



HAL
open science

STAG3 homozygous missense variant causes primary ovarian insufficiency and male non-obstructive azoospermia

Sylvie Jaillard, Kenneth Mcelreavy, Gorjana Robevska, Linda Akloul, Farah Ghieh, Rajini Sreenivasan, Marion Beaumont, Anu Bashamboo, Joelle Bignon-Topalovic, Anne-Sophie Neyroud, et al.

► **To cite this version:**

Sylvie Jaillard, Kenneth Mcelreavy, Gorjana Robevska, Linda Akloul, Farah Ghieh, et al.. STAG3 homozygous missense variant causes primary ovarian insufficiency and male non-obstructive azoospermia. *Molecular Human Reproduction*, 2020, 26 (9), pp.665-677. 10.1093/molehr/gaaa050 . hal-02929984

HAL Id: hal-02929984

<https://hal.science/hal-02929984v1>

Submitted on 11 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

STAG3 homozygous missense variant causes primary ovarian insufficiency and male non-obstructive azoospermia

Running title: *STAG3* variants in female and male infertility

Sylvie Jaillard^{1,2,3}, Kenneth McElreavy⁴, Gorjana Robevska¹, Linda Akloul⁵, Farah Ghieh⁶, Rajini Sreenivasan¹, Marion Beaumont³, Anu Bashamboo⁴, Joelle Bignon-Topalovic⁴, Anne-Sophie Neyroud⁸, Katrina Bell¹, Elisabeth Veron-Gastard⁸, Erika Launay³, Jocelyn van den Bergen¹, Bénédicte Nouyou³, François Vialard^{6,7}, Marc-Antoine Belaud-Rotureau^{2,3,8}, Katie L Ayers^{1,9}, Sylvie Odent⁵, Célia Ravel^{2,8}, Elena J Tucker^{1,9*}, Andrew H Sinclair^{1,9*}

- 1- Murdoch Children's Research Institute, Royal Children's Hospital, Melbourne, VIC Australia
- 2- Univ Rennes, CHU Rennes, INSERM, EHESP, IRSET (Institut de recherche en santé, environnement et travail) – UMR_S 1085, F-35000 Rennes, France
- 3- CHU Rennes, Service de Cytogénétique et Biologie Cellulaire, F-35033 Rennes, France
- 4- Institut Pasteur, CNRS – UMR_3738, Paris, France
- 5- CHU Rennes, Service de Génétique Clinique, CLAD Ouest, F-35033 Rennes, France
- 6- Université Paris-Saclay, UVSQ-INRA-ENVA, UMR-BREED, Montigny le Bretonneux, 78180, France
- 7- Fédération de Génétique, Laboratoire de Biologie Médicale, CHI de Poissy-St Germain en Laye, Poissy, 78300, France
- 8- CHU Rennes, Service de Biologie de la Reproduction-CECOS, F-35033 Rennes, France
- 9- Department of Paediatrics, University of Melbourne, Melbourne, VIC Australia

*These authors should be regarded as co-senior authors of this study

Corresponding authors: Sylvie Jaillard, sylvie.jaillard@chu-rennes.fr; Elena Tucker, elena.tucker@mcri.edu.au

Abstract

Infertility, a global problem affecting up to 15% of couples, can have varied causes ranging from natural aging to the pathological development or function of the reproductive organs. One form of female infertility is premature ovarian insufficiency (POI), affecting up to 1 in 100 women and characterised by amenorrhea and elevated follicle stimulating hormone before the age of 40. POI can have a genetic basis, with over 50 causative genes identified. Non-obstructive azoospermia (NOA), a form of male infertility characterised by the absence of sperm in semen, has an incidence of 1% and is similarly heterogeneous. The genetic basis of male and female infertility is poorly understood with the majority of cases having no known cause. Here, we study a case of familial infertility including a proband with POI and her brother with NOA. We performed whole-exome sequencing (WES) and identified a homozygous *STAG3* missense variant that segregated with infertility. *STAG3* encodes a component of the meiosis cohesin complex required for sister chromatid separation. We report the first pathogenic homozygous missense variant in *STAG3* and the first *STAG3* variant associated with both male and female infertility. We also demonstrate limitations of WES for the analysis of homologous DNA sequences, with this variant being ambiguous or missed by independent WES protocols and its homozygosity only being established via long-range nested PCR.

Keywords

premature ovarian insufficiency, non-obstructive azoospermia, whole-exome sequencing, *STAG3*

Introduction

Premature ovarian insufficiency (POI) represents one of the main causes of female infertility with a prevalence of 1-3%, depending on population characteristics such as ethnicity. This condition is characterised by the occurrence of menstrual disturbance (primary or secondary amenorrhea or oligomenorrhea) for at least four months before the age of 40 with high follicle stimulating hormone (FSH) levels (> 25 UI/l on two occasions > 4 weeks apart) and low estradiol levels (European Society of Human Reproduction and Embryology <https://www.eshre.eu/>). The mechanism leading to POI can be an impaired formation of primordial follicles leading to a reduced number of their pool, an impaired recruitment and/or an altered maturation of the follicles, and/or an increased follicular atresia (Laissue, 2015). POI can be caused by infections, medical treatments, or metabolic or autoimmune diseases, but often has a genetic basis with over 50 causative genes reported, affecting various processes such as gonadal development, meiosis, DNA repair, folliculogenesis, hormonal signaling, steroidogenesis, metabolism, mitochondrial function and immune regulation (Tucker *et al.*, 2016). Such a genetic cause can be identified in 10-15% of the POI patients (Caburet *et al.*, 2014). Depending on the gene involved, POI can occur as part of a syndromic disorder (e.g. *AIRE*, *ATM*, *FOXL2*) or as an isolated disease (non-syndromic POI: e.g. *NOBOX*, *GDF9*). In order to decipher the genetic heterogeneity of POI, recent studies using massive parallel sequencing (MPS) approaches (targeted panels and whole-exome sequencing (WES)), mainly in individuals with non-syndromic POI, have reported single nucleotide variants in novel dominant or recessive candidate genes such as *MCM9* (Fauchereau *et al.*, 2016), *POLR2C* (Moriwaki *et al.*, 2017), *NUP107* (Ren *et al.*, 2018) or *TP63* (Tucker *et al.*, 2019). *STAG3* was first described as a POI gene in 2014 (Caburet *et al.*, 2014), and recessive high impact variants have since been described as a rare but recurrent

cause of non-syndromic POI (Caburet *et al.*, 2014; Le Quesne Stabej *et al.*, 2016; Colombo *et al.*, 2017; He *et al.*, 2018). *STAG3* is a key gene essential for meiosis, and is required for gametogenesis and fertility. Its implication in male infertility has also been strongly suggested (Le Quesne Stabej *et al.*, 2016), and two male patients with non-obstructive azoospermia (NOA) and biallelic high impact variants in *STAG3* have recently been reported (Riera-Escamilla *et al.*, 2019; van der Bijl *et al.*, 2019). We report here the first description of a homozygous missense *STAG3* variant associated with both female and male infertility, in a woman with POI (primary amenorrhea) and her infertile brother with NOA. This study confirms the recent data expanding the phenotypic spectrum of variants within *STAG3* to include male infertility, and demonstrates the involvement of missense variants in contrast to loss-of-function variants previously reported.

