Human Fertility, 2014; Early Online: 1–8
© 2014 The British Fertility Society
ISSN 1464-7273 print/ISSN 1742-8149 online
DOI: 10.3109/14647273.2014.907506

informa
healthcare

ORIGINAL ARTICLE

SPACA3 gene variants in a New Zealand cohort of infertile and fertile couples

DEBORAH PRENDERGAST[1], KATIE J. WOAD[3], LAWRENCE W. CHAMLEY[2], OLIVIA J. HOLLAND[2] & ANDREW N. SHELLING[2]

[1]*Faculty of Medical and Health Sciences, The University of Auckland, Auckland, New Zealand,* [2]*Department of Obstetrics and Gynaecology, Faculty of Medical and Health Sciences, The University of Auckland, Auckland, New Zealand, and* [3]*School of Veterinary Medicine and Science, Sutton Bonington Campus, University of Nottingham, Loughborough, UK*

Abstract

SPRASA (also referred to as SLLP1) is a protein identified in the acrosome of human sperm and encoded by the gene *SPACA3*. SPRASA is associated with sperm-oocyte recognition and binding, and may play a role in fertility. In order to determine whether variants in the *SPACA3* gene are associated with human infertility, we undertook a genetic analysis of 102 infertile and 104 fertile couples. Three gene variants were identified using PCR-based DNA sequencing; 1) an insertion of TGC within a quadruple trinucleotide (TGC) repeat region in the 5′ untranslated region (UTR) (g.–22TGC(4_5), 2) a guanine to adenosine transition at position 239 (c.239G>A) resulting in a non-synonymous amino acid substitution from cysteine to tyrosine (p.C80Y) at position 80 in the putative transmembrane region, and 3) a novel nucleotide variant (c.691G>C) located in the 3′UTR. A functional effect of the g.–22TGC (4_5) was confirmed by a luciferase expression assay, while the effects of the variants c.239G>A and c.691G>C were predicted using *in silico* analysis. Although the frequencies of these variants were not significantly different between the infertile and fertile populations, we present evidence that the variants could affect the expression levels or function of SPRASA, thereby affecting a couple's fertility. Larger populations, especially individuals/couples with unexplained infertility, need to be screened for these variants to validate a relationship with fertility.

Keywords: *Reproduction, gene mutation, unexplained infertility*

Introduction

The prevalence of infertility in Western countries such as New Zealand is estimated to be one in six couples, while one in eight couples require some form of medical assistance to achieve a pregnancy (HFEA, 2006). Of these infertile cases, the precise cause of infertility for approximately 5–15% of couples is unknown, and is referred to as unexplained infertility (Farquhar, 1998; Liu et al., 2007). Standard investigations, including testing for ovulation, tubal patency and semen analysis are routine in these couples (Kamath & Bhattacharya, 2012). The causes of unexplained infertility and infertility, in general, are likely to be heterogeneous, and may include endocrine, immunological, and genetic factors (Kamath & Bhattacharya, 2012; Pellicer et al., 1998). It is also common to find multiple potential causes of infertility in an infertile couple (Collins & Crosignani, 1992; Marrero & Ory, 1991; Quaas & Dokras, 2008).

SPRASA (also referred to as SLLP1) is a protein identified in the acrosome of human sperm (Chi et al., 2004). SPRASA is a member of the c-type lysozyme/alpha-lactalbumin family (Chiu et al., 2004; Mandal et al., 2003), and human SPRASA is encoded by a gene (*SPACA3*) at chromosomal locus 17q11.2 (Mandal et al., 2003). SPRASA is known to be the target for antisperm antibodies in some infertile men, and this protein has been shown to have roles in sperm–oocyte binding and early embryo development in murine, hamster, or bovine models (Chiu et al., 2004; Mandal et al., 2003; Zhang et al., 2005; A. Wagner, unpublished). SPRASA expression was thought to be limited to the sperm/testis and some tumour cells, but we have recently demonstrated it is also expressed by oocytes (A. Wagner, unpublished). This pattern of gonad/gamete-specific expression also supports the concept that SPRASA has an important role in fertility (Herrero et al., 2005;

Correspondence: Olivia Holland, Department of Obstetrics and Gynaecology, Faculty of Medical and Health Sciences, The University of Auckland, Private Bag 92019 Auckland, New Zealand. Tel: + 64-9-9232314. E-mail: olivia.holland@auckland.ac.nz

(Received 4 September 2013; accepted 3 December 2013)

PROOF ONLY

THUF_A_907506.indd 1 4/10/2014 1:49:43 PM

Mandal et al., 2003; Wagner, 2009). While we discovered SPRASA as the antigen for antisperm antibodies in some infertile men, the association of SPRASA with human infertility has not been investigated further. Because of SPRASA's putative importance in fertility, naturally occurring genetic variation that disrupts the function of SPRASA may have a role in human infertility. This study was undertaken to investigate whether genetic variants in the SPRASA-encoding *SPACA3* gene are associated with infertility in humans.

Methods and materials

Ethics statement and study population

This study was approved by the Northern Regional Ethics Committee (Auckland, New Zealand). Blood samples obtained by venipuncture were collected following written informed consent from 102 infertile couples, (recruited through Fertility Associates, Auckland and Fertility Plus, Auckland; Supplementary Table I available online at: http//informahealthcare.com/doi/abs/10.3109/14647273.2014.907506) and 104 fertile couples (recruited from the general Auckland population through media advertisement; Supplementary Table II available online at: http//informahealthcare.com/doi/abs/10.3109/14647273.2014.907506). Couples were defined as infertile if they were unable to conceive after one year of regular intercourse without the use of contraception. The clinically diagnosed disorders affecting the infertile cohort recruited for this study are summarised in Supplementary Table I available online at: http//informahealthcare.com/doi/abs/10.3109/14647273.2014.907506. Couples were defined as fertile if they had given birth in the previous 2 years.

Mutation nomenclature

In accordance with current mutation nomenclature (den Dunnen & Antonarakis, 2001), the adenosine of the ATG (initiator methionine codon) of *SPACA3* is denoted nucleotide + 1. Nucleotide numbers refer to GenBank accession number NM_173847.

