TITLE PAGE

Exome sequencing identifies variants in *FKBP4* that are associated with recurrent fetal loss in humans

Authors: Charalambos Demetriou[1*], Estelle Chanudet[2], GOSgene[2], Agnel Joseph[3], Maya Topf[3], Anna C. Thomas[1], Maria Bitner-Glindzicz[1], Lesley Regan[4], Philip Stanier[1], Gudrun E. Moore[1]

Affiliations: [1]Genetics and Genomic Medicine, UCL Great Ormond Street Institute of Child Health, University College London, London, UK. [2]Centre for Translational Omics - GOSgene, UCL Great Ormond Street Institute of Child Health, University College London, London, UK. [3]Institute of Structural and Molecular Biology, Birkbeck College, London, UK. [4]Department of Obstetrics and Gynaecology, St. Mary's Campus, Imperial College London, London, UK.

Philip Stanier and Gudrun Moore contributed equally.

[*]**Corresponding author:**

Charalambos Demetriou

Address: UCL Great Ormond Street Institute of Child Health

30 Guilford Street, Lonodn, WC1N 1EH

Tel: 0044796 081 3545

Fax: 0044207 905 2832

Email address: c.demetriou@ucl.ac.uk

ABSTRACT

Recurrent pregnancy loss (RPL) is defined as two or more consecutive miscarriages and affects an estimated 1.5% of couples trying to conceive. RPL has been attributed to genetic, endocrine, immune and thrombophilic disorders, But many cases remain unexplained. We investigated a Bangladeshi family where the proband experienced 29 consecutive pregnancy losses with no successful pregnancies from three different marriages. Whole exome sequencing identified rare genetic variants in several candidate genes. These were further investigated in Asian and White European RPL cohorts, and in Bangladeshi controls. *FKBP4,* encoding the immunophilin FK506 binding protein 4, was identified as a plausible candidate, with three further novel variants identified in Asian patients. None were found in European patients or controls. *In silico* structural studies predicted damaging effects of the variants in the structure-function properties of the FKBP52 protein. These were located domains reported to be involved in Hsp90 binding and peptidyl-prolyl cic-trans isomerase (PPIase) activity. Profound effects on PPIase activity were demonstrated in transiently transfected HEK293 cells comparing wildtype and mutant *FKBP4* constructs. Mice lacking *Fkbp4* have been previously reported as infertile through implantation failure. This study therefore strongly implicates *FKBP4* as associated with fetal losses in humans, particularly in the Asian population.

INTRODUCTION

Miscarriage is the commonest complication of pregnancy and it is defined as the spontaneous loss of a fetus before it has reached viability[1]. Studies have shown that the risk of a pregnancy loss increases after each successive miscarriage and can reach as high as 45% after three consecutive losses[2]. A diagnosis of recurrent pregnancy loss (RPL) is usually considered following the loss of at least two pregnancies and has been attributed to many risk factors as parental age or lifestyle factors, as well as genetic causes such as chromosomal abnormalities, anatomical disorders including uterine abnormalities and endocrine defects[3]. Moreover, RPL can be caused by abnormalities in the immune system, for example the antiphospholipid syndrome (APLS)[3] or thrombophilic disorders leading to thrombosis, caused for example by the Factor V Leiden mutation (FVL)[3,4]. However, for many patients, the cause of RPL remains a clinical dilemma, since even after detailed investigation their miscarriages remain unexplained, which is traumatic for patients[3].

Here we describe a genomic approach to identify a new candidate gene for RPL. This study focused on a single family from Bangladesh where the patient suffered a total of 29 early RPLs, before the 12th week of gestation, all of whom were unexplained. We performed whole-exome sequencing in the family members, which led to the identification of two candidate genes with variants in the proband. These candidate genes *FKBP4* (encoding FK506 binding protein 4, also known as FKBP52) and *SERPINB2* (serpin peptidase inhibitor clade b2) were investigated in greater depth by sequencing them in cohorts of Asian RPL (n=100) and White European RPL (n=120) patients, all of which had no successful pregnancies, as well as in a normal Bangladeshi cohort. Three novel variants were identified in *FKBP4* in three Asian patients while no variants were identified in *SERPINB2*.

FKBP52 is common to several vertebrate species and the respective mRNA transcripts are expressed in a variety of human tissues such as brain, heart and placenta[5]. FKBP52 has been described as a co-chaperone for steroid hormone nuclear receptors and binds to the progesterone receptor (PR) to optimise progesterone-PR signalling. Therefore success of transcriptional activation depends on the binding of PR to FKBP52. Female mice lacking the *Fkbp4* gene have compromised PR functions leading to a total failure of the uterus to support blastocyst implantation[6-8], strongly supporting the candidacy of *FKBP4* for a role in RPL. We conclude that *FKBP4* is a novel candidate gene associated with RPL in humans and may be specifically associated with the Asian population.

RESULTS

The family pedigree

The patient (II-3) and her family (Fig. 1) originate from Bangladesh but live in London. The family is under the care of the Recurrent Miscarriage team at St Mary's Hospital, Imperial College London. This patient, who is now 46 years old, had a total of 29 pregnancy losses all of which were early, between 9-11 weeks of gestation, with no live births. She had 17 miscarriages with her first husband (II-1), five miscarriages with her second husband (II-2) and seven miscarriages with her third husband (II-4). All of these pregnancies were naturally conceived and the miscarriages were unexplained as the patient tested normal for APLS and lupus anticoagulant (LA) and negative for the FVL mutation. The patient's mother had three successful pregnancies and three late miscarriages all of which were in the second or third trimester. The patient's brothers (II-5 and II-6), each had two healthy children.

FKBP4 variants in recurrent miscarriage patients

We performed whole-exome sequencing (WES) in the proband as well as her third husband and her two brothers. Following comprehensive variant analyses and prioritization, two genes with novel variants present only in the proband were identified. The model we used was a dominant model that considered the RPL patient as an affected individual and her two brothers and husband as unaffected. WES data from four Asian females provided by Great Ormond Street Hospital's Centre for translational research (GOSgene) were used as additional unaffected controls in this analysis.