Materials and methods

Patients

The proband (Figure 1, II-3), born in 1984, was first seen at 15 years old for delayed puberty. At this time, her height was 164.1 cm, weight was 61.2 kg and secondary sex characteristics were underdeveloped (Tanner I-II for breast, I for pubic and axillary hair). Genitals were normal and ultrasonographic examination showed a small uterus; the right ovary was present with a cyst suggesting a residual retained function and the left ovary was not visualised. Hormonal assessment evaluated growth factors (normal IGF-1 and IGFBP-3, sub-normal GH stimulation test by ornithine), thyroid hormones (normal T4 and TSH, and TRH stimulation test), prolactin (normal) and reproductive hormones. High levels were detected for follicle stimulating hormone (FSH: 46.5 UI/l) and luteinising hormone (LH: 24.5 UI/l) with peaks at 78.5 UI/l and 87.7 UI/l respectively after gonadotrophin releasing hormone (GnRH)

stimulation. Decreased levels were noted for estradiol (33 pg/ml) and sex hormone-binding globulin (SHBG: 23 nmol/l, N: 28-80 nmol/l). Bone age was 12 years old and three months. Autoimmunity was discounted (negative thyroid antibodies, ovarian antibodies, ACA/21-OH antibodies). Karyotype showed 46,XX constitution. Estrogen therapy was initiated, and the patient was evaluated six months later. Development of the secondary sex characteristics started (Tanner III for breast, II for pubic hair and I for axillary hair), bone age was 13 years old, hormonal assessment showed FSH: 43.1 UI/l, LH: 4.1 UI/l, estradiol: 26 pg/ml, and normal testosterone. Estrogen therapy was progressively increased and hormone replacement therapy was initiated in 2002. Other medical history included appendectomy and surgical treatment for inguinal hernia, and hysteroscopia which revealed endometrial hypotrophy but no malformation. Repeat ultrasonographic assessments in 2004 and 2008 confirmed uterine hypoplasia, small sized right ovary with some small follicles, and atrophic left ovary. She was enrolled in an oocyte donation program in 2013 and had a child in 2017. At this time, height was 1.76 m and weight was 80 kg. When last seen in 2019, she was pregnant following another oocyte donation from the same woman.

In her familial history (Figure 1), her parents (I-1 and I-2) are healthy and consanguineous (her mother's great grandfather and her father's grandfather are brothers). Maternal menopause occurred at 50-year-old. She has two brothers, and a third (II-2) who died at four months (sudden infant death). One of the brothers (II-1) also has reproductive problems with NOA, and had two children after assisted reproductive techniques using donor sperm. Assessment included hormonal testing with normal FSH (11.13 UI/l), testosterone (6.84 µg/l), prolactin (8.68 ng/ml) and thyroid stimulation hormone (TSH) (1.21 UI/l), and karyotype (46,XY). The other brother (II-4) is fertile and had two children following spontaneous pregnancies. No other history of infertility is noted in the family.

Written informed consent was obtained from all participants. All procedures were in accordance with the ethical standards of the Ethics Committee of Rennes University Hospital and the French law.

Detection of the variants by sequencing

Whole-exome sequencing (WES)

- WES assays

DNA from the proband underwent WES at the Australian Genome Research Facility (AGRF). Library preparation was performed with Agilent SureSelect Human All Exon V6 (Agilent Technologies, Santa Clara, CA, USA) and sequencing was with the NovaSeq™ 6000 Sequencing System (Illumina Inc., San Diego, CA, USA). All WES data were processed using the Cpipe pipeline (Sadedin *et al.*, 2015) designed according to the GATK guidelines and deposited into SeqR for analysis (<https://seqr.broadinstitute.org/>).

The proband and her family were also previously tested by exome sequencing, which was performed as described elsewhere (Murphy *et al.*, 2015). Briefly, exon enrichment was performed using Agilent SureSelect Human All Exon V4 (Agilent Technologies, Santa Clara, CA, USA). Paired-end sequencing was performed on the Illumina HiSeq2000 platform (Illumina Inc., San Diego, CA, USA) with an average sequencing coverage of x50. Read files were generated from the sequencing platform via the manufacturer's proprietary software. Reads were mapped using the Burrows–Wheeler Aligner and local realignment of the mapped reads around potential insertion/deletion (indel) sites was carried out with the GATK version 1.6. SNP and indel variants were called using the GATK Unified Genotyper for each sample. SNP novelty was determined against dbSNP138. Datasets were filtered for novel or rare (MAF<0.01) variants.

- WES analysis

We performed two phases of analysis for WES performed with Agilent SureSelect Human All Exon V6, based on our previously described method (Tucker *et al.*, 2019), with the first focused on gene priority and the second focused on variant priority. For gene-centric analysis, we considered the potential pathogenicity of all “moderate to high impact” coding variants within diagnostic or candidate genes (gene list as per Tucker *et al.*, 2019). For variant-centric analysis, we considered the potential pathogenicity of “high impact” variants in any gene, and “moderate to high impact” recessive-type variants in any gene. High priority variants are rare (<0.01 MAF) variants affecting essential splice sites, introducing frameshifts or premature stop codons, whereas moderate priority variants are rare (<0.01 MAF) missense variants and in-frame codon deletions/insertions. MAF and tolerance of genes to missense and/or loss-of-function variation were assessed in the public databases ExAC (<http://exac.broadinstitute.org/>) and gnomAD (<https://gnomad.broadinstitute.org/>).

STAG3 variants observed by sequencing were described based on the Genbank NM_012447.3 cDNA reference sequence and NP_036579.2 protein reference sequence, according to the Sequence Variant Nomenclature for the Human Genome (HGVS nomenclature) (<http://varnomen.hgvs.org/>). For greater clarity, the variants are reported as c. instead of NM_012447.3:c. and p. instead of NP_036579.2:p.

Sanger sequencing

Long-range PCR was performed to amplify specifically the genomic region of interest, with the Expand Long Range, dNTPack kit (Roche Life Science, Penzberg, Germany). The PCR mix included 10 µl of Expand Long Range Buffer with MgCl₂, 2.5 µl of PCR Nucleotide Mix, 0.25 µM of forward primer (CACTCTACCTAGGTTTCACTTCTGC), 0.25 µM of reverse primer (AATCTATCTCCCTACCTTCTCCAAGC), 6% DMSO, 0.7 µl of Expand Long Range Enzyme mix

and 150 ng of patient genomic DNA in a final volume of 50 µl. PCR conditions were an initial denaturation at 94°C for 2 min, 35 cycles with denaturation at 92°C for 10 sec, annealing at 62°C for 15 sec, elongation at 68°C for 8 min, and a final elongation at 68°C for 7 min. The PCR product was visualised by electrophoresis on a 1% agarose gel (size 4,387 bp). Samples were analysed by direct sequencing using the ABI Prism Big Dye Terminator Cycle Sequencing Ready Reaction kit and the ABI 3100 Genetic Analyser (Applied Biosystems, CA, USA). Sanger sequencing was performed (AGRF) with primers surrounding a 327 bp region including the variant (forward: TCCATGAGAGGGAGTTATCTGGT / reverse: GGGTTCATCATTGGTCACTAGGT). DNA sequences were aligned using the Benchling tool (<https://benchling.com/>).

Functional studies

In-silico analyses

The effects of the missense variant identified were assessed using HOPE database (Venselaar *et al.*, 2010) (<https://www.cmbi.ru.nl/hope/>).

Variant pathogenicity was predicted in silico using Polyphen-2 (<http://genetics.bwh.harvard.edu/pph2/>), Mutation Taster (<http://www.mutationtaster.org/>), SIFT/Provean (<http://provean.jcvi.org/>), DANN (Deleterious Annotation of genetic variants using Neural Networks) score (Quang *et al.*, 2015) and CADD (Combined Annotation-Dependent Depletion) score (<https://cadd.gs.washington.edu/snv>). The conservation of affected nucleotides and residues in mammals, birds and fish was also taken into account with GERP (Genomic Evolutionary Rate Profiling) score and Multiz Alignments of 100 vertebrates (UCSC Genome Browser <https://genome.ucsc.edu/>). Variants in known diagnostic genes were curated using

the American College of Medical Genetics and Genomics (ACMG) criteria (Richards *et al.*, 2015).

Testis biopsy staining

Immunohistochemistry (IHC) was performed using the Benchmark XT Ventana Roche system with the XT ultraView DAB v3 revelation kit (Roche Life Science, Penzberg, Germany). First, 4 µm sections of each testis block were made using a microtome to obtain fine paraffin ribbons. These ribbons were deposited on IHC slides (SuperFrost Plus type 25x75x1.0 mm, Thermo Fisher Scientific, Waltham, MA, USA). The slides were then dried at 56°C for 24 hours to ensure good adhesion of the tissue to the slide before staining. After inhibition of endogenous peroxidases, sections of testicular tissue from a fertile control and from the infertile brother were incubated with the primary STAG3 antibody (1:20 dilution of rabbit polyclonal antibody) (PA5-63556, Invitrogen, Carlsbad, CA, USA) for 40 min. The primary antibody was detected with a biotinylated secondary antibody. Staining was visualised with the use of 3,3'-diaminobenzidine and hematoxylin as a counterstain. The slides were then dehydrated in successive alcohol baths and xylene so as to be covered with a coverslip stuck with pertex, a mounting medium.