Polymerase chain reaction

Genomic DNA was extracted from 5 to 10 mL of blood using the salting out protocol (Miller et al., 1988). Eight sets of Polymerase chain reaction (PCR) primers were designed to flank exons and the putative promoter regions of *SPACA3* (Supplementary Table III available online at: http//informahealthcare.com/doi/abs/10.3109/14647273.2014.907506). Genomic DNA (100ng) was amplified using standard PCR conditions (Supplementary Tables III and IV available online at: http//informahealthcare.com/doi/abs/10.3109/14647273.2014.907506). All PCR amplifications were performed in an iCycler (BioRad, Hempstead, UK). The PCR products were analysed using agarose gel electrophoresis, stained with ethidium bromide (10mg/ml; Sigma, Australia) and visualised under UV light.

DNA sequencing

The PCR products were purified with the Roche High Pure PCR Product Purification Kit (Roche, Basel, Switzerland) following the manufacturer's instructions. Sequencing reactions were performed using the ABI prism Big Dye Terminator Sequencing Kit v3.1 (ABI, Foster City, USA) following the manufacturer's instructions in a 20 μL volume with approximately 5 ng per 100 bp of PCR product. All primers used in the sequencing reactions allowed for 100% coverage of the coding region for each sample. Sequencing reactions for exon four were supplemented with betaine at a final concentration of 1.4M to assist sequencing through secondary structures (Haqqi et al., 2002). The variants were confirmed by sequence replication. Sequencing of the promoter variants was carried out in both directions using internal primers (Supplementary Table III available online at: http//informahealthcare.com/doi/abs/10.3109/14647273.2014.907506). Sequencing was amplified using standard conditions (Supplementary Table VI available online at: http//informahealthcare.com/doi/abs/10.3109/14647273.2014.907506). Sequencing extension products were precipitated by $MgSO_4$. Capillary electrophoresis was analysed on a 3100 Genetic Analyser (Applied Biosystems, Life Technologies New Zealand Limited, Auckland) at the Genome Research Facility at the School of Biological Sciences, The University of Auckland.

Luciferase assay for promoter function: variant g.–22TGC(4_5)

The three putative promoters isolated in this study, Promoter A (located five-prime of exon one; major allele; 983bp), Promoter A g.–22TGC(4_5) variant (located five-prime of exon one; 986bp) and Promoter B (located in intron one; invariant; Figure 1), were investigated for their capacity to support transcriptional activity using luciferase assay.

Promoter A, Promoter A g.–22TGC(4_5) variant and Promoter B were cloned upstream of the firefly luciferase gene of the pGL3-enhancer reporter plasmid (Promega, Wisconsin, USA). The cell lines HEK-293T (non-*SPACA3* expressing negative control; human embryonic kidney; American Type Culture Collection, Manassas, USA), NCCIT (*SPACA3* expressing teratocarcinoma; human pluripotent embryonal carcinoma; American Type Culture Collection, Manassas, USA) and KGN (*SPACA3* expressing granulosa; human granulosa cell tumour originated from a stage III granulosa cell tumour of the ovary; provided by Dr Ashwini Chand, Prince Henry's Institute of Medical Research, Victoria, Australia) were co-transfected with the RL-Tk vector (Promega, In vitro Technologies, Auckland) and either the (1) pGL3-Promoter A construct, (2) pGL3-Promoter A g.–22TGC(4_5) variant construct,

PROOF ONLY

THUF_A_907506.indd 2 4/10/2014 1:49:44 PM

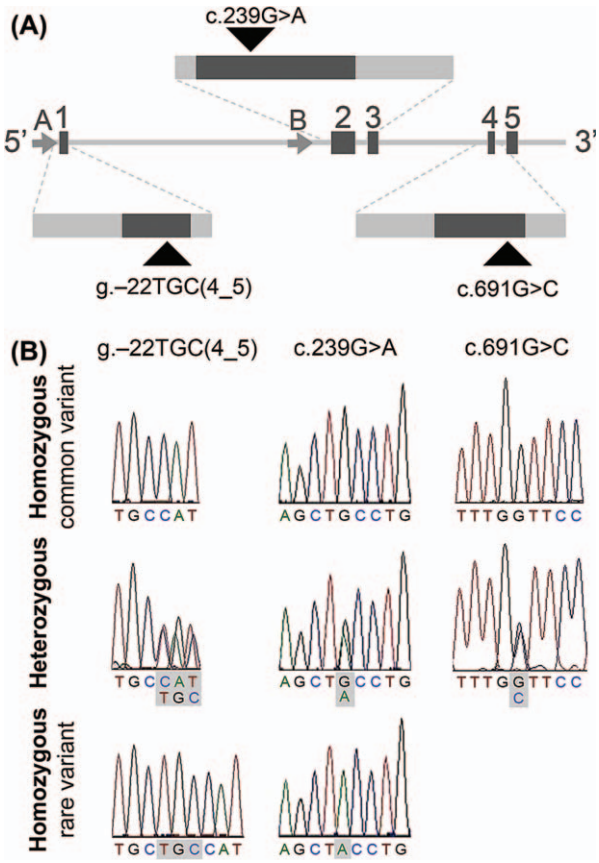

Figure 1. The genomic organisation of human *SPACA3* and the location of variants identified in this study. (A) Schematic diagram of *SPACA3*; exons are shown in *dark grey*, introns are shown in *light grey* and the putative promoters (A and B) are indicated by *grey arrows*. The location of the variants g.–22TGC(4_5), c.239G>A and c.691G>C are indicated by *black triangles* on magnified sections. (B) Representative electropherograms of the variants g.–22TGC(4_5), c.239G>A and c.691G>C.

(3) Promoter B construct or (4) pGL3 empty vector. Cells were transfected using Lipofectamine 2000 (Invitrogen, Auckland) according to the manufacturer's instructions. Cells were cultured for 24 h in RPMI 1640 medium (Invitrogen, Auckland) supplemented with 2mM L-glutamine (Invitrogen, Auckland) and 10% fetal calf serum (Invitrogen, Auckland). All cell lines were maintained at 37°C in 5% CO_2.

Luciferase assays were performed using the Dual Luciferase Reporter Assay System (Promega, In vitro Technologies, Auckland) following the manufacturer's instructions. The promoter assays were normalised to Renilla luciferase activity from the co-transfected pRL-Tk vector. The means and standard error (SE) were calculated from identical transfections in triplicate from three independent experiments, and the two *SPACA3* promoter variants were compared to the empty control plasmid transfected into the appropriate cell lines.