After excluding common variants that were present more than 1% in public databases and including variants that are predicted to be possibly pathogenic by *in silico* tools, two heterozygous variants were detected in the proband: a rare missense heterozygous change

c.47C>A (Ala16Glu) in *FKBP4* at position chr12:2904352 and a rare missense heterozygous change c.328C>G (Leu110Val) in *SERPINB2* at position chr18:61564364. Both variants were confirmed by Sanger sequencing. Segregation analysis was performed by PCR and Sanger sequencing in the parents. This revealed that both variants were inherited from the patient's father. Coding regions of these genes were completely sequenced in two larger RPL cohorts (100 Asians and 120 White-Europeans) that had no successful pregnancies, as well as a control cohort of 100 Bangladeshi mothers who had no previous history of miscarriage and at least one successful pregnancy. While no further missense variants present exclusively in patients were identified in *SERPINB2*, three more heterozygous missense variants were identified in *FKBP4* in the Asian RPL cohort, which were not present in the RPL White-Europeans or the Bangladeshi controls tested. The missense variants detected in three different Asian patients were: c.374A>G (Asn125Ser) at position chr12:29047018, c.1142A>T (Gln381Leu) at position chr12:2910392 and c.1196G>A (Arg399Gln) at position chr12:2910446. A list of all the variants identified in both *FKBP4* and *SERPINB2* in our control and RPL cohorts with population frequency data from gnomAD (v2.1.1) is shown in Supplementary Table 1. Overall, a statistically significant difference was observed between Asian RPL patients and Bangladeshi fertile controls (chi-squared test; p=0.04).

The pathogenicity scores of the variants identified in *FKBP4* was predicted using online tools (SIFT[10], Polyphen2[11], PROVEAN[12], nsSNPAnalyzer[13], Mutation Taster[14] and Variant Effect Predictor[15]). All the variants were predicted to be pathogenic by multiple programmes. Amino acid sequence conservation across different species was investigated using Clustal W2[16], a multiple sequence alignment program for DNA or proteins. This protein shows a high degree of conservation between species and the four variant residues are at highly conserved positions across mammals. Sequence chromatograms showing the four

variants in *FKBP4*, pathogenicity predictions of different online tools, as well as protein conservation are shown in Figure 2.

Structural and functional effects of FKBP52 missense variants

FKBP52 is a member of the immunophilin protein family FK506-binding proteins, which are based on their ability to bind to the immunosuppressive drug FK506 (Tacrolimus). Human FKBP52 is composed of four distinct and functional domains, which have been previously described[17-28]. A16E and N125S variants are located in Domain I, whereas Q381L and R399Q are located in Domain III (Fig. 3). Structural studies were undertaken to predict damaging effects in the structure-function properties of FKBP52 (Fig. 4).

Although A16E, was not amenable to accurate structural modelling due to the lack of known structural information for this region, the N125S mutation occurs in FK1, in a loop that is part of the peptidyl-prolyl cis-trans isomerase (PPIase) active site. This loop also binds rapamycin and FK506 by closing over the drug-binding pocket and makes multiple contacts with the drug. This loop has four conserved prolines and its conformation is likely to be important for PPIase activity, receptor binding and potentiation. Mutation of P119L in FKBP52 has previously been shown to dramatically reduce signaling compared to wild-type FKBP52[25] and mutation of FKBP12 amino acids 89 and 90 (equivalent to positions 120 and 121 in FKBP52) in the loop region, disrupted binding of FKBP12 to the TGF-β receptor. Further, deletion of the PPIase domain eliminates FKBP52's ability to enhance receptor function[25]. The drug FK506, which binds the PPIase active site and inhibits enzymatic activity, inhibits FKBP52-dependent potentiation of receptor function in yeast[26]. However, PPIase activity per se is not required for potentiation. By analogy to other well-characterized FKBPs, FK506 may sterically interfere with recognition of an FKBP52 binding partner[25]. It

has also been reported that FKBP52 interacts with the motor protein dynein through its PPIase domain[27].

It is proposed that FKBP52 FK1 forms a specific contact with the receptor ligand binding domain (LBD) of Heat shock protein 90 (Hsp90) heterocomplexes via the notch formed by P119 in the overhanging loop. This contact potentiates receptor responses to hormone[28]. Also, interaction with the LBD is likely weak, while Hsp90 interaction stabilizes a dynamic LBD conformational state that otherwise is transiently favorable for hormone binding. N125 forms a hydrogen bond with S115 and this interaction is not present in the rapamycin (and FK506) bound forms of FKBP52 suggesting that this interaction might help in mediating the conformational changes required in the loop for interaction with partner receptors and for PPIase activity (Fig. 4a).

The Q381 variant is located at the end of helix 6 while R399 is located in the helix 7 at the C-terminal of the tetratricopeptide (TPR) domain (Fig 4b). The core TPR domain and the helix H7 extension at least up to residue 414 were shown to be essential for Hsp90 binding and assembly into receptor complexes[28]. Despite retention of the core TPR domain, deletion of residues beyond 406 abrogates Hsp90 binding and PR association. Additionally, the final 30 amino acids of FKBP51 enhance binding to Hsp90, whereas the corresponding region of FKBP52 moderates binding to Hsp90. The H7 helix of FKBP52 shifts away by 30° from the Hsp90-binding interface relative to FKBP51. With respect to the R399Q variant, R399 is at the site where the helix 7 begins to shift away from the conformation in FKBP51, reflecting a probable role in maintaining the structure of this region. Both R399 and Q381 are found at the interface between different FKBP52 monomers in the crystal structure[29]. But there have been no clear experimental evidence to support the multimerization state of FKBP52 or its importance in its function. R399Q results in the loss of a positive charge at the

surface while Q381L replaces a polar side chain with a hydrophobic group potentially affecting the interactions mediated by these two residues (Fig. 4c).

The region of FKBP52 located between aa 267 and 400, which includes the TPR domain, is also required for tubulin binding[23]. FKBP52 can have multiple tubulin binding sites based on the observation that the deletion mutants with segments 1-375 and 375-458 both bind tubulin. FKBP52 prevents tubulin polymerization and the 84 residue sequence located in the C-terminal part of the molecule (aa 375–458) is necessary and sufficient for its microtubule de-polymerization activity, although with lower efficiency.

Next, the functional effects of each variant were tested in a transient overexpression assay measuring PPIase activity[9] (Fig. 5). Minimal endogenous PPIase activity was initially detected in HEK293 cells (data not shown). HEK293 cells were then transfected with wildtype (WT) and mutant FKBP52 constructs, each representing sequence variants identified in the RPL patients described above as well as empty vector an mock transfected controls. The WT construct showed a highly significant, 3-fold higher PPlase activity than empty vector or mock transfected cells ($p=<0.002$). The N125S mutant, located in the PPIase active site showed reduced activity compared to the controls which was significantly ($p=<0.0001$) less than the WT construct. The other three variants tested, A16E, Q381L and R399Q all showed similar activity to controls, again, all significantly ($p=<0.002$) reduced from the WT construct suggesting a deleterious effect of each variant.