Results

The filtering pipeline identified a total of 558 “moderate to high impact” variants in any gene. For gene-centric analysis, eight “moderate to high impact” variants were within diagnostic genes, and seven were within candidate genes. For variant-centric analysis, we identified 46 “high impact” variants in any gene, and 184 “moderate to high impact” recessive-type variants in any gene.

The most relevant variants within diagnostic genes included (Table 1): a heterozygous missense variant in **HS6ST1** (OMIM 604846, hypogonadotropic hypogonadism 15 with or without anosmia), a homozygous missense variant in **STAG3** (OMIM 608489, autosomal recessive premature ovarian failure 8), a heterozygous missense variant in **RECQL4** (OMIM 603780, Baller-Gerold / Rothmund-Thomson / RAPADILINO syndromes), compound heterozygous missense variants in **ATM** (OMIM 607585, autosomal recessive ataxia-telangiectasia), a heterozygous variant in **RCBTB1** (OMIM 607867, autosomal recessive retinal dystrophy with or without extraocular anomalies), a heterozygous splice site variant in **RSPH1** (OMIM 609314, autosomal recessive primary ciliary dyskinesia 24), and a heterozygous missense variant in **AIRE** (OMIM 607358, autosomal dominant or recessive autoimmune polyendocrinopathy syndrome type I, with or without reversible metaphyseal dysplasia). Six genes were ruled out because they were unlikely to be causal for the phenotype based on their inheritance (heterozygous variant involved in recessive disease: *RECQL4*, *RCBTB1*, *RSPH1* or inheritance from one parent), and/or the absence of phenotypic match due to their association with syndromic POI (*ATM*, *AIRE*, *RECQL4*, *RCBTB1*) or with a discordant phenotype (*HS6ST1*), and/or their unlikely pathogenic prediction (*HS6ST1*, *RECQL4*). For *ATM* variants, a compound heterozygous status was observed with one variant inherited from each parent, but both variants were curated as benign and the phenotype was not an exact match so they are not further discussed. Further interesting variants in candidate genes are listed in Table 1.

A homozygous missense variant in *STAG3* NM_012447.3:c.962G>A / NP_036579.2:p.(Arg321His) (NC_000007.13:g.99794799G>A, GRCh37) was predicted to be “likely pathogenic” after variant curation using ACMG-based criteria. This variant was submitted to ClinVar (ClinVar accession SCV000995939). A deletion on one allele making

the *STAG3* variant appear homozygous was discounted with microarray and manual inspection of the sequencing coverage with Integrative Genomics Viewer (IGV). It is a rare variant not present in ExAC or in gnomAD and not referenced in ClinVar. At the protein level, the variant lies within a highly conserved region and is thus probably damaging, which is in accordance with the in-silico predictions (probably damaging in PolyPhen-2, disease causing in Mutation Taster, damaging in SIFT/Provean, and DANN, pathogenic CADD score). The *STAG3* protein contains a STAG domain and a stromalin conservative domain (SCD) (86 amino acid motif present in all proteins of the family). Two armadillo repeats (ARM), consisting of superhelical structures, are suitable for binding large substrates such as proteins and nucleic acids. One of the ARM repeats is located in the SCD (Figure 2). The variant introduces a smaller amino acid with neutral charge in place of the positively charged wild-type amino acid within the ARM repeat of the SCD. This can disturb the domain with loss of interaction with other molecules, abolishing the protein's function (HOPE database with the UniProtKB Q9UJ98 entry).

Interestingly, on IGV, many sequencing reads across this variant did not uniquely align to the genome, with a number of sequencing reads aligning non-specifically, thereby indicating a low mapping quality corresponding to highly repetitive and homologous genomic region (Supp Figure 1B). In fact, the family was studied independently by two groups, with WES performed in two different centres and data being analysed using two different bioinformatics pipelines. One group failed to detect the variant due to a capture with no baits targeting this exon of *STAG3* (Supp Figure 1A, C). Initial attempts at sanger validation cast further doubt on the variant being a true variant. Initial sanger sequencing indicated a heterozygous variant at this site (Supp. Figure 2). However, this erroneous sequencing was due to poor specificity of the sequencing primers. Persisting, we attempted a long-range

nested PCR. The first PCR was performed with primers that captured a very large but specific region of genomic DNA, followed by nested PCR closely flanking the variant site. There were no common/reported SNPs affecting the binding site of any of the primers, making allele-specific drop-out unlikely. Allele-specific drop-out was also discounted by sequencing parental DNA and demonstrating heterozygosity, indicating successful amplification of both alleles. This long-range nested PCR validated that the patient was truly homozygous for the *STAG3* missense variant (Supp. Figure 2). The difficulty in validating the variant and the repeated transient belief that the variant was a false positive indicates the need to carefully investigate and persist when reads do not align uniquely to the genome. Familial studies targeting *STAG3* showed that the variant was present in a heterozygous state in the parents and the healthy brother whereas it was present in a homozygous state in the infertile brother (Figure 1). This familial segregation with infertility adds weight to the argument that this variant is causative.

For the infertile brother bearing the homozygous *STAG3* variant, testis biopsy showed the conservation of the testicular tissue architecture. On the right and on the left, the presence of germ cells was noted in the seminiferous tubules, but no spermatozoa were identified, indicating incomplete spermatogenesis, not beyond the spermatocyte stage and corresponding to a spermatogenic arrest. IHC staining performed on testis biopsies (Figure 3) showed that in the control testis, *STAG3* protein was present in the seminiferous tubules in the nucleus of spermatogonia and spermatocytes. In the testicular biopsies from the NOA brother, *STAG3* expression was undetectable in the seminiferous tubules.

Discussion

Homozygous *STAG3* missense variant associated with POI

In our study, we identified the first likely pathogenic homozygous missense variant in *STAG3* in a patient presenting with isolated POI and her brother with NOA in a consanguineous family. *STAG3* encodes a meiosis-specific component of the cohesion ring. Other meiosis-specific subunits include, *SMC1 β* , *RAD21L*, and *REC8*. The meiotic cohesion complex, wraps around the centromeres and arms of chromosomes and ensures sister chromatid cohesion. After cohesion cleavage, it plays a key role in proper chromosomal segregation during meiosis. Several distinct cohesion complexes are thought to exist, showing differences in spatiotemporal distribution throughout the meiotic divisions (Brooker and Berkowitz, 2014). Variants in genes involved in the cohesin network have been associated with developmental conditions known as cohesinopathies (e.g. Roberts syndrome, Cornelia de Lange syndrome). Recent studies by whole-exome sequencing in patients with POI have identified putative causative variants in *SMC1 β* and *REC8* (Bouilly *et al.*, 2016), and established causative variants in *STAG3*.

STAG3 causative variants include a homozygous 1 bp deletion (Caburet *et al.*, 2014), a homozygous 2 bp duplication (Le Quesne Stabej *et al.*, 2016), a homozygous nonsense variant (Colombo *et al.*, 2017), a homozygous splice site variant (He *et al.*, 2018), and two homozygous in-frame variants (Xiao *et al.*, 2019), which have been described in five different consanguineous pedigrees affected by POI (OMIM: Premature ovarian failure 8, #615723). Furthermore, compound heterozygous variants (missense and 1 bp deletion) have recently been identified in two sisters of a non-consanguineous family (Heddar *et al.*, 2019). Two additional loss-of-function variants have been identified in an adopted woman with POI, although compound heterozygosity was not confirmed (França *et al.*, 2019), and three other *STAG3* variants associated with POI are reported in the ClinVar database: one nonsense variant and two canonical splice-site variants without complete information on