In silico analysis for function: variants g. –22TGC(4_5), c.239G>A and c.691G>C

The putative effects of the sequence variations on genetic association were predicted using the SHEsis programme (Shi & He, 2005). The TESS (transcription element search software) prediction tool was used to determine whether the g.–22TGC(4_5) variant affected transcription factor binding (Schug & Overton, 1997). The possible impact of the amino acid substitution p.C80Y (genetic variant c.239G>A) on the structure, function and expression of the SPRASA protein was tested using the three prediction tools; SIFT (sorts intolerant from tolerant) scores were classified as intolerant (0.00–0.05), potentially intolerant (0.051–0.10), borderline (0.101–0.20) or tolerant (0.201–1.00) (Ng & Henikoff, 2003; Xi et al., 2004); PolyPhen II (polymorphism phenotyping), scores were designated probably damaging (≥ 2.00), possibly damaging (1.50–1.99), potentially damaging (1.25–1.49), borderline (1.00–1.24) or benign (0.00–0.99) (Xi et al., 2004); Grantham matrix scores categorise codon replacement into classes of increasing chemical dissimilarity and are designated conservative (0–50), moderately conservative (51–100), moderately radical (101–150) or radical (≥ 151) (Grantham, 1974; Li et al., 1984; Rudd et al., 2005). The potential effect on microRNA (miRNA) binding to the variant c.691G>C was analysed using the TargetScanHuman version 6.2 programme. TargetScan predicts biological targets of miRNAs by searching for the presence of conserved 8mer and 7mer sites that match the seed region of each miRNA (Lewis et al., 2005), also identified are sites with mismatches in the seed region that are compensated by conserved 3′ pairing (Friedman et al., 2009). In mammals, predictions are ranked based on the predicted efficacy of targeting as calculated using the context scores of the sites (Grimson et al., 2007).

Sequence alignments

In order to obtain a measure of the level of evolutionary conservation of the variants, multiple sequence alignments of *SPACA3* were performed as relevant for each variant using known orthologs with complete sequences in the programme Clustal W2. Thirteen higher primates possess known orthologs to human exon one containing the g.–22TGC(4_5) variant, 42 mammalian species possess known orthologs to human exon two containing the c.239G>A variant, and 23 mammalian species possess known orthologs to human exon five containing the c.691G>C variant.

Statistical analysis

The chi-squared statistic was used to detect the Hardy-Weinberg equilibrium (HWE) for the three identified sequence variations. The odds ratios and 95% confidence interval were used to measure the association in the genotype and allele distribution of the three sequence variations between the infertile and fertile groups. The p values were estimated using chi-square test. Statistical analysis from the results of the luciferase assay comparing the promoter variants was determined using unpaired Student's T-test in the Graph Pad Prism

© 2014 The British Fertility Society

PROOF ONLY

5.02 programme. All p values were two-tailed and considered significant if $p < 0.05$.

Results

In order to investigate the population level variation in *SPACA3* and to determine whether this variation is associated with infertility, 102 infertile couples ($n = 204$ individuals) and 104 fertile couples ($n = 208$ individuals) were analysed. Three sequence variants were identified in this population; g.–22TGC(4_5), c.239G>A and c.691G>C (Figure 1; Table I).

SPACA3 is well conserved in the cohort investigated; the minor allele frequencies for the variants were g.–22TGC(4_5) 0.12, c.239G>A 0.02 and c.691G>C 0.001. The variant c.691G>C was found at a population frequency of less than 0.01, and is thus defined as a mutation (den Dunnen & Antonarakis, 2001). There were a number of couples from the infertile and fertile cohorts who had a combination of multiple variants (Supplementary Table VII available online at: http//informahealthcare.com/doi/abs/10.3109/14647273.2014.907506). The variants identified consisted of one in-frame insertion and two single nucleotide polymorphisms (details below).

SPACA3/SPRASA is highly evolutionarily conserved. A multiple sequence alignment of *SPACA3* orthologs showed 100% conservation of the major variant in exons one and five (Figure 2). Analysis of *SPACA3* orthologs in mammals showed that the guanosine/cytosine is conserved (100% conservation in 42 species; Figure 2).

The first variant in the *SPACA3* gene consists of the insertion of an additional TGC within a quadruple trinucleotide TGC repeat region, g.–22TGC(4_5), and is situated in the 5' UTR of exon one (Figure 1). Allele frequencies for infertile and fertile individuals are shown in Table II. Diagnoses of infertile individuals with this variant are shown in Supplementary Table VIII available online at: http//informahealthcare.com/doi/abs/10.3109/14647273.2014.907506. The frequencies of this variant in the infertile and fertile individuals were calculated to be in HWE and were not significantly different between the infertile and fertile individuals ($p = 0.24$).

In order to investigate the potential consequences of this *SPACA3* variant on promoter function, we conducted luciferase reporter experiments using three cell lines. *In vitro* luciferase reporter assays comparing Promoter A (located five-prime of exon one; major allele) and Promoter A g.–22TGC(4_5) variant (located five-prime of exon one) found expression from Promoter A in the *SPACA3*-expressing granulosa cell line KGN, with an approximately 1.4-fold increase in expression compared to the control. There was no expression above the normalisation-control from Promoter A in the *SPACA3*-expressing seminoma cell line NCCIT or the non-*SPACA3*-expressing embryonic kidney cell line HEK-293T. Promoter A g.–22TGC(4_5) variant did not induce expression above the normalisation-control in any cell line; there was a significant difference between Promoter A and Promoter A g.–22TGC(4_5) variant expression in the reproductively relevant KGN cell line ($p = 0.01$; Figure 3). The function of Promoter B (located in intron one; invariant) was also assessed. Expression from Promoter B was found in the NNCIT (approximately 2.7-fold increase compared to the control) and KGN (approximately 5.6-fold increase compared to the control) cell lines. There was no expression above the normalisation-control from Promoter B in the HEK-293T cell line (Figure 3).

The second variant, a guanosine to adenosine transition at position 239, c.239G>A, was identified in exon two, which encodes the putative transmembrane region of the SPRASA protein (Figure 1). This transition occurred at the second residue of the codon and resulted in a non-synonymous amino acid substitution at codon position 80 from a cysteine to a tyrosine (p.C80Y). Allele frequencies for infertile and fertile individuals are shown in Table II. Diagnoses of infertile individuals with this variant are shown in Supplementary Table VIII available online at: http//informahealthcare.com/doi/abs/10.3109/14647273.2014.907506. The frequencies of this variant in the infertile and fertile individuals were calculated to be in HWE and were not significantly different between the infertile and fertile individuals ($p = 0.36$).

