 (C)**.) Downloaded model-single.py script from Modeller website, and modelling p.G22E by using wildtype FZD4 structure as the reference (**Figure A3 (D)**.) Selected the best model based on the lowest DOPE score that generated in the modelling process, and downloaded model_evaluation.py script from the Modeller website to evaluate the modelling quality of the best structure (**Figure A3 (E)**.)

```
(A) >P1;qseq
Sequence:qseq::::::::::0.00: 0.00
MAWRGAGPSVPGAPGGVGLSLELLLQLLLLLLPARGFGDEEERRCDPIRISMCQNLGYNV
TKMPNLVGHELQTD AELQLTTFTPLIQYGCSSQLQFFLCSVYVPMCTEKINIPIGPCGGM
CLSVKRRCEPVLKEFGFAWPESLNCSKFPPQNDHNHMCMEGPGDEEVPLPHKTPIQPGEE
CHSVGTNSDQYIWKRLNLCVLCGYDAGLYSRSAKEFTDIWMAVWASLCFISTAFTVLT
FLIDSSRFSYPERPIIFLSMCYNIYSIAYIVRLTVGRERISCDFEEAAEPVLIQEGLKNT
GCAIIFLLMYFFGMASIIWVILTTLTWFLAAGLKWGHEAIEMHSSYFHIAAWAIPAVKTI
VILIMRLVDADEL TGLCYVGNQNLDAL TGFVVAPLFTYLVIGTLFIAAGLVALFKIRSNL
QKDGTKTDKLERLMVKIGVFSVLYTVPATCVIACYFYEISNWALFRYSADDSNMAVEMLK
IFMSLLVGITSGMWIWSAKTLHTWQKCSNRLVNSGKVKREKRGNGWVKPGKGETVV*
```

```
(B) from modeller import *
log.verbose()
env = environ()

#-- Prepare the input files

#-- Read in the sequence database
sdb = sequence_db(env)
sdb.read(seq_database_file='pdb_95.pir', seq_database_format='PIR',
         chains_list='ALL', minmax_db_seq_len=(30, 4000), clean_sequences=True)

#-- Write the sequence database in binary form
sdb.write(seq_database_file='pdb_95.bin', seq_database_format='BINARY',
         chains_list='ALL')

#-- Now, read in the binary database
sdb.read(seq_database_file='pdb_95.bin', seq_database_format='BINARY',
         chains_list='ALL')

#-- Read in the target sequence/alignment
aln = alignment(env)
aln.append(file='qseq.ali', alignment_format='PIR', align_codes='ALL')

#-- Convert the input sequence/alignment into
#   profile format
prf = aln.to_profile()

#-- Scan sequence database to pick up homologous sequences
prf.build(sdb, matrix_offset=-450, rr_file='${LIB}/blosum62.sim.mat',
         gap_penalties_1d=(-500, -50), n_prof_iterations=1,
         check_profile=False, max_aln_evalue=0.01)

#-- Write out the profile in text format
prf.write(file='build_profile.prf', profile_format='TEXT')

#-- Convert the profile back to alignment format
aln = prf.to_alignment()

#-- Write out the alignment file
aln.write(file='build_profile.ali', alignment_format='PIR')
```

```
(C) from modeller import *
env = environ()
aln = alignment(env)
mdl = model(env, file='FZD4', model_segment=('FIRST:A','LAST:A'))
aln.append_model(mdl, align_codes='FZD4', atom_files='FZD4.pdb')
aln.append(file='qseq.ali', align_codes='qseq')
aln.align2d()
aln.write(file='qseq-FZD4.ali', alignment_format='PIR')
aln.write(file='qseq-FZD4.pap', alignment_format='PAP')
```

```
(D) from modeller import *
from modeller.automodel import *
#from modeller import soap_protein_od

env = environ()
a = automodel(env, alnfile='qseq-FZD4.ali',
             knowns='FZD4', sequence='qseq ',
             assess_methods=(assess.DOPE,
                            #soap_protein_od.Scorer(),
                            assess.GA341))

a.starting_model = 1
a.ending_model = 5
a.make()
```

```
(E) from modeller import *
from modeller.scripts import complete_pdb

log.verbose() # request verbose output
env = environ()
env.libs.topology.read(file='${LIB}/top_heav.lib') # read topology
env.libs.parameters.read(file='${LIB}/par.lib') # read parameters

# read model file
mdl = complete_pdb(env, 'qseq .B99990001.pdb')

# Assess with DOPE:
s = selection(mdl) # all atom selection
s.assess_dope(output='ENERGY_PROFILE NO_REPORT', file='qseq.profile',
              normalize_profile=True, smoothing_window=15)
```

Figure A 2. Example of the screenshots of the scripts used in Modeller9.24 for modelling p.G22E by using wildtype FZD4 structure as the reference. **(A).** The sequence format of p.G22E mutant of FZD4. **(B).** Screenshot of the script of Biuld_profile.py file used in modelling p.G22E. **(C).** Screenshot of the script of Align.py file used in aligning the target p.G22E sequence on wildtype FZD4 structure. **(D).** Screenshot of the script of model.py used in generating the models in Modeller9.24. **(E).** Screenshot of the script of evaluate_model.py used in evaluate the mutant model quality.