DISCUSSION

Infertility and spontaneous pregnancy losses are sadly a common outcome for many women. The establishment of pregnancy depends on successful implantation, where a complex series of interactions occur between the uterus and blastocyst. Multiple early pregnancy loss is a significant clinical problem for women and their health care providers.

In this study, a woman with an unusually severe history of miscarriages was investigated. Exome sequencing was carried out and led to the identification of a missense variant in the *FKBP4* gene of the female patient, which was not present in the patient's husband, brothers or the four unrelated Asian female controls used in the prioritization. Segregation analysis was performed by PCR and Sanger sequencing in the patient's parents. This revealed that the variant was inherited from the father. This would be consistent with an inheritance model that suggests that the RPL effect is primarily expressed in the mother, perhaps creating an environment that is not compatible with implantation. A maternal effect is described when an offspring shows the phenotype expected from the genotype of the mother, irrespective of its own genotype. This can be due to the mother supplying gene products (mRNA or proteins) to the embryo[30]. Functional studies have reported the essential role of these products during gametogenesis and early embryonic development[31]. Loss of their function has no effect on male offspring or their fertility, or on the females offspring themselves, but females are sterile because their oocytes do not support early development. Maternal mRNAs accumulated in the oocyte play a critical role for successful development of the embryo and some of them are also required for embryo development after implantation[32].

Although *Fkbp4* heterozygous null animals were all fertile and without reproductive failures, both male and female homozygous mice, were reported to be infertile, highlighting the importance of FKBP52 in reproduction[6]. Interestingly, male null mice were found to

produce viable spermatozoa but had defects in reproductive tissues consistent with androgen insensitivity. Female null mice were anatomically normal but infertility was found to be a consequence of either implantation failure or pregnancy loss following implantation, which was associated with impaired progesterone function[7]. While this implantation failure phenotype was observed in both CD1 and C57BL6/129 knockout mice, daily progesterone supplementation rescued implantation with subsequent decidualization in CD1 *Fkbp4*$^{-/-}$ females but was ineffective in rescuing implantation in C57BL6/129 *Fkbp4*$^{-/-}$ females demonstrating the importance of genetic background[7.]

Further experiments showed that uterine levels of an antioxidant, peroxiredoxin-6 (PRDX6) were significantly lower in *Fkbp4*$^{-/-}$ mice and that these mice having reduced uterine PRDX6 levels were susceptible to oxidative stress leading to implantation failure even with progesterone supplementation[8]. FKBP52 may also promote endometriosis, as women with endometriosis show reduced FKBP52 expression and the progesterone resistance observed in the FKBP52 deficient mice results in increased inflammation, cell proliferation and angiogenesis which encourages the growth of endometriotic lesions. This suggests a vital role of FKBP52 in female reproduction and uterine signalling[33].

FKBP52 is a member of the immunophilin protein family, which is known for their roles in immunoregulation and basic cellular processes involving protein-protein interaction, protein folding and trafficking. FKBPs range in size from 12 to 135kDa[34] but all share a common domain for immunosuppressant ligand binding. However, they differ in their FK506 affinity, in other domain structures and their subcellular localization, suggesting that each FKBP may have a distinct function[23]. Human FKBP52 is composed of four distinct functional domains (shown in Fig. 3). The four variants identified in our study locate to Domains I and III, both of which are involved in Hsp90 binding. Domain I has PPIase activity, which is inhibited by FK506 and rapamycin, whose immunosuppressant drug action

is mediated partly by binding to immunophilins[5,35]. The N125S variant is located in the PPIase active site, suggesting a potentially useful functional assay for mutational effect. It was notable that this variant and indeed each of the other variants identified, all showed profound loss of activity, indicating either structural or conformational disturbance to the PPIase active site.

Domain III includes three TPRs which mediate protein-protein interactions involved in FKBP52-regulated cellular functions[36]. FKBP52 associates with Hsp90 in steroid receptor complexes, suggesting a role in the nuclear translocation of steroid hormone receptors such as the glucocorticoid receptor (GR). The hormone-free GR is localized in the cytoplasm and it translocates to the nucleus after steroid binding[37]. Studies suggest that the Hsp90-based chaperone system and the Hsp90-binidng FKBP52 are involved in the movement of the GR to the nucleus[38]. In fact, *in vivo* studies have shown that FKBP52 is incorporated into steroid receptor complexes by binding to Hsp90 via the TPR domain and it then facilitates translocation of the activated receptor to the nucleus via PPIase domain binding to dynein[18,39]. In addition, the importance of the role of *FKBP4* has been emphasized in the horse embryo[40], as well as in the human embryo, as FKBP52 protein levels were decreased in placenta of RPL patients suggesting an association with pregnancy loss[41].

The four *FKBP4* variants described here were all identified in RPL Asian patients but were not present in a Bangladeshi fertile control cohort. Neverthless, all four variants were present as rare variants in the ExAC/gnomAD databases, mainly in the South Asian population. The A16E mutation was heterozygous in four South Asian females, R399Q was heterozygous in six South Asian female and one homozygote, Q381L was heterozygous in two South Asian females, while N125S was heterozygous in one African and one Latino female. These databases contained data from more than 57,000 females but no information is available regarding the clinical status of individual variant carriers, so it is not feasible to

determine if any of these variant carrying females might have experienced miscarriages. There remains a possibility that this apparent population bias might suggest an Asian specific cause of RPL. We don't currently have any compelling explanation for this other than the fact that the particular variants identified do appear to be largely Asian-specific. The lack of likely-causal European variants identified in this study may simply reflect the relatively small size of the cohort investigated and suggests an urgent need for a larger study population to be investigated.

Our study identified missense variants at conserved residues within two functional domains of FKBP52. The variants were predicted to have damaging effects to structure-function properties and were shown to abrogate PPIase activity in a cell based assay. Coupled to evidence from mouse studies demonstrating the importance of Fkbp52 for embryo implantation, we suggest a potentially important role for this gene in the maintenance of pregnancy in humans, which requires further investigation. The nature of the immunophilin FK506 binding protein and the functional role of the encoded prolyl isomerase may present novel targets for therapeutic intervention in RPL.

MATERIALS AND METHODS

Ethics

DNA samples and medical records for this study were obtained from the Recurrent Miscarriage Clinic at St Mary's Hospital, part of Imperial College Healthcare NHS Trust, which contains blood samples from over seven thousand women who agreed to participate in research with signed, informed consent. The study was approved by the Imperial College Hospital Ethics Committee (REC ref EC0081).