To determine possible effects that the c.239G>A variant may have on the function of the transmembrane region, online prediction tools were used to examine the effect of the amino acid substitution p.C80Y. Bioinformatic analysis by SIFT (Ng & Henikoff, 2001; Xi et al., 2004) and PolyPhen II (Ramensky et al., 2002; Xi et al., 2004) programmes suggested that the substitution is not tolerated, and is predicted to be deleterious to protein

Table I. Variations in *SPACA3* identified in all subjects.

Nucleotide change	Amino acid change	Domain	Previous report	dbSNP ID	PolyPhen II[a] prediction	SHIFT[a] prediction	Grantham[b] score	Sequence conservation[c]
g.–22TGC(4_5)	–	5′UTR	Yes	rs3052914	–	–	–	–
c.239G>A	p.C80Y	Transmembrane	Yes	rs16967845	0.96	0.02	194	High (100% in 42 species)
c.691G>C	–	3′UTR	No	–	–	–	–	–

UTR, untranslated region

[a]PolyPhen II score (0 least; 1 most) and SIFT score (0–1; ≤ 0.05 damaging, > 0.05 tolerated) predicted the damage of the variant.

[b]Grantham score classifies the chemical dissimilarity of the variant (0–50 conservative, 51–100 moderately conservative, 101–150 moderately radical, ≥ 151 radical).

[c]Protein sequence alignment of all known orthologs ($n = 42$).

PROOF ONLY

THUF_A_907506.indd 4 4/10/2014 1:49:45 PM

mono for print colour online

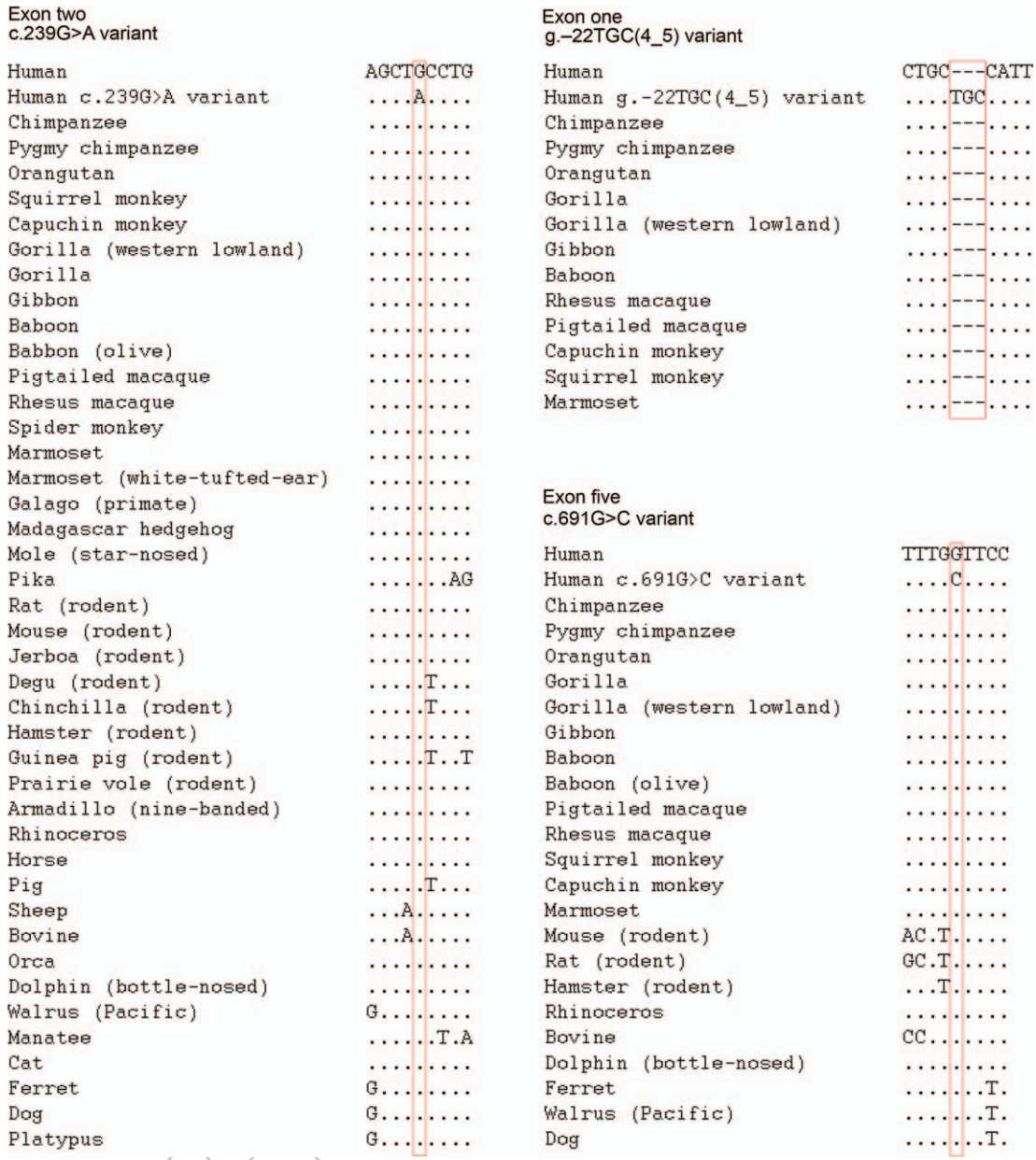

Figure 2. Multiple sequence alignments of SPACA3 orthologs. Sequence alignments of all known SPACA3 othologs were performed in the programme ClustalW2 on exons one containing the g.–22TGC(4_5) variant, exon two containing the c.239G>A variant, and exon five containing the c.691G>C variant. Four nucleotides either side of each variant are shown. The variant region is boxed. *Dots* indicate agreement with the header sequence (human major allele).

structure or function. A similar prediction was observed using the Grantham matrix score (Abkevich et al., 2004; Grantham, 1974; Lee et al., 2008).

A third variant was identified in exon five and resulted in a guanosine to cytosine transversion at nucleotide position 691, c.691G>C (Figure 1; Table I). This novel SNP occurred in the 3′UTR in the heterozygous form in one individual, an infertile female (Table II) diagnosed as having unexplained infertility (Supplementary Table VIII available online at: http//informahealthcare.com/doi/abs/10.3109/14647273.2014.907506). The observed frequencies for this variant were not significantly different between the infertile and fertile individuals ($p = 0.31$). The TargetScanHuman programme version 6.2 (Lewis et al., 2005) identified that the novel variant c.691G>C is located in the seed region of the human miRNA 873 (hs-miR873).