their homozygous or heterozygous status (Table 2). In contrast to the *STAG3* variants previously reported in association with POI, our patient is the first case of a homozygous missense variant. We provide evidence that this *STAG3* variant impacts protein function by aberrant *STAG3* expression in testicular tissue from the affected brother. Furthermore, the strong phenotypic match between reported patients with *STAG3* biallelic variants (Table 3) strengthened the hypothesis of the involvement of the homozygous missense variant in POI. All patients bearing biallelic *STAG3* variants had indeed a severe POI phenotype with especially early onset POI, primary amenorrhea, undeveloped secondary sex characteristics, and streak gonads. The presence of streak gonads has been proposed to be a major feature in POI patients bearing a deleterious variant with high impact on protein. The patient of our study, similarly presented with primary amenorrhea and undeveloped secondary sex characteristics. However, she had development of at least one ovary albeit non-functioning, and had evidence of estrogen albeit at a low level, indicating that the missense variant may result in some residual *STAG3* activity and a slightly less severe phenotype. One of the 13 reported patients with confirmed or supposed biallelic *STAG3* variant had a germ-cell tumour (Caburet *et al.*, 2014), although there is not enough evidence to suggest a statistically significant increased risk. Additional monitoring for germ-cell tumour development in patients with *STAG3* variants may be warranted, and could improve patient prognosis. Functional studies involving inactivation of *Stag3* in a murine model were performed by Caburet *et al.* (Caburet *et al.*, 2014). Female *Stag3*^{-/-} mice had ovarian dysgenesis and were sterile, with fetal oocytes showing an arrest at early prophase I due to defects in synapsis, and in centromeric sister chromatid cohesion. *Stag3* was therefore proposed to be essential for the assembly of the meiotic cohesin ring and the synaptonemal complex. For the homozygous splicing variant identified by He *et al.* in two sisters with POI,

a truncated *STAG3* transcript was identified in mRNA extracted from patient's blood, confirming the deleterious impact of the variant (He *et al.*, 2018). Recently Xiao *et al.* confirmed the deleterious impact of two homozygous in-frame variants with aberrant REC8 localisation and loss of interactions with REC8 or SMC1A as a consequence the variants (Xiao *et al.*, 2019).

***STAG3* variants in male infertility**

Furthermore, male *Stag3*^{-/-} mice were also infertile (Caburet *et al.*, 2014). Male mice lacking *Stag3* showed a meiotic arrest with shortening of their chromosome axial element and loss of centromeric cohesion, and arrest of spermatogenesis at the first spermatocyte level, leading to NOA (Pezzi *et al.*, 2000; Llano *et al.*, 2014). *STAG3* loss-of-function is thus postulated to lead to infertility, not only in women, but also in men (Llano *et al.*, 2014). At the time of WES analysis, there had been no report of *STAG3* variants associated with human male infertility, however, two biallelic variants in *STAG3* leading to NOA have since been described (Riera-Escamilla *et al.*, 2019; van der Bijl *et al.*, 2019). In the first report, the patient was affected by complete bilateral meiotic arrest and carried two loss-of-function variants in *STAG3* (Figure 2). Further analyses performed on the patient's testis biopsy showed that meiotic entry occurred with normal frequency but was followed by a failure to complete chromosome pairing. Thus, spermatocytes were lost before they reached meiotic metaphase I. In the second report, the patient presented drastic aberrations in chromosome architecture, with a meiotic arrest during prophase I, and no cells developing beyond late zygotene (corresponding to spermatocyte I stage). This patient carried compound heterozygous variants in *STAG3* with one loss-of-function and one missense (van der Bijl *et al.*, 2019) (Figure 2). Interestingly, the proband of our study had an infertile brother with

NOA also bearing the *STAG3* variant in a homozygous state, with testis biopsy staining confirming reduced *STAG3* stability and an early meiotic arrest. This is the third male patient described, and the first with a homozygous missense variant, confirming the involvement of *STAG3* in male infertility manifesting as NOA. According to the classification established for meiotic arrest in human male meiosis (Jan *et al.*, 2018), *STAG3* biallelic variants are associated with type I meiotic arrest where severe asynapsis of the homologous chromosomes is a major feature, as observed in our patient. We report for the first time a *STAG3* variant leading to POI and NOA in a single pedigree, strengthening the existence of shared genetic factors causing both these infertility conditions.

Impact of disturbance of the stromalin conservative domain (SCD)

STAG genes (*STAG1*, *STAG2* and *STAG3*) encode members of the highly conserved family of stromalin nuclear proteins harbouring particular structural elements: 1) the stromalin conservative domain or SCD, an 86 amino acid region highly conserved from yeast to humans, 2) the STAG domain present in *Schizosaccharomyces pombe* mitotic cohesin Psc3, and the meiosis specific cohesin Rec11 (Ellermeier and Smith, 2005), 3) the amino-terminal and carboxyl-terminal domains, and 4) the armadillo-type fold corresponding to a superhelical structure adapted for binding large substrates such as proteins and nucleic acids. It is noteworthy that inconsistency was observed for the reported *STAG3* variants when collecting the molecular data of the literature, with some misleading information (e.g. the study of Colombo *et al.* describes a nonsense variant whereas a missense variant is indicated in the title of the article). We have amended the description of variants in the literature using HGVS nomenclature, as indicated in Table 2. Previous *STAG3* homozygous or compound heterozygous variants reported are located in the amino-terminal domain, the

STAG domain, the carboxyl-terminal domain or the ARM-type fold. It is also interesting to note a recurrence of the affected amino acid position 650 p.(Tyr650*) with two different loss-of-function substitutions resulting in the same protein damage (Le Quesne Stabej *et al.*, 2016; França *et al.*, 2019). All but four variants were predicted to generate nonsense-mediated decay, leading to *STAG3* loss-of-function. Of the remaining variants, functional experiments performed for two demonstrated also a complete loss of function, with absence of any binding to interacting proteins (Xiao *et al.*, 2019). The two missense reported variants were both associated with a severe variant in a compound heterozygous state (Heddar *et al.*, 2019; van der Bijl *et al.*, 2019). We report the first homozygous missense variant in the SCD. The strong conservation of this domain among the STAG proteins, and the pathogenic in-silico predictions, suggest that a missense variant at this position can deleteriously impact protein function.

***STAG3*-like genes lead to difficulties in identification of *STAG3* variants by massive parallel sequencing**

STAG3 is a meiosis-specific gene which consists of 34 exons and encompasses more than 30 kb of the genomic DNA. The *STAG3* gene is the precursor of a family of truncated genes or *STAG3*-like (*STAGL*) genes (*STAG3L1* to *STAG3L6*), all on chromosome 7, displaying 90-98% identity. These paralogs correspond to truncated copies of *STAG3* that have originated through genomic duplications from the ancestral *STAG3* gene, with most of them transcribed normally although the production of a functional protein is not demonstrated. For some of the *STAG3L* genes, 85% similarity to the middle part of *STAG3*, including the entire SCD, is predicted (Pezzi *et al.*, 2000). These data are in line with the difficulty in detecting the *STAG3* variants, which introduce a change in the ARM repeat of the SCD of the

protein. We can deduce that the variant identified in our patient is located in the similar region of *STAG3/STAG3L1-2-3*. The 124bp exon of *STAG3* in which the variant falls, has 100% identity with regions of *STAG3L1*, *STAG3L2* and *STAG3L3*. This led to failure of variant detection by one WES protocol and ambiguous results with another WES protocol as well as subsequent sanger sequencing. We solved the ambiguity by the development of a long-range nested PCR. For variants in multi-mapping regions such as these, the most conclusive data comes from long-range nested PCR, and not WES. Allele-specific dropout due to SNVs in primer-binding sites remains a theoretical possibility, however, in this case is unlikely given the lack of common/reported SNPs at primer-binding sites, the fact that both alleles successfully amplified in each heterozygous parent, the strong gene:phenotype match and the lack of *STAG3* in patient biopsy, providing functional evidence to support the involvement of this gene.

Conclusion

This study reports for the first time a *STAG3* missense variant leading to POI and NOA, strengthening the existence of shared genetic factors causing both these infertility conditions. Furthermore, this study highlights the role of meiosis-specific genes of the cohesion complex in POI pathogenesis but also in male infertility, particularly NOA. Interpretation of *STAG3* variants can be challenging due to a lack of uniformity of variant nomenclature in the literature, as well as the limitations of WES for detecting variants in repetitive regions of the genome. Exome sequencing is now a major tool allowing the elucidation of key genetic factors leading to infertility, however, it requires careful analysis when considering genes within homologous regions of DNA, such as *STAG3*. In such cases,

manual inspection of aligned sequences and complementary techniques such as long-range PCR should be performed to establish variant inheritance and pathogenicity.