Patient samples

Blood samples were available from the RPL patient (proband) who had 29 miscarriages, her third husband and her two brothers. Buccal swabs were collected from the patient's parents. The RPL cohort studied consisted of 100 Asian female patients and 120 White European female patients. These patients were idiopathic and to be included in the study must have had three or more miscarriages and no live births. The average age of the Asian and European patient cohorts at the time of sample collection was 33.2 ± 5 and 35.3 ± 4.6 years, respectively. DNA samples from 100 Bangladeshi controls had no history of RPL and had at least one live birth.

Whole-exome sequencing

Whole-exome sequencing was performed on DNA derived from peripheral blood with GOSgene (UCL genomic partner http://www.ucl.ac.uk/ich/services/lab-services/gosgene). Genomic DNA was subjected to array capture with the SureSelect Human Exon Kit v.4 (Agilent Technologies) according to the manufacturer's instructions. Adapters were ligated, and paired-end sequencing was performed on an Illumina HiSeq 2000, which generated 2 x 50 bp reads. The mean exome coverage was 81 to 91-fold, and 98% and 86% of the target

sequence was covered at least 10 and 30 times, respectively. Sequencing reads were aligned to the reference human genome (GRCh37/hg19, UCSC Genome Browser) with the BWA consensus, variant bases were called with GATK and variants were annotated with ANNOVAR.

Ingenuity Variant Analysis

Candidate gene variants were identified using Ingenuity Variant Analysis (Quiagen Bioinformatics). We applied a dominant model considering the RPL patient to be an affected individual with her two brothers and husband as unaffected. Recessive inheritance was considered unlikely due to the continued recurrence of miscarriages with two further, unrelated partners. Four Asian females were used as additional unaffected controls. The following filters were used to choose variants that were present in the patient but not present in the remaining samples: Common variants filter**:** excluded variants with an allele frequency of at least 1% in both the 1000 Genomes Projectand public Complete Genomics genomes. We considered as rare variants those with a frequency below 1% in public databases (ExAC, Exome Variant Server and 1000 Genomes); Predicted deleterious filter**:** kept variants that are experimentally observed to be pathogenic, likely pathogenic or uncertain significance; Genetic Analysis filter: kept variants that are homozygous, heterozygous compound heterozygous, haploinsufficient and occur in the case sample and excluded variants that occur in at least 1 of the control samples; Biological context filter**:** kept only variants that are known or predicted to affect genes implicated in the following diseases, processes, pathways or phenotypes: implantation, thrombosis or miscarriage.

Sanger sequencing

Sanger sequencing on *FKBP4* and *SERPINB2* was used to confirm WES variants and screen their entire coding sequences in 220 patients and 100 controls (see Supplementary

Table 2 for primers). Exons were amplified using Taq DNA polymerase (Bioline) and sequencing was performed using BigDye terminator v1.1 (Life technologies). Sequencing reactions were run on an ABI ABI 3730xl DNA Analyzer and the read-out was then analysed with Sequencher™ v4.8 (Gene Codes Corporation).

In silico prediction of variant pathogenicity

The pathogenicity scores of the *FKBP4* and *SERPINB2* variants identified was predicted using the online tools SIFT (http://sift.jcvi.org/), nsSNP Analyzer (http://snpanalyzer.uthsc.edu/), Polyphen2 (http://genetics.bwh.harvard. edu/pph2/), Mutation Taster (http://www.mutationtaster.org/), Variant Effect Predictor (http://www.ensembl.org/Tools/VEP) and PROVEAN (http://provean.jcvi.org/index.php), applied with standard procedures and settings.

Structure analysis and evolutionary conservation

UCSF Chimera was used for visualization and alignment of FKBP52 structures, and to carry out residue interaction studies. ConSurf server was used to calculate evolutionary conservation of amino acids and map the conservation scores onto the structures. The mapped scores were then used to color the structures in Chimera.

Constructs, Cell culture and transfection

A WT full length FKBP52-FLAG plasmid and matching empty vector were provided by Dr Felix Hausch (Max Planck Institute of Psychiatry, Department of Chemical Genomics, Munich, Germany). The p.A16E, p.N125S, p.Q381L and p.R399Q variants were introduced using site-directed mutagenesis using the Quickchange II XL site-directed mutagenesis kit according to manufacturer's instructions. Primers used for mutagenesis are available on request. Sequence integrity of each construct was confirmed by Sanger sequencing.

HEK293 cells were cultured in 10% FBS (Thermo Fisher, #10500064) prepared in DMEM (Thermo Fisher, #41966052) supplemented with penicillin and streptomycin (Thermo Fisher, #15140122). Cells were passaged with 0.25% Trypsin-EDTA (Thermo Fisher, #25200056) when they approached 80-90% confluency. Transfections were performed in a T25 flask using FuGene (Promega) according to the manufacturer's instructions.

PPIase activity

The effect of each variant was assayed compared to WT and empty vector controls by monitoring the peptidyl-prolyl *cis-trans* isomerase (PPIase) activity of FKBP52. Essentially, 48 hours following transfection of constructs (in triplicate), proteins were extracted from cells using standard procedures. Pellets were immediately frozen and sent to Eurofins Selcia Drug Discovery (Ongar, Essex, UK) where PPIase activity was measured spectrophotometically according to the method described in Janowski et al (1997)[9]. A two-tailed, equal variance t-test was used to examine statistical significance. Protein loading was confirmed by western bloting, carried out using anti-FLAG M2 antibody (Sigma F1804) and GAPDH as a control.

ACKNOWLEDGMENTS

We would like to thank all the patients and control participants who took part in this study. We thank Dr Carol Austin and Dr Johnathan Sanvoisin at Eurofins Selcia Drug Discovery for their invaluable assistance with the PPIase assay. C.D. is funded by Save the Baby Unit (http://www.vhh.co.uk/our-work/save-the-baby-unit-st-mary-s-hospital-paddington) and the Medical Research Council (MRC). The work was supported by the National Institute for Health Research Biomedical Research Centre at Great Ormond Street Hospital for Children NHS Foundation Trust and University College London. P.S. is also supported by Great Ormond Street Children's Charity and the MRC. G.E.M.'s research is funded by MRC, Wellbeing of Women and Sparks. Funding sources had no involvement in study design or data interpretation.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

REFERENCES

1. Stirrat, G.M. (1990) Recurrent miscarriage. *Lancet.* **336**, 673-675.

2. Regan, L., Braude, P.R., Trembath, P,L. (1989) Influence of past reproductive performance on risk of spontaneous abortion. *BMJ*. **299**, 541-545.