Discussion

Currently there is only limited knowledge about the specific proteins that are involved in human fertility, and how genetic variation in these proteins may affect fertility. In this study we investigated variations in the gene *SPACA3* which encodes SPRASA, a little studied protein that has a role in sperm–oocyte binding (Chiu & Chamley, 2002; Chiu et al., 2004; Herrero et al., 2005; Mandal et al., 2003) and early embryonic development

© 2014 The British Fertility Society

THUF_A_907506.indd 5 4/10/2014 1:49:45 PM

Table II. Allele frequencies and levels of homozygosity/heterozygosity in *SPACA3* variants identified in all subjects.

	Minor allele frequency		Number in infertile patients ($n = 204$)		Frequency in infertile patients		Number in fertile controls ($n = 208$)		Frequency in fertile controls	
Nucleotide change	Infertile patients	Fertile controls	HO	HE	HO	HE	HO	HE	HO	HE
g.–22TGC(4_5)	0.137	0.098	5	46	0.025	0.226	3	35	0.014	0.168
c.239G>A	0.032	0.017	1	11	0.005	0.054	0	7	0	0.034
c.691G>C	0.002	0	0	1	0	0.005	0	0	0	0

HO, homozygote; HE, heterozygote

(A. Wagner, unpublished). Levels of population genetic variation in *SPACA3* were determined in a cohort of New Zealand males and females ($n = 412$). Limited variation was found in *SPACA3*, suggesting that this gene may be well conserved, possibly because its potential role in fertilisation (Herrero et al., 2005; Mandal et al., 2003) means that variation would be detrimental to fertility.

To explore whether variants in *SPACA3* are associated with infertility in the New Zealand population, variation was investigated in infertile and fertile couples. Twenty-nine variants in *SPACA3* have previously been recorded (NCBI SNP database http://www.ncbi.nlm.nih.gov/). In the New Zealand population, three variants were found. All variants were seen at higher frequencies in individuals belonging to infertile, compared to fertile, couples. However, none of these differences were statistically significant. When investigating infertility, couples should be considered together (Campana et al., 1995). Thus, analysis of couples found three variant combinations in the infertile couples only, suggesting that these variant combinations are possibly detrimental to fertility, although the variant combinations were not statistically associated with infertility. Given that infertility has many causes and is often multifactorial, it may be difficult to show an association with infertility in such a broad collection of individuals. The variants identified may therefore be important in the pathogenesis of infertility in the affected individuals. Indeed, we determined that the variant encoding a cysteine to a tyrosine (p.C80Y) transition in the putative transmembrane region may affect the function of SPRASA. Further, the other two variants may affect the expression levels of SPRASA.

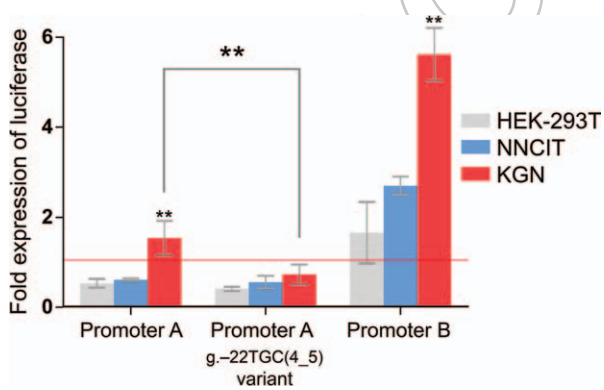

Figure 3. Functional analysis of the human *SPACA3* gene promoter variants isolated in this study. Luciferase reporter assay of Promoter A (located five-prime of exon one; major allele; 983bp), Promoter A g.–22TGC(4_5) variant (located five-prime of exon one; 986bp) and Promoter B (located in intron one; invariant) transfected into HEK-293T (negative control; *grey*), NCCIT (*SPACA3*-expressing teratocarcinoma; *blue*) and KGN (*SPACA3*-expressing granulose; *red*) cell lines. Data are presented as fold luciferase activity calibrated to the appropriate cell line transfected with the empty pGL3-enhancer. The *red line* shows the level of the normalisation control. Results shown as mean ± SE of three independent experiments performed in triplicate. $^{**}p \leq 0.01$.

The functional activity of Promoter A (major allele) in the KGN cell line was abolished when Promoter A g.–22TGC(4_5) variant was transfected instead suggesting that the g.–22TGC(4_5) variant may reduce the expression of *SPACA3*, and could lead to an effect on fertility. Further, the major allele at this locus is highly conserved, suggesting that any variation in this region may be deleterious. However, the g.–22TGC(4_5) variant was found in the homozygous state in three fertile couples, indicating that it does not lead to frank infertility in all individuals, and may have a subtle effect on fertility. A second promoter, Promoter B, is located in intron one. This promoter may allow the continuing expression of an isoform of SPRASA and, therefore, the continuation of some functions of SPRASA in fertility.

The c.239G>A variant results in a non-conservative change of a cysteine to tyrosine and is located in the putative transmembrane region of SPRASA. *In silico* analyses suggest that this substitution is deleterious to the function of SPRASA, and analysis of *SPACA3* orthologs confirms that the cysteine is highly conserved, further indicating that it may be crucial for protein function. The c.239G>A variant was only seen in the homozygous state in a single infertile female, which may indicate an effect on fertility of non-functional SPRASA in this individual. The allele frequency of the c.239G>A variant in the New Zealand population (0.02) is similar to the European and Asian frequencies (0.03 and 0.01, respectively; refer to cluster reports ss227519113, ss23645285, and ss237222685 on dpSNP http://www.ncbi.nlm.nih.gov/snp), indicating the potential of this variant to affect fertility in multiple populations.

The heterozygous variant c.691G>C was identified in a single female with unexplained infertility. The major variant at this locus is conserved across all known *SPACA3* orthologs, indicating its importance to the functioning of SPRASA. This variant is located within the seed region for hs-miR873, a miRNA

mono for print colour online

PROOF ONLY

THUF_A_907506.indd 6 4/10/2014 1:49:48 PM

(Pasten-Hidalgo et al., 2008) that can specify post-transcriptional repression by transcript destabilisation/translational repression (Friedman et al., 2009). The identification of hs-miR873 and the conservation of this region suggest that this locus may be important in the regulation of *SPACA3* by affecting the stability and/or translation of the mRNA, thereby affecting levels of the SPRASA protein.