Authors' Contributions

SJ, EJT, and AHS conceived and designed the study. LA, ASN, EVG, SO, and CR were involved in the patients care and evaluation. SJ, KM, and EJT acquired the data. SJ, EJT, KM, AB, KLA, and RS analysed and interpreted the data. MB, GR, EL and JVDB supplied technical support. KB and BN supplied bioinformatics support. FG and FV were involved in the acquisition of additional data. MABR, KLA, AHS and EJT participated in the project supervision. SJ and EJT wrote the manuscript. FV and AHS critically reviewed the manuscript and all authors read and approved the final manuscript.

Acknowledgments

We would like to thank the patients for taking part in our research.

Funding statement

This work was supported by CHU Rennes and Rennes 1 University, Faculty of Medicine in France, a Peter Doherty Early Career Fellowship (1054432; to EJT), a National Health and Medical Research Council program grant (1074258; to AHS), a fellowship (1062854; to AHS) from the Australian National Health and Medical Research Council, the Victorian Government's Operational Infrastructure Support Program, a research grant from the European Society of Pediatric Endocrinology (to AB), and the Agence Nationale de la Recherche (ANR), ANR-10-LABX-73 REVIVE and ANR-17-CE14-0038-01 (to KM).

Conflict of interests

The authors declare no conflicts of interest.

Accepted Manuscript

Figure Legends

Figure 1: Familial pedigree and *STAG3* genotype

The female proband (arrow, II-3) is homozygous for the *STAG3* missense variant, as is her infertile brother (II-1). Both affected individuals conceived children (III-1, III-2, III-3, III-4) using assisted reproduction with donor gametes. Year of birth is indicated for the siblings (generation II). Familial segregation of the variant was assessed with long-range nested PCR and sanger sequencing.

Figure 2: *STAG3* variants identified in infertile patients and/or reported in ClinVar (NP_036579.2)

Variants reported in POI patients are in black with single line, variants reported in the NOA patient are in bold with double line, variants observed in ClinVar are in grey with dotted line, variant observed in our study is underlined and indicated by an arrow. Reported variants are either homozygous (hom.) or heterozygous (het.). ARM-type fold: armadillo-type fold. N: amino-terminal domain, STAG: STAG domain, SCD: stromalin conservative domain, C: carboxyl-terminal domain.

Figure 3: Immunohistochemistry on testis biopsy (magnification x10, x20, x40 from left to right)

3a: Normal testis. *STAG3* protein is present in the seminiferous tubules in the nucleus of spermatogonia and spermatocytes (arrows).

3b: Testis from brother with NOA. *STAG3* protein is absent from the seminiferous tubules (arrows).

References

- Bijl N van der, Röpke A, Biswas U, Wöste M, Jessberger R, Kliesch S, Friedrich C, Tüttelmann F. Mutations in the stromal antigen 3 (STAG3) gene cause male infertility due to meiotic arrest. *Hum Reprod* 2019;**34**(11):2112-2119.
- Bouilly J, Beau I, Barraud S, Bernard V, Azibi K, Fagart J, Fèvre A, Todeschini AL, Veitia RA, Beldjord C, *et al.* Identification of Multiple Gene Mutations Accounts for a new Genetic Architecture of Primary Ovarian Insufficiency. *J Clin Endocrinol Metab* 2016;**101**(12):4541–4550.
- Brooker AS, Berkowitz KM. The roles of cohesins in mitosis, meiosis, and human health and disease. *Methods Mol Biol* 2014;**1170**:229–266.
- Caburet S, Arboleda VA, Llano E, Overbeek PA, Barbero JL, Oka K, Harrison W, Vaiman D, Ben-Neriah Z, García-Tuñón I, *et al.* Mutant cohesin in premature ovarian failure. *N Engl J Med* 2014;**370**(10):943–949.
- Colombo R, Pontoglio A, Bini M. A STAG3 missense mutation in two sisters with primary ovarian insufficiency. *Eur J Obstet Gynecol Reprod Biol* 2017;**216**:269–271.
- Ellermeier C, Smith GR. Cohesins are required for meiotic DNA breakage and recombination in *Schizosaccharomyces pombe*. *Proc Natl Acad Sci USA* 2005;**102**(31):10952–10957.
- Fauchereau F, Shalev S, Chervinsky E, Beck-Fruchter R, Legois B, Fellous M, Caburet S, Veitia RA. A non-sense MCM9 mutation in a familial case of primary ovarian insufficiency. *Clin Genet* 2016;**89**(5):603–607.
- França MM, Nishi MY, Funari MFA, Lerario AM, Baracat EC, Hayashida SAY, Maciel GAR, Jorge AAL, Mendonca BB. Two rare loss-of-function variants in the STAG3 gene leading to primary ovarian insufficiency. *Eur J Med Genet* 2019;**62**(3):186–189.
- He W-B, Banerjee S, Meng L-L, Du J, Gong F, Huang H, Zhang X-X, Wang Y-Y, Lu G-X, Lin G, *et al.* Whole-exome sequencing identifies a homozygous donor splice-site mutation in STAG3 that causes primary ovarian insufficiency. *Clin Genet* 2018;**93**(2):340–344.
- Heddar A, Dessen P, Flatters D, Misrahi M. Novel STAG3 mutations in a Caucasian family with primary ovarian insufficiency. *Mol Genet Genomics* 2019;**294**(6):1527–1534.
- Jan SZ, Jongejan A, Korver CM, Daalen SKM van, Pelt AMM van, Repping S, Hamer G. Distinct prophase arrest mechanisms in human male meiosis. *Development* 2018;**145**(16).
<http://molehr.oxfordjournals.org/>

- Laissue P. Aetiological coding sequence variants in non-syndromic premature ovarian failure: From genetic linkage analysis to next generation sequencing. *Mol Cell Endocrinol* 2015;**411**:243–257.
- Le Quesne Stabej P, Williams HJ, James C, Tekman M, Stanescu HC, Kleta R, Ocaka L, Lescai F, Storr HL, Bitner-Glindzicz M, *et al.* STAG3 truncating variant as the cause of primary ovarian insufficiency. *Eur J Hum Genet* 2016;**24**(1):135–138.
- Llano E, Gomez-H L, García-Tuñón I, Sánchez-Martín M, Caburet S, Barbero JL, Schimenti JC, Veitia RA, Pendas AM. STAG3 is a strong candidate gene for male infertility. *Hum Mol Genet* 2014;**23**(13):3421–3431.
- Moriwaki M, Moore B, Mosbrugger T, Neklason DW, Yandell M, Jorde LB, Welt CK. POLR2C Mutations Are Associated With Primary Ovarian Insufficiency in Women. *J Endocr Soc* 2017;**1**(3):162–173.
- Murphy MW, Lee JK, Rojo S, Gearhart MD, Kurahashi K, Banerjee S, Loeuille G-A, Bashamboo A, McElreavey K, Zarkower D, *et al.* An ancient protein-DNA interaction underlying metazoan sex determination. *Nat Struct Mol Biol* 2015;**22**(6):442–451.
- Pezzi N, Prieto I, Kremer L, Pérez Jurado LA, Valero C, Del Mazo J, Martínez-A C, Barbero JL. STAG3, a novel gene encoding a protein involved in meiotic chromosome pairing and location of STAG3-related genes flanking the Williams-Beuren syndrome deletion. *FASEB J* 2000;**14**(3):581–592.
- Quang D, Chen Y, Xie X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics* 2015;**31**(5):761–763.
- Ren Y, Diao F, Katari S, Yatsenko S, Jiang H, Wood-Trageser MA, Rajkovic A. Functional study of a novel missense single-nucleotide variant of NUP107 in two daughters of Mexican origin with premature ovarian insufficiency. *Mol Genet Genomic Med* 2018;**6**(2):276–281.
- Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, *et al.* 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* 2015;**17**(5):405–424.
- Riera-Escamilla A, Enguita-Marruedo A, Moreno-Mendoza D, Chianese C, Sleddens-Linkels E, Contini E, Benelli M, Natali A, Colpi GM, Ruiz-Castañé E, *et al.* Sequencing of a “mouse azoospermia” gene panel in azoospermic

men: identification of RNF212 and STAG3 mutations as novel genetic causes of meiotic arrest. *Hum Reprod* 2019;**34**(6):978–988.