3. The ESHRE Guideline Group on RPL. (2018) ESHRE guideline: reccurrent pregnancy loss. Hum. Reprod. Open. 1-12.

4. Bertina, R.M., Koeleman, B.P., Koster, T., Rosendaal, F.R., Dirven, R.J., de Ronde, H., van der Velden, P.A., Reitsma, P.H. (1994) Mutation in the blood coagulation factor V associated with resistance to activated protein C. *Nature*, **369**, 64-67.

5. Peattie, D.A. (1992) Expression and characterization of human FKBP52, an immunophilin that associates with the 90-kDa heat shock protein and is a component of steroid receptor complexes. *Proc. Natl. Acad. Sci. USA.* **89**, 10974-10978.

6. Tranguch, S., Cheung-Flynn, J., Daikoku, T., Prapapanich, V., Cox, M.B., Xie, H., Wang, H., Das, S.K., Smith, D.F., Dey, S.K. (2005) Cochaperone immunophilin FKBP52 is critical to uterine receptivity for embryo implantation. *Proc. Natl. Acad. Sci. USA* **102**, 14326-14331.

7. Tranguch, S., Wang, H., Daikoku, T., Xie, H., Smith, D.F., Dey, S.K. (2007) FKBP52 deficiency-conferred uterine progesterone resistance is genetic background and pregnancy stage specific. *J. Clin. Invest.* **117**, 1824-1834.

8. Hirota, Y., Acar, N., Tranguch, S., Burnum, K.E., Xie, H., Kodama, A., Osuga, Y., Ustunel, I., Friedman, D.B., Caprioli, R.M. et al. (2010) Uterine FK506-binding protein 52 (FKBP52)-peroxiredoxin-6 (PRDX6) signaling protects pregnancy from overt oxidative stress. *Proc. Natl. Acad. Sci. USA*. **107**, 15577-15582.

9. Jankowski, B., Wollner, S., Schutowski, M., Fischer, G. (1997) A protease-free assay for peptidyl prolyl cis/trans isomerases using standard peptide substrates. *Anal. Biochem.* **252**, 299–307.

10. Ng, P.C., Henikoff, S. (2011) Predicting deleterious amino acid substitutions. *Genome Res.* **11**, 863-874.

11. 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. *Nat. Methods.* **7**, 248-249.

12. Choi, Y., Sims, G.E., Murphy, S., Miller, J.R., Chan, A.P. (2012) Predicting the functional effect of amino acid substitutions and indels. *PLoS One*. **7**, e46688.

13. Bao, L., Zhou, M., Cui, Y. (2005) nsSNPAnalyzer: identifying disease-associated nonsynonymous single nucleotide polymorphisms. *Nucleic Acids Res.* **33**, W480-482.

14. Schwarz, J.M., Cooper, D.N., Schuelke, M., Seelow, D. (2014) MutationTaster2: mutation prediction for the deep-sequencing age. *Nat. Methods.* **11**, 361-362.

15. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P., Cunningham, F. (2016) The Ensembl Variant Effect Predictor. *Genome Biology*. **17**, 122.

16. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*. **23**, 2947-2948.

17. Bermingham, N.A., Rauf, S., Katsanis, N., Martin, J.E., Hunter, A.J., Fisher, E.M. (1998) The immunophilin FKBP4 (FKBP52/FKBP59) maps to the distal short arm of human chromosome 12. *Mamm. Genome*. **9**, 268.

18. Scammell, J.G., Hubler, T.R., Denny, W.B., Valentine, D.L. (2003) Organization of the human FK506-binding immunophilin FKBP52 protein gene (FKBP4). *Genomics.* **81**, 640-643.

19. Callebaut, I., Renoir, J.M., Lebeau, M.C., Massol, N., Burny, A., Baulieu, E.E., Mornon, J.P. (1992) An immunophilin that binds M(r) 90,000 heat shock protein: main structural features of a mammalian p59 protein. *PNAS USA.* **19**, 6270-6274.

20. Chambraud, B., Rouviere-Fourmy, N., Radanyi, C., Hsiao, K., Peattie, D.A., Livingston, D.J., Baulieu, E.E. (1993) Overexpression of p59-HBI (FKBP59), full length and domains, and characterization of PPlase activity. *Biochem. Biophys. Res. Commun.* **196**, 160-166.

21. Le Bihan, S., Renoir, J.M., Radanyi, C., Chambraud, B., Joulin, V., Catelli, M.G., Baulieu, E.E. (1993) The mammalian heat shock protein binding immunophilin (p59/HBI) is an ATP and GTP binding protein. *Biochem. Biophys. Res. Commun.* **195**, 600-607.

22. Radanyi, C., Chambraud, C., Baulieu, E.E. (1994) The ability of the immunophilin FKBP59-HBI to interact with the 90-kDa heat shock protein is encoded by its tetratricopeptide repeat domain. *Proc. Natl. Acad. Sci. USA.* **91**, 11197–11201.

23. Chambraud, B., Belabes, H., Fontaine-Lenoir, V., Fellous, A., Baulieu, E.E. (2007) The immunophilin FKBP52 specifically binds to tubulin and prevents microtubule formation. *FASEB J.* **21**, 2787-2797.

24. Massol, N., Lebeau, M.C., Renoir, J.M., Faber, L.E., Baulieu, E.E. (1992) Rabbit FKBP59-heat shock protein binding immunophillin (HBI) is a calmodulin binding protein. *Biochem. Biophys. Res. Commun.* **187**, 1330-1335.

25. Riggs, D.L., Cox, M.B., Tardif, H.L., Hessling, M., Buchner, J., Smith, D.F. (2007) Noncatalytic role of the FKBP52 peptidyl-prolyl isomerase domain in the regulation of steroid hormone signaling. *Mol. Cell. Biol.* **27**, 8658-8669.

26. Riggs, D.L., Roberts, P.J., Chirillo, S.C., Cheung-Flynn, J., Prapapanich, V., Ratajczak, T., Gaber, R., Picard, D., Smith, D.F. (2003) The Hsp-90 binding pepridylprolyl isomerase FKBP52 potentiates glucocorticoid signaling in vivo. *EMBO J.* **22**, 1158-1167.

27. Silverstein, A.M. (1999) Different regions of the immunophilin FKBP52 determine its association with the glucocorticoid receptor, hsp90, and cytoplasmic dynein. *J. Biol. Chem.* **274**, 36980-36986.

28. Cheung-Flynn, J., Roberts, P.J., Riggs, D.L., Smith, D.F. (2003) C-terminal sequences outside the tetratricopeptide repeat domain of FKBP51 and FKBP52 cause differential binding to Hsp90. *J. Biol. Chem.* **278**, 17388-17394.