Summary

In this investigation, we have determined that *SPACA3* has low levels of variation in our study population. Three variants were detected; although none were statistically associated with infertility; this is not a surprising finding, given the multifactorial nature of infertility. Furthermore, because infertility has multiple causes, the explanation of an individual's infertility may not be associated with infertility when examined on a population basis. All three of the variants were found at higher frequencies in the infertile cohort and one of the variants was only found in the homozygous state in an infertile individual, suggesting that this could be associated with the pathogenesis of infertility in some individuals/couples. Two of the variants may affect the expression levels of *SPACA3*, and the other variant may affect the function of SPRASA. Results from multiple models indicate that SPRASA is important in fertility (Chiu et al., 2004; Mandal et al., 2003; Zhang et al., 2005; A. Wagner, unpublished), and thus variants that affect SPRASA's function or expression levels may compromise fertility for some couples. Further functional studies are required to determine whether the presence of these variants has a role in infertility.

Acknowledgement

We would like to thank the infertile and fertile couples for their involvement in this study, staff at Fertility Plus and Fertility Associates, as well as Irma Heijboer for recruiting the couples. This work was supported by a grant from the Marsden Fund of the Royal Society of New Zealand.

Declaration of interest: The authors report no declarations of interest. The authors alone are responsible for the content and writing of the paper.

References

Abkevich, V., Zharkikh, A., Deffenbaugh, A.M., Frank, D., Chen, Y., Shattuck, D., et al. (2004). Analysis of missense variation in human BRCA1 in the context of interspecific sequence variation. *Journal of Medical Genetics,* 41, 492–507.

Campana, A., de Agostini, A., Bischof, P., Tawfik, E., & Mastrorilli, A. (1995). Evaluation of infertility. *Human Reproduction Update*, 1, 586–606.

Chiu, W.W. & Chamley, L.W. (2002). Use of antisperm antibodies in differential display Western blotting to identify sperm proteins important in fertility. *Human Reproduction,* 17, 984–989.

Chiu, W.W., Erikson, E.K., Sole, C.A., Shelling, A.N., & Chamley, L.W. (2004). SPRASA, a novel sperm protein involved in immune-mediated infertility. *Human Reproduction*, 19, 243–249.

Collins, J.A. & Crosignani, P.G. (1992). Unexplained infertility: a review of diagnosis, prognosis, treatment efficacy and management. *International Journal of Gynaecology and Obstetrics,* 39, 267–275.

den Dunnen, J.T. & Antonarakis, S.E. (2001). Nomenclature for the description of human sequence variations. *Human Genetics*, 109, 121–124.

Dieffenbach, C.W. & Dveksler, G.S. (eds.) (1995). *PCR Primer: A Laboratory Manual*, 1st ed. New York: Cold Spring Harbor Laboratory Press.

Farquhar, C. (1998). Infertility: causes, investigations and treatments. *New Zealand Family Physician*, 25, 11–15.

Friedman, R.C., Farh, K.K., Burge, C.B., & Bartel, D.P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research*, 19, 92–105.

Grantham, R. (1974). Amino acid difference formula to help explain protein evolution. *Science*, 185, 862–864.

Grimson, A., Farh, K.K., Johnston, W.K., Garrett-Engele, P., Lim, L.P., & Bartel, D.P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Molecular Cell*, 27, 91–105.

Haqqi, T., Zhao, X., Panciu, A., & Yadav, S.P. (2002). Sequencing in the presence of betaine: Improvement in sequencing of the localized repeat sequence regions. *Journal of Biomolecular Techniques*, 13, 265–271.

Herrero, M.B., Mandal, A., Digilio, L.C., Coonrod, S.A., Maier, B., & Herr, J.C. (2005). Mouse SLLP1, a sperm lysozyme-like protein involved in sperm-egg binding and fertilization. *Developmental Biology*, 284, 126–142.

HFEA. (2006). Facts and Figures: Fertility Problems and Treatment. http://www.hfea.gov.uk/5777.html

Kamath, M.S. & Bhattacharya, S. (2012). Demographics of infertility and management of unexplained infertility. *Best Practice & Research. Clinical Obstetrics & Gynaecology*, 26, 729–738.

Lee, E., McKean-Cowdin, R., Ma, H., Chen, Z., Van Den Berg, D., Henderson, B.E., et al. (2008). Evaluation of unclassified variants in the breast cancer susceptibility genes BRCA1 and BRCA2 using five methods: results from a population-based study of young breast cancer patients. *Breast Cancer Research*, 10, R19.

Lewis, B.P., Burge, C.B., & Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, 120, 15–20.

Li, W.H., Wu, C.I., & Luo, C.C. (1984). Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. *Journal of Molecular Evolution*, 21, 58–71.

Liu, D.Y., Liu, M.L., Garrett, C., & Baker, H.W. (2007). Comparison of the frequency of defective sperm-zona pellucida (ZP) binding and the ZP-induced acrosome reaction between subfertile men with normal and abnormal semen. *Human Reproduction*, 22, 1878–1884.

Mandal, A., Klotz, K.L., Shetty, J., Jayes, F.L., Wolkowicz, M.J., Bolling, L.C., et al. (2003). SLLP1, a unique, intra-acrosomal, non-bacteriolytic, c lysozyme-like protein of human spermatozoa. *Biology of Reproduction*, 68, 1525–1537.

Marrero, M.A. & Ory, S.J. (1991). Unexplained infertility. *Current Opinion in Obstetrics & Gynecology*, 3, 211–218.

Miller, S.A., Dykes, D.D., & Polesky, H.F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research*, 16, 1215.

Ng, P.C. & Henikoff, S. (2001). Predicting deleterious amino acid substitutions. *Genome Research,* 11, 863–874.

Ng, P.C. & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research,* 31, 3812–3814.

Pasten-Hidalgo, K., Hernandez-Rivas, R., Roa-Espitia, A.L., Sanchez-Gutierrez, M., Martinez-Perez, F., Monrroy, A.O., et al. (2008). Presence, processing, and localization of mouse ADAM15 during sperm maturation and the role of its disintegrin domain during sperm-egg binding. *Reproduction,* 136, 41–51.

Pellicer, A., Albert, C., Mercader, A., Bonilla-Musoles, F., Remohí, J., & Simón, C. (1998). The follicular and endocrine environment in women with endometriosis: local and systemic cytokine production. *Fertility and Sterility*, 70, 425–431.

Quaas, A. & Dokras, A. (2008). Diagnosis and treatment of unexplained infertility. *Reviews in Obstetrics and Gynecology*, 1, 69–76.