Sadedin SP, Dashnow H, James PA, Bahlo M, Bauer DC, Lonie A, Lunke S, Macciocca I, Ross JP, Siemering KR, *et al.* Cpipe: a shared variant detection pipeline designed for diagnostic settings. *Genome Med* 2015;**7**(1):68.

Tucker EJ, Grover SR, Bachelot A, Touraine P, Sinclair AH. Premature Ovarian Insufficiency: New Perspectives on Genetic Cause and Phenotypic Spectrum. *Endocr Rev* 2016;**37**(6):609–635.

Tucker EJ, Jaillard S, Grover SR, Bergen J den, Robevska G, Bell KM, Sadedin S, Hanna C, Dulon J, Touraine P, *et al.* TP63-truncating variants cause isolated premature ovarian insufficiency. *Human Mutation* 2019;**40**(7):886-892..

Venselaar H, Te Beek TAH, Kuipers RKP, Hekkelman ML, Vriend G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics* 2010;**11**:548.

Xiao W-J, He W-B, Zhang Y-X, Meng L-L, Lu G-X, Lin G, Tan Y-Q, Du J. In-Frame Variants in STAG3 Gene Cause Premature Ovarian Insufficiency. *Front Genet* 2019;**10**:1016.

Accepted Manuscript

Table 1. Relevant variants within diagnostic and candidate genes in the proband.

Gene	Condition (inheritance)	Chromosomal variant GRCh37	Variant	Type of variant	Amino acid change	MAF (gnomAD)	Prediction	Previous description	Inheritance	ACMG-based criteria
<i>HS6ST1</i>	Hypogonadotropic hypogonadism (susceptibility, AD)	NC_000002.11: g.129025851C>T	NM_004807.2: c.1121G>A	Heterozygous Missense	NP_004798.3: p.(Ser374Asn)	T: 0.0008421 EP: 0.001439	MT: polymorphism PP-2: benign SIFT: tolerated	dbSNP: rs61732021 ClinVar: benign	Father heterozygous	Likely benign
<i>STAG3</i>	Premature ovarian insufficiency (AR)	NC_000007.13: g.99794799G>A	NM_012447.3: c.962G>A	Homozygous Missense	NP_036579.2: p.(Arg321His)	0	MT: disease causing PP-2: probably damaging SIFT: damaging	dbSNP: NR ClinVar: NR	Father and mother heterozygous	Likely pathogenic
<i>RECQL4</i>	Baller-Gerold, RAPADILINO, Rothmund-Thomson syndromes (AR)	NC_000008.10: g.145737131C>G	NM_004260.3: c.3435G>C	Heterozygous Missense	NP_004251.3 (LRG_277p1): p.(Gln1145His)	T:0.002964 EP: 0.004466	PP-2: benign SIFT: tolerated	dbSNP: rs61755066 ClinVar: benign / likely benign	Father heterozygous	Likely benign
<i>ATM</i>	Ataxia-telangiectasia (AR)	NC_000011.9: g.108123551C>T	NM_000051.3: c.1810C>T	Heterozygous Missense	NP_000042.3 (LRG_135p1): p.(Pro604Ser)	T: 0.003179 EP: 0.001549	MT: disease causing PP-2: possibly damaging / benign SIFT: tolerated	dbSNP: rs2227922 ClinVar: Benign(5) / Likely benign(7) / VUS(1)	Mother and brother II-1 heterozygous	Likely benign
<i>ATM</i>	Ataxia-telangiectasia (AR)	NC_000011.9: g.108175463A>T	NM_000051.3: c.5558A>T	Heterozygous Missense	NP_000042.3 (LRG_135p1): p.(Asp1853Val)	T: 0.004883 EP: 0.007144	MT: disease causing PP-2: possibly damaging / benign SIFT: damaging	dbSNP: rs1801673 ClinVar: Benign(5) / Likely benign(7) / VUS(1)	Father and brother II-1 heterozygous	Likely benign
<i>RCBTB1</i>	Retinal dystrophy with or without extraocular anomalies (AR)	NC_000013.10: g.50115874T>A	NM_018191.3: c.1262A>T	Heterozygous Missense	NP_060661.3: p.(Tyr421Phe)	T: 0.00009192 EP: 0.0004800	MT: disease causing PP-2: benign SIFT: tolerated	dbSNP: rs146955596 ClinVar: NR	Mother and brother II-1 heterozygous	VUS
<i>RSPH1</i>	Primary ciliary dyskinesia (AR)	NC_000021.8: g.43906573T>G	NM_080860.3: c.275-2A>C	Heterozygous Splice site	NP_543136.1: p.?	T: 0.0003535 EP: 0.0006736	Loss of function	dbSNP: rs151107532 ClinVar: pathogenic	Mother heterozygous	Pathogenic
<i>AIRE</i>	Autoimmune polyendocrinopathy syndrome (AD, AR)	NC_000021.8: g.45710692G>A	NM_000383.3: c.880-286G>A	Heterozygous Missense	NP_000374.1 (LRG_18p1):p.?	T: 0.000004002 EP: 0.000008884	MT: disease causing	dbSNP: rs556919503 ClinVar: NR	Father and brother II-1 heterozygous	Likely pathogenic
<i>HTRA3</i>		NC_000004.11: g.8295839C>T	NM_053044.4: c.962C>T	Heterozygous Missense	NP_444272.1: p.(Thr321Met)	T: 0.0002940 EP: 0.0005521	MT: disease causing PP-2: probably damaging SIFT: damaging	dbSNP: rs151074010 ClinVar: NR	Father heterozygous	VUS
<i>CITED2</i>		NC_000006.11: g.139694603T>A	NM_001168388.2: c.479A>T	Heterozygous Missense	NP_001161860.1: p.(His160Leu)	T: 0.001576 EP: 0.002906	MT: disease causing PP-2: possibly damaging SIFT: damaging	dbSNP: rs111814036 ClinVar: NR	Mother heterozygous	VUS

<i>IGF2R</i>		NC_000006.11: g.160493009C>T	NM_000876.3 : c.4511C>T	Heterozygous Missense	NP_000867.2 (LRG_1036p1): p.(Thr1504Ile)	T: 0.001099 EP: 0.001525	MT: polymorphism PP-2: benign SIFT: tolerated	dbSNP: rs77999960 ClinVar : NR	Mother heterozygous	Likely benign
<i>ADGRA2</i>		NC_000008.10: g.37693106C>T	NM_032777.9 : c.1868C>T	Heterozygous Missense	NP_116166.9: p.(Pro623Leu)	T: 0.00001990 EP: 0.00002640	MT: disease causing PP-2: possibly damaging / benign SIFT: damaging	dbSNP: rs370919357 ClinVar : NR	Mother and brother II-1 heterozygous	VUS
<i>TNFRSF1A</i>		NC_000012.11: g.6438570G>A	NM_001065.3 : c.1276C>T	Heterozygous Missense	NP_001056.1 (LRG_193p1): p.(Arg426Cys)	0	MT: polymorphism PP-2: benign SIFT: tolerated	dbSNP: rs56279759 ClinVar : NR	Father heterozygous	VUS
<i>SMAD9</i>		NC_000013.10: g.37427805G>C	NM_0011272 17.2: c.1011C>G	Heterozygous Missense	NP_001120689. 1 (LRG_703p1): p.(His337Gln)	T: 0.0004252 EP: 0.0005984	MT: disease causing PP-2: benign SIFT: tolerated	dbSNP: rs149015682 ClinVar: VUS	Father heterozygous	VUS
<i>MLH3</i>		NC_000014.8: g.75513934T>C	NM_014381.2 : c.2425A>G	Heterozygous Missense	NP_055196.2: p.(Met809Val)	T: 0.001864 EP: 0.002765	MT: polymorphism PP-2: benign SIFT: tolerated	dbSNP: rs61752722 ClinVar: likely benign	Father, mother, and brother II-1 heterozygous	Likely benign

3 White: variants in diagnosis genes, bold: variants in *STAG3*, light grey: variants in candidate genes

4 T: total, EP: European population, MT: Mutation Taster, PP-2: PolyPhen-2, SIFT: Sorting Intolerant From Tolerant, VUS: variant of uncertain significance, NR: not reported, ACMG: American College
5 of Medical Genetics and Genomics.