29. Wu, B., Li, P., Liu, Y., Lou, Z., Ding, Y., Shu, C., Ye, S., Bartlam, M., Shen, B., Rao, Z., (2004) 3D structure of the human FK506-binding protein 52: implications for the assembly of the glucocorticoid receptor/Hsp90/immunophilin heterocomplex. *Proc. Natl. Acad. Sci. USA.* **101**, 8348-8353.

30. Kim, K.H., Lee, K.A. (2014) Maternal effect genes: Findings and effects on mouse embryo development. *Clin. Exp. Reprod.* **41**, 47-61.

31. Dadé, S., Callebaut, I., Paillisson, A., Bontoux, M., Dalbiès-Tran, R., Monget, P. (2004) In silico identification and structural features of six new genes similar to MATER specifically expressed in the oocyte. *Biochem. Biophys. Res. Commun.* **324**, 547-553.

32. Hamatani, T., Carter, M.G., Sharov, A.A., Ko, M.S. (2004) Dynamics of global gene expression changes during mouse preimplantation development. *Dev. Cell*. **6**, 117-131.

33. Hirota, Y., Tranguch, S., Daikoku, T., Hasegawa, A., Osuga, Y., Taketani, Y., Dey, S.K. (2008) Deficiency of immunophilin FKBP52 promotes endometriosis. *Am. J. Pathol.* **173**, 1747-1757.

34. Galat, A. (2004) A note on clustering the functionally-related paralogues and orthologues of proteins: a case of the FK506-binding proteins (FKBPs). *Comput. Biol. Chem.* **28**, 129-140.

35. Schreiber, S.L. (1991) Chemistry and biology of the immunophilins and their immunosuppressive ligands. *Science*. **251**, 283-287.

36. Pratt, W.B., Toft, D.O. (1997) Steroid receptor interactions with heat shock protein and immunophilin chaperones. *Endocr. Rev.* **18**, 306-360.

37. Picard, D., Yamamoto, K.R. (1987) Two signals mediate hormone-dependent nuclear localization of the glucocorticoid receptor. *EMBO J.* **11**, 3333-3340.

38. Davies, T.H., Ning, Y.M., Sánchez, E.R. (2002) A new first step in activation of steroid receptors: hormone-induced switching of FKBP51 and FKBP52 immunophilins. *J. Biol. Chem.* **277**, 4597-4600.

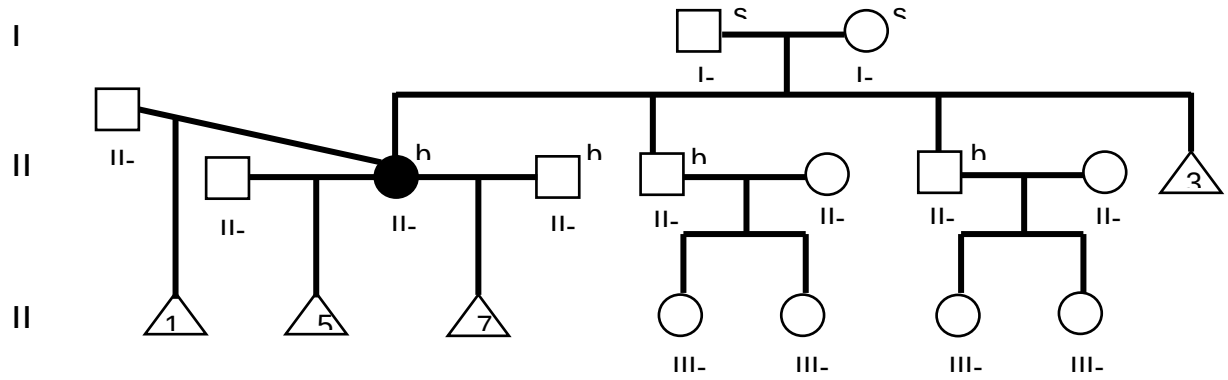
39. Galigniana, M.D., Radanyi, C., Renoir, J.M., Housley, P.R., Pratt, W.B. (2001) Evidence that the peptidylprolyl isomerase domain of the hsp90-binding immunophilin FKBP52 is involved in both dynein interaction and glucocorticoid receptor movement to the nucleus. *J. Biol. Chem.* **276**, 14884-14889.

40. Swegen, A., Grupen, C.G., Gibb, Z., Baker, M.A., de Ruijter-Villani, M., Smith, N.D., Stout, T.A.E., Aitken, R.J. (2017) From Peptide Masses to Pregnancy Maintenance: A Comprehensive Proteomic Analysis of The Early Equine Embryo Secretome, Blastocoel Fluid, and Capsule. *Proteomics*. doi: 10.1002/pmic.201600433.

41. Chen, H.Y., Li, O.Y., Pang, L.H., Xu, H., Fan, X.J., Liang, H.F., Chen, X.F., Qing, J.Z., Huang, R.D., Deng, B.Y. (2015) Expression of FK506-binding protein 52 (FKBP52) in chorionic villi with early recurrent spontaneous abortion. *J. Matern. Fetal Neonatal. Med.* **28**, 1165-1169.

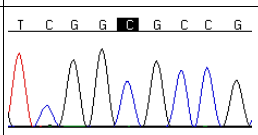
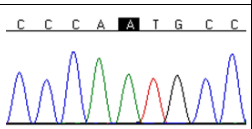
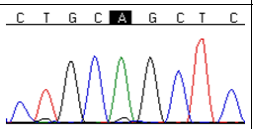
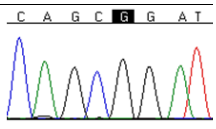
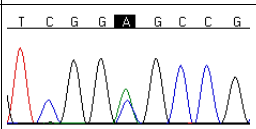
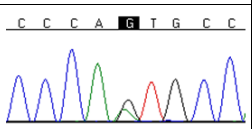
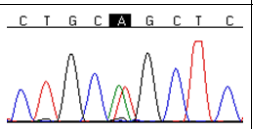
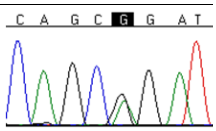
FIGURES

Figure 1: Family pedigree.