© 2014 The British Fertility Society

PROOF ONLY

Ramensky, V., Bork, P., & Sunyaev, S. (2002). Human non-synonymous SNPs: server and survey. *Nucleic Acids Research*, 30, 3894–3900.

Rudd, M.F., Williams, R.D., Webb, E.L., Schmidt, S., Sellick, G.S., & Houlston, R.S. (2005). The predicted impact of coding single nucleotide polymorphisms database. *Cancer Epidemiology, Biomarkers & Prevention*, 14, 2598–2604.

Schug, J. & Overton, G.C. (1997). TESS: Transcription Element Search Software on the WWW CBIL-TR-1997-1001-v0.0.

Shi, Y.Y. & He, L. (2005). SHEsis, a powerful software platform for analyses of linkage disequilibrium, haplotype construction, and genetic association at polymorphism loci. *Cell Research,* 15, 97–98.

[AQ8]

Supplementary material available online

Supplementary Tables I–IX.

Wagner, A. (2009). The Role of *SPRASA* in Fertility Obstetrics and Gynaecology. 2009. University of Auckland, Auckland, pp. 247.

Xi, T., Jones, I.M., & Mohrenweiser, H.W. (2004). Many amino acid substitution variants identified in DNA repair genes during human population screenings are predicted to impact protein function. *Genomics*, 83, 970–979.

Zhang, K., Gao, R., Zhang, H., Cai, X., Shen, C., Wu, C., et al. (2005). Molecular cloning and characterization of three novel lysozyme-like genes, predominantly expressed in the male reproductive system of humans, belonging to the c-type lysozyme/alpha-lactalbumin family. *Biology of Reproduction*, 73, 1064–1071.

PROOF ONLY

THUF_A_907506.indd 8 4/10/2014 1:49:51 PM

Supplementary material for Prendergast, D., et al. (2014). *SPACA3* gene variants in a New Zealand cohort of infertile and fertile couples, Human Fertility, doi: 10.3109/14647273.2014.907506.

Supplementary Table I. Infertility diagnosis, duration of infertility and age of the infertile couples analysed in this study.

Classification	Sex/diagnosis	Range (mean)	Number/percentage
Age (years)	Female	24–42 (35.0)	n = 102
	Male	26–52 (36.8)	n = 102
Duration of infertility (years)		1–14 (4)	
Male factor infertility	Oligospermia		n = 38
	Teratospermia		37.3%
	Asthenospermia		
	Testicular cancer		
	Vasectomy reversal		
	Antisperm antibodies		
Female factor infertility	Endometriosis		n = 17
	PCOS		16.7%
	Raised FSH		
Couples with both a female and male factor infertility			n = 18
			17.6%
Couples with unexplained infertility			n = 29
			28.4%

The percentages indicate the number of male, females or couples with a diagnosed cause for their infertility. Male n = 102, female n = 102 and couples n = 102.
Oligospermia, teratospermia and asthenospermia as defined by World Health Organization standards.
PCOS = polycystic ovary syndrome, FSH = follicle stimulating hormone.

Supplementary Table II. Age of the fertile couples analysed in this study.

Diagnosis	Range (mean)
Female	23–46 (35)
Male	23–51 (37)

Male n = 104, female n = 104 and couples n = 104.

Supplementary Table III. Summary of PCR primers used for mutation screening and promoter amplification.

Fragment	Size	Primers (5′_3′)	Location
Exon 1	253 bp	FOR GTGGCGCTGTTTGTGGAAGATGAG	60349–60372
		REV TGTTAACAGCCCCAGGAAGAAGG	60556–60578
Exon 2	544 bp	FOR GCCTTCTGCCCACCCCTTCTCT	63939–36960
		REV GGTTGCCTGCTGCCTGCCTACT	64461–64482
Exon 3	223 bp	FOR CTGGGGTGGCTGTAACCATCTGAC	65360–65383
		REV AGGCCCAGCTACCTGAGCAGTA	65561–65582
Exon 4	268 bp	FOR ACAGGATTGGATTTAGGCGAGTGG	65857–65882
		REV ACCGCTGCGGGGCTCCAG	66108–66125
Exon 5	366 bp	FOR GTGGGCAGCAGCAGGGAACAAAC	66141–66163
		REV AACGGAGGTGCTCTGGCTCTGACA	66483–66506
Promoter A	983 or 986	FOR TATATAACGCGTCCAACACTACCATCCCTG	59514–59531
		REV ATATACTCGAGGTGACAATGGCAGCAGCA	60479–60496
ProInt	616	FOR CTATTCTGGGCACCAACCA	59881–59898
Promoter A		REV ATATACTCGAGGTGACAATGGCAGCAGCA	60479–60496
Promoter A	537	FOR TATATAACGCGTCCAACACTACCATCCCTG	59514–59531
ProInt		REV GGGATGTTAACAGGTGTGCAA	60030–60050
Promoter BF	926	FOR GGTATAACGCGTGGTGAGGACAAGAAGCAG	63224–63241
Promoter BR		REV TATATCTCGAGATCCCAGCTGGGCACCA	64133–64149

Forward (FOR) and reverse (REV) PCR primers spanning the five exons of the SPRASA gene and the promoter region. Size of PCR products generated by each set of primer pairs is indicated. Locations of fragments are indicated with reference to the reverse complements of Genbank accession number AC003687. Bases flanking restrictions sites Mlu I and Xho I are highlighted in grey, the restriction sites Mlu I and Xho I are in green and yellow respectively and the *SPACA3* gene sequence is in black.

PROOF ONLY

© 2014 The British Fertility Society

THUF_A_907506.indd Sec1:1 4/10/2014 1:49:51 PM

Supplementary Table IV. Summary of PCR conditions.

Cycle	Step	Temperature & Duration	Function
One (1x)		94.0°C for 5.0 minutes	Initial denaturation
Two (30x)	One	94.0°C for 1.0 minutes	Denaturation
	Two	64.5°C for 1.0 minutes	Annealing
	Three	72.0°C for 1.0 minutes	Elongation
Three (1x)		72.0°C for 10.0 minutes	Final elongation
Four (1x)		7.0°C for ∞	Hold

Supplementary Table V. Summary of Touchdown PCR conditions.