Table 2. Molecular data of reported STAG3 variants.

	Chromosomal variant GRCh37 NC_000007.13	Variant NM_012447.3	Type of variant	Amino acid change NP_036579.2	Expectation	Exon/Intron	Functional studies	Human phenotype
Caburet et al., 2014	g.99786486del	c.[562delC];[562delC] ENST00000426455: c.968delC ¹	Homozygous Frameshift by deletion 2 premature stop codons	p.[(Gln188Argfs*8)];[(Gln188Argfs*8)] p.Phe187fs*7	Truncated protein or NMD (LoF)	Exon 7/34	<i>Stag3</i> ^{-/-} mouse (POI female and infertile male)	POI
Le Quesne Stabej et al., 2016	g.99798478_99798479 dup	c.[1947_1948dupCT];[1947_1948dupCT]	Homozygous Frameshift by duplication 2 premature stop codons	p.[(Tyr650Serfs*22)];[(Tyr650Serfs*22)]	Truncated protein or NMD (LoF)	Exon 19/34	NP	POI
He et al., 2017	g.99796995G>A	c.[1573+5G>A];[1573+5G>A]	Homozygous Frameshift by splice site 2 premature stop codons	p?	Truncated protein or NMD (LoF)	Intron 15/33	Truncated mRNA identified from patients by RT-PCR (p.Leu490Thrfs*10 predicted)	POI
Colombo et al., 2017	g.99786601C>G	c.[677C>G];[677C>G]	Homozygous Nonsense 2 premature stop codons	p.[(Ser226*)];[(Ser226*)] p.(Ser227*) ²	Truncated protein or NMD (LoF)	Exon 7/34	NP	POI
França et al., 2019	g.g.99780417dup g.99798481C>A	c.291dupC(;);1950C>A	Compound heterozygous? Frameshift by duplication + nonsense 2 premature stop codons	p.(Asn98Glnfs*2)(;)(Tyr650*)	Truncated protein or NMD (LoF)	Exon 4/34 Exon 19/34	NP	POI
Heddar et al., 2019	g.99786583T>G g.99802728del	c.[659T>G;3052delC]	Compound heterozygous Missense + Frameshift by deletion 1 premature stop codon	p.[(Leu220Arg)];[(Arg1018Aspfs*14)]	Impaired protein function and truncated protein or NMD (LoF)	Exon 7/34 Exon 28/34	NP	POI
Xiao et al., 2019	g.99792941_99792949 del g.99792955_99792957 dup	c.[877_885del;c.891_893dupTGA];[877_885del;c.891_893dupTGA]	Two homozygous variants Inframe	p.[(His293_Glu295del;Ile297_Glu298insAsp)];[(His293_Glu295del;Ile297_Glu298insAsp)]	Loss of protein function	Exon 9/34	Aberrant REC8 localization and loss of interactions with REC8 or SMC1A	POI
Our study	g.99794799G>A	c.[962G>A];[962G>A]	Homozygous Missense http://molehr.oxfordjournals.org/	p.[(Arg321His)];[(Arg321His)]	Impaired protein	Exon 10/34	Spermatogenic arrest in male patient,	POI and NOA

					function		aberrant STAG3 protein expression in patient testicular biopsy	
Riera-Escamilla et al., 2019	g.99798467dup g.99799665G>A	c.[1936dupG];[2394+1G>A] c.1759dupG	Compound heterozygous Frameshift by insertion + splice site 2 premature stop codons	p.[(Ala646Glyfs*9)];[?] p.(Ala586fs) ³	Truncated protein or NMD (LoF)	Exon 17/34 Intron 23/33	Spermatogenic arrest in patient	NOA
Van der Bijl et al., 2019	g.99796115T>G g.99796165C>T	c.[1262T>G];[1312C>T]	Compound heterozygous Missense + Nonsense 1 premature stop codon	p.[(Leu421Arg)];[(Arg438*)]	Impaired protein function and truncated protein or NMD (LoF)	Exon 13/34	Spermatogenic arrest in patient, aberrations in chromosome architecture	NOA
ClinVar	g.99783821A>C	c.337-2A>C	ND allelic status rs1296715259	p.?	Likely pathogenic	Intron 4/33	NP	ND
ClinVar	g.99794903G>C	c.1065+1G>C	ND allelic status rs1554406947	p.?	Likely pathogenic	Intron 10/33	NP	ND
ClinVar	g.99801719C>T	c.2776C>T	ND allelic status rs764841861	p.(Arg926*)	Likely pathogenic	Exon 26/34	NP	POI

1. Caburet *et al.* describes the *STAG3* variant in terms of the ENST00000426455 Ensembl reference transcript, and does not begin numbering with “c.1” at the A of the ATG start codon as per HGVS guidelines. The protein nomenclature has also been amended because the protein is encoded normally up to and including p.Phe187. The first altered amino acid is the substitution of p.Gln188Arg and there are a total of 8 altered codons, including the new stop codon (Caburet et al., 2014).
2. Colombo *et al.* describe the c.677C>G variant as a missense variant affecting codon number 227, whereas it is in fact a nonsense variant affecting codon 226 (Colombo et al., 2017).
3. Riera-Escamilla *et al.* describe a NM_001282718:c.1759dupG variant, however position c.1759 is not a G. Electropherograms were provided by the authors upon request, which clarified the variant is in fact NM_001282718:c.1762dupG, which is the equivalent to NM_012447.3:c.1936dupG:p.(Ala646Glyfs*9).

DNA variant numbering is based on GenBank reference DNA sequence NM_012447.3, with the A of the ATG initiation codon designated +1. Chromosomal variants are described according to the nucleotide sequence NC_000007.13 (GRCh37). Predicted protein annotations are based on NP_036579.2.

In variant and amino acid change columns, correct nomenclature with reference to NM_012447.3 and NP_036579.2 is provided in bold. Erroneous nomenclature or variants described using a different transcript in the literature are indicated underneath.

NMD: nonsense mediated decay, POI: premature ovarian insufficiency, NOA: non-obstructive azoospermia, ND: not determined, NP: not performed.

Table 3. Clinical data of the reported POI patients with *STAG3* variants.