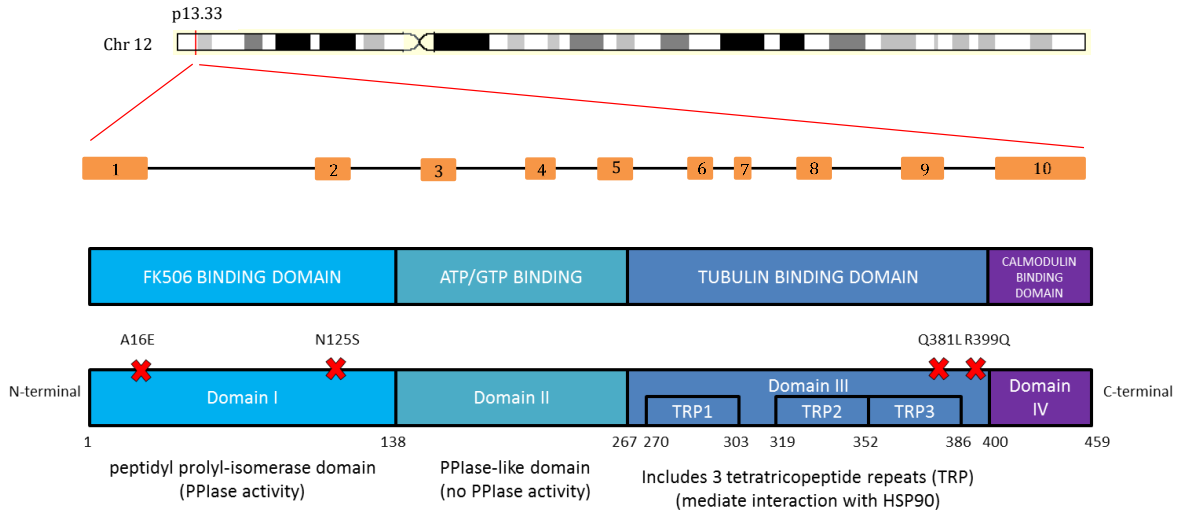
The affected patient (II-3) had a total of 29 early miscarriages: 17 with her first husband (II-1), five with her second husband (II-2) and seven with her third husband (II-4). Her two brothers (II-5 and II-7) had each two successful pregnancies. Her mother (I-2) had suffered three late miscarriages. Blood was collected from individuals marked with a “b” and buccal swab was collected from individuals marked with an “s”.

Figure 2: Analysis of variants in *FKBP4*.

		1) c.47C>A; p.A16E	2) c.374A>G; p.N125S	3) c.1142A>T; p.Q381L	4) c.1196G>A; p.R39
A) Sanger sequencing chromatograms	Wildtype				
	Patient				
B) *In silico* pathogenicity predictions	Variant Effect predictor-LoF score	Probably damaging	Probably damaging	Probably damaging	Probably damaging
	Variant Effect predictor-Condel	Deleterious	Neutral	Deleterious	Deleterious
	Mutation taster	Neutral	Disease causing Splice site changes	Disease causing	Disease causing Splice site changes
	PROVEAN	Neutral	Deleterious	Deleterious	Neutral
	SIFT	Damaging	Tolerated/Damaging*	Damaging	Damaging
	Polyphen-2	Benign	Benign	Benign/Possibly damaging*	Possibly damaging
	nsSNP Analyzer	Neutral	Disease	Disease	Neutral

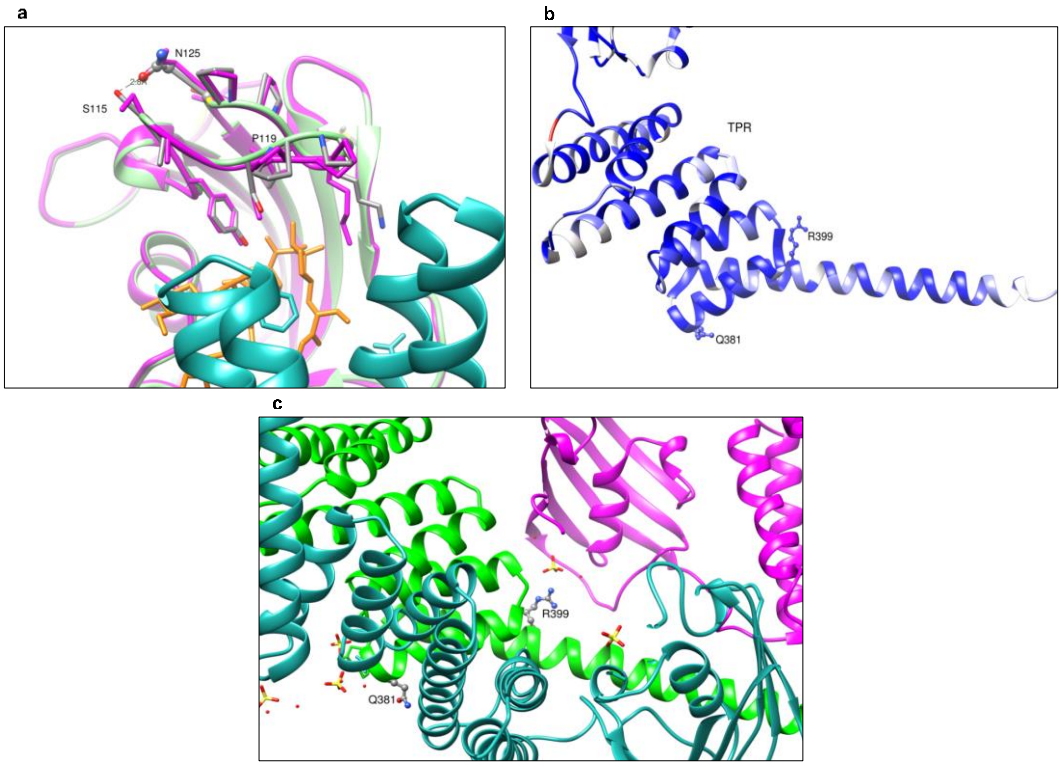
A) Four missense variants were identified in four RPL patients. Both WT and mutant DNA sequence is displayed. Patient 1 (exome patient) had a mutation in exon 1 (c.47C>A; p.A16E). Patient 2 had a mutation in exon 3 (c.374A>G; p.N125S). Patients 3 and 4 each had a mutation in exon 9 (c.1142A>T; p.Q381L) and (c.1196G>A; p.R399Q). B) *In silico* predictions were used to predict pathogenicity of the four variants (LoF: Loss of Function). C) Amino acid conservation of the human FKBP52 protein and its orthologues in selected species. The aa residues at each of the identified variants (p.A16E, p.N125S, p.Q381L and p.R399Q) are highlighted in yellow.

Figure 3: Graphical view of *FKBP4* chromosomal position, gene structure and protein product, along with the distribution of four identified missense variants.