Cycle	Step	Temperature & Duration	Function
One (1x)		94.0°C for 5 minutes	Initial denaturation
Two (20x)	One	94.0°C for 0.45 minutes	Denaturation
	Two	70.0°C for 0.45 minutes	Annealing
		Decrease temperature after cycle one by 0.5°C every 1 cycle	Annealing
	Three	72.0°C for 1.0 minutes	Elongation
Three (15x)	One	94.0°C for 0.45 minutes	Denaturation
	Two	60.0°C for 0.45 minutes	Annealing
	Three	72.0°C for 1.0 minutes	Elongation
Four (1x)		72.0°C for 10.0 minutes	Final elongation
Five (1x)		7.0°C for ∞	Hold

Due to non-specific primer binding, exon two, four and the promoter were amplified by Touchdown PCR 70°C–60°C.[1] The promoter amplicons followed the same Touchdown PCR protocol as exon two and four with the exception that the time for the first 20 cycles was extended to one minute.

[1]Dieffenbach CW, Dveksler GS (eds.): PCR Primer: A Laboratory Manual, First Edition: Cold Spring Harbor Laboratory Press; 1995.

Supplementary Table VI. Summary of sequencing reaction conditions.

Cycle	Step	Temperature & Duration	Function
One (1x)		96.0°C for 1.0 minutes	Initial denaturation
Two (25x)	One	96.0°C for 0.1 minutes	Denaturation
	Two	50.0°C for 0.05 minutes	Annealing
	Three	60.0°C for 1.0 minutes	Elongation
Three (1x)		15.0°C for ∞	Hold

Primers (5μM) used for sequencing the human SPRASA exons were the forward PCR primers for exons one, two, three and five. Due to the presence of repetitive sequences in the 5′ end of the amplicon, the reverse primer for exon four was used.

PROOF ONLY

THUF_A_907506.indd Sec1:2 4/10/2014 1:49:51 PM

Supplementary Table VII. Variant frequencies of g.–22TGC(4_5) and c.239G>A in fertile and infertile couples.

Variant frequencies in couples				
Variant combination		Number		
Male	Female	Fertile	Infertile	p-values
Wildtype	Wildtype	69	56	0.43
Wildtype	5x	11	17	0.32
Wildtype	A	1	2	0.64
Wildtype	5x/A	1	0	0.50
5x	Wildtype	13	15	0.84
A	Wildtype	1	2	0.62
5x/A	Wildtype	2	0	0.25
5x	5x	4	3	1.00
5x	5x/A	1	0	0.50
5x	A	0	2	0.25
5x/A	5x	0	3	0.25
A	5x	1	1	1.00
5x/A	A	0	1	0.50
5x/A	5x/A	0	0	–
A	A	0	0	–
A	5x/A	0	0	–

Possible combinations and the frequencies of the 5x TGC repeat (g.–22TGC(4_5)) and the A allele (c.239G>A) in the fertile and infertile couples.
Statistics were performed with Fisher's exact t-test (2-tailed).

Supplementary Table VIII. Diagnosis of infertility in infertile patients with identified *SPACA3* variants.

	g.–22TGC(4_5)		c.239G>A		c.691G>C	
Diagnosis	HO	HE	HO	HE	HO	HE
Male factor infertility[a]	0	9	0	5	0	0
Unexplained (male)	4	11	0	2	0	0
Endometriosis	0	4	1	2	0	0
PCOS	1	3	0	0	0	0
Unexplained (female)	0	19	0	2	0	1
TOTAL	5	47[b]	1	11	0	1

[a]Testicular cancer, oligospermia, antisperm antibodies, teratospermia or asthenospermia.
[b]Forty six individuals were heterozygous for g.–22TGC(4_5), one female was diagnosed with both endometriosis and PCOS.
PCOS = polycystic ovary syndrome; HO = homozygote, HE = heterozygote.

PROOF ONLY

© 2014 The British Fertility Society

THUF_A_907506.indd Sec1:3 4/10/2014 1:49:51 PM

Supplementary Table IX. The genotype analysis of the three variants identified in the infertile and fertile cohorts.

			Genotype analysis						
		Cohort	4_4	4_5	5_5	5X allele	OR	95% CI	P Value
g.–22TGC(4_5)	Male	Fertile Controls	84 (0.808)	18 (0.173)	2 (0.019)	22 (0.106)	0.71	0.40–1.29	0.55
		Infertile Cases	77 (0.755)	21 (0.206)	4 (0.039)	29 (0.142)			
	Female	Fertile Controls	86 (0.827)	17 (0.163)	1 (0.010)	19 (0.091)	0.66	0.35–1.23	0.35
		Infertile Cases	76 (0.745)	25 (0.245)	1 (0.010)	27 (0.132)			
	Couples	Fertile Controls	170 (0.817)	35 (0.168)	3 (0.014)	41 (0.099)	0.69	0.45–1.05	0.24
		Infertile Cases	153 (0.750)	46 (0.225)	5 (0.025)	56 (0.137)			
c.239G>A			GG	GA	AA	A allele			
	Male	Fertile Controls	100 (0962)	4 (0.038)	0 (0.000)	4 (0.019)	0.55	0.16–1.1	0.34
		Infertile Cases	95 (0.931)	7 (0.069)	0 (0.000)	7 (0.034)			
	Female	Fertile Controls	101 (0.971)	3 (0.029)	0 (0.000)	3 (0.014)	0.48	0.12–1.96	0.55
		Infertile Cases	97 (0.951)	4 (0.039)	1 (0.010)	6 (0.029)			
	Couples	Fertile Controls	201 (0.966)	7 (0.034)	0 (0.000)	7 (0.017)	0.52	0.21–1.32	0.36
		Infertile Cases	192 (0.941)	11 (0.054)	1 (0.005)	13 (0.032)			
c.691G>C			GG	GC	CC	C allele			
	Male	Fertile Controls	104 (1.000)	0 (0.000)	0 (0.000)	0 (0.000)			1.00
		Infertile Cases	102 (1.000)	0 (0.000)	0 (0.000)	0 (0.000)			
	Female	Fertile Controls	104 (1.000)	1 (0.010)	0 (0.000)	0 (0.000)			0.31
		Infertile Cases	201 (0.990)	0 (0.000)	0 (0.000)	1 (0.005)			
	Couples	Fertile Controls	208 (1.000)	0 (0.000)	0 (0.000)	0 (0.000)			0.31
		Infertile Cases	203 (0.995)	1 (0.005)	0 (0.000)	1 (0.002)			

OR = odds ratio; CI = confidence interval

PROOF ONLY

THUF_A_907506.indd Sec1:4 4/10/2014 1:49:51 PM