	Caburet et al., 2014	Le Quesne Stabej et al., 2016		He et al., 2017		Colombo et al., 2017		França et al., 2018	Heddar et al., 2019		Xiao et al., 2019		Our study
Age at diagnosis	17-20 (four sisters)	11	11 (younger sister)	19	ND (younger sister)	28	31 (elder sister)	21	13	17 (younger sister)	18	21 (elder sister)	15
Menses	PA	PA Menarche with ET	PA Menarche with ET	PA Menarche with ET	PA	PA	PA	PA	PA	PA	PA	PA	PA
Secondary sex characters	Undeveloped breast	No spontaneous puberty Undeveloped breast Development with ET	No spontaneous puberty Development with ET	Undeveloped breast Sparse pubic / axillary hair	Undeveloped breast	Breast Tanner I / II Sparse pubic hair	Delayed puberty Development with ET	Development with ET Breast Tanner III Pubic hair Tanner III	No spontaneous puberty	No spontaneous puberty	Breast Tanner I	ND	Delayed puberty Breast Tanner I-II Pubic and axillary hairs Tanner I
Height	N	N	N	ND	N	-1.2 SD	-0.9 SD	N	-1.5 SD	N	N	ND	+1 SD
Weight	ND	ND	ND	ND	ND	+0.6 SD	+0.4 SD	N	-2 SD	N	N	ND	+1 SD
US	Streak gonads	Normal uterus Streak gonads	Small uterus Streak gonads	Small uterus Streak gonads	Streak gonads	Streak gonads (fibrous tissue)	ND	Small uterus Ovaries not seen	Small uterus Streak gonads	ND	Uterus and ovaries not seen	Small uterus Small ovaries	Small uterus Small right ovary with small follicles, atrophic left ovary (not seen)
FSH IU/l	> 45	136	130	48.69	48.38	48	Similar	89	86	ND (POI level)	51	72.19	46.5
LH IU/l	> 18	31	62	26	25.51	23	Similar	37	27.5	ND (POI level)	11.08	18.58	24.5
Estradiol pg/ml	< 22	< 10	< 10	N	N	19	Similar	<13	<3	ND (POI level)	<10	<10	33
Other hormonal levels		INHB undetectable	INHB undetectable	AMH: 0.03 ng/ml Prolactin N Thyroid hormone N INHB undetectable	Prolactin N Progesterone N	Testosterone N INHB undetectable	Similar to that of sister	INHB low	ND	ND	AMH undetectable	ND	SHBG: 23 nmol/l (low)
Karyotype	46,XX	46,XX	46,XX	46,XX	46,XX	46,XX	46,XX	46,XX	46,XX	ND	46,XX	46,XX	46,XX
Auto-Ab	Auto-Ab -ve	Ovarian auto-Ab -ve	Ovarian auto-Ab -ve	Auto-Ab -ve	ND	Ovarian auto-Ab -ve	Ovarian auto-Ab -ve	Adrenal or thyroid auto-Ab -ve	Ovarian auto-Ab -ve	ND	Adrenal or thyroid auto-Ab -ve	Adrenal or thyroid auto-Ab -ve	Auto-Ab -ve
<i>FMR1</i>	N	ND	ND	N	ND	ND	ND	N	N	ND	N	N	N
Other	Consanguineous parents Gonadoblastoma, complex tumor in one of the sisters	Consanguineous parents		Consanguineous parents Unaffected parents and brother het. carriers		Consanguineous parents Het. carriers with premature menopause or irregular menstruation / oligoamenorrhea		Adopted woman Parental DNA non-available	Non-consanguineous parents		Consanguineous parents		Consanguineous parents One infertile brother (non-obstructive azoospermia)

PA: primary amenorrhea, POI: primary ovarian insufficiency, US: ultrasonographic examination, FSH: follicle stimulating hormone, LH: luteinising hormone, INHB: inhibin beta, AMH: anti-müllerian hormone, SHBG: sex hormone-binding globulin Ab: antibodies, ET: estrogen therapy, N: normal, SD: standard deviation, ND: not determined/described, het.: heterozygous
Hormonal levels were converted to a unique unit of measure.

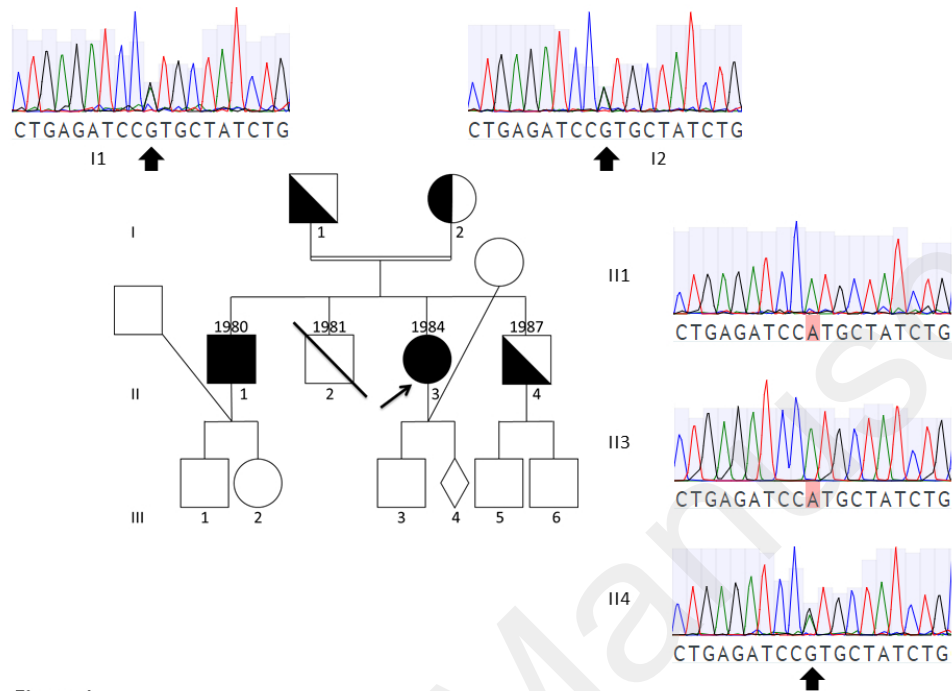


Figure 1

Figure 1. Familial pedigree and STAG3 genotype

254x190mm (96 x 96 DPI)

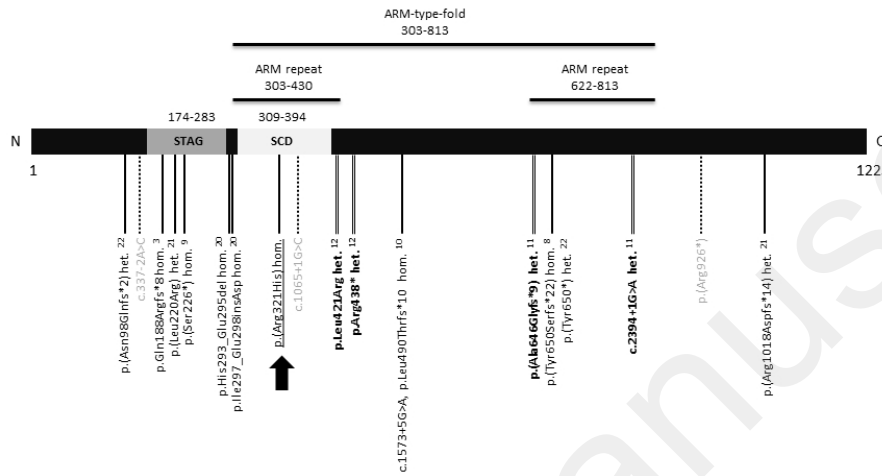


Figure 2

Figure 2. STAG3 variants identified in infertile patients and/or reported in ClinVar (NP_036579.2)

254x190mm (96 x 96 DPI)

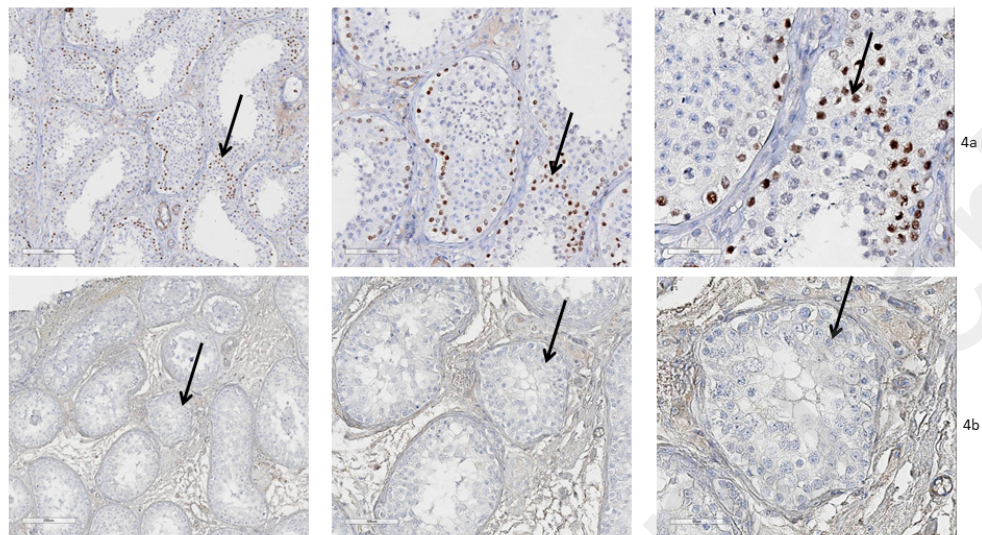


Figure 4

Immunohistochemistry on testis biopsy (magnification x10, x20, x40 from left to right)
3a: normal testis. STAG3 protein is present in the seminiferous tubules in the nucleus of spermatogonia and spermatocytes (arrows).

3b: Testis from brother with NOA. STAG3 protein is absent from the seminiferous tubules (arrows).

254x190mm (96 x 96 DPI)