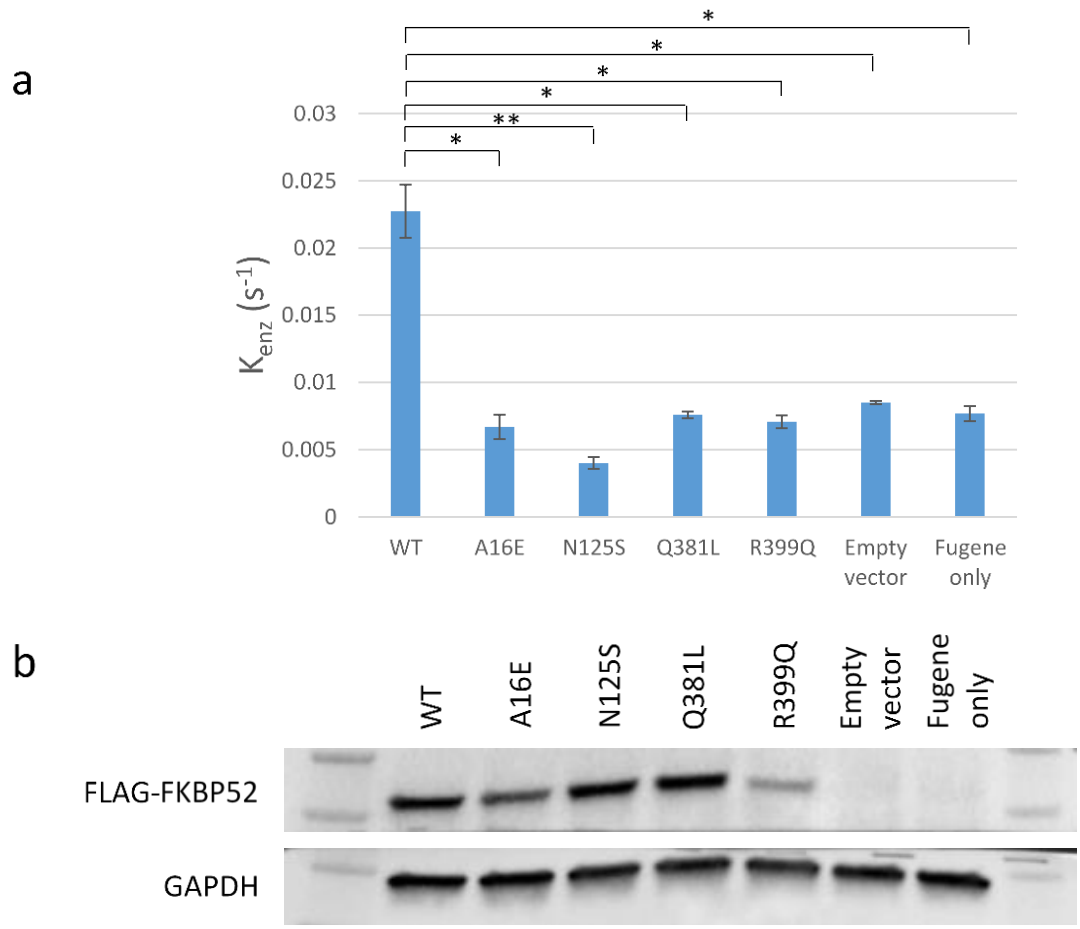
FKBP4 maps to chromosomal region 12p13.33 and encodes a 10-exon transcript that spans approximately 10 kb of genomic DNA[16,17]. Grey boxes represent exons and black lines represent introns. The translated protein (459 amino acids) includes four individual and functional domains[18]. Domain I or FK1 includes the N-terminal portion of FKBP52 and consists of the FK506-binding site, which contains peptidyl-prolyl *cis-trans* isomerase (PPIase) activity[19]. Domain II or FK2 contains the PPIase-like domain, which shares 32% sequence homology to domain I, but exhibits no PPIase activity[18,19]. Domain II contains an ATP/GTP binding sequence[20]. Domain III possesses three tetratricopeptide repeat (TPR) domains that serve as binding sites for 90-kDa heat shock protein (Hsp90), which is also a component of the steroid receptor complexes[21]. The region located between aa 267 and 400 is required for tubulin binding[22]. Domain IV is the C-terminal domain of FKBP52 and it is a short helix that contains a putative binding site for calmodulin[23]. Black crosses indicate the position and protein variant of each mutation identified in the four RPL patients. The structure of the *FKBP4* transcript and organization of the protein domains are not drawn to exact scale.

Figure 4: Structure analysis of FKBP52 variants.



a) Hydrogen bond between N125 and S115 in the FK1 domain of FKBP52 in the unbound form (PDB ID: 1Q1C [28]). Crystal structure of FKBP52 (magenta) bound to Rapamycin (orange) and FRB domain of mTOR (blue) (PDB ID: 4DRJ [40]) is superposed. N125S mutation can result in the loss of hydrogen bond interaction with S115, which appears to be important for conformation changes involved in the active site loop (residues 112-127). b) Conservation of residues in the TPR domain and C-terminal helix. Red to Blue shows increasing degree of conservation. Both R399 and Q381 are highly conserved among FKBP52 family and are surface exposed in the structure, strongly suggesting a role in interactions with binding partners. c) Crystal structure of C-terminal domains (FK2, TPR and C-term extension) of FKBP52 (PDB ID: 1P5Q [28]). R399Q results in the loss of a positive charge at the surface while Q381L replaces a polar side chain with a hydrophobic group potentially affecting the interactions mediated by these two residues.

Figure 5: PPIase activity of *FKBP4* constructs.



a) HEK293 cells were transiently transfected with constructs containing WT or mutant full-length *FKBP4* cDNAs, empty vector and Fugene reagent transfection only controls. Following 48 hour incubation, proteins were extracted and measured for PPIase activity. * = $p<0.002$; **= $p<0.001$. b) Western blot showing protein loading using anti-FLAG (top) and anti-GAPDH (bottom).

ABBREVIATIONS

APLS	Antiphospholipid syndrome
FKBP4	FK506 binding protein 4
FVL	Factor V Leiden
GR	Glucocorticoid receptor
Hsp90	Heat shock protein 90
LA	Lupus anticoagulant
LBD	Ligand binding domain
LoF	Loss of function
PPIase	Peptidyl-prolyl cis-trans isomerase
PR	Progesterone receptor
PRDX6	Peroxiredoxin-6
RPL	Recurrent pregnancy loss
SERPINB2	Serpin peptidase inhibitor clade b2
TPR	Tetratricopeptide
WES	Whole-exome sequencing
WT	Wildtype