Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Caburet S, Arboleda VA, Llano E, et al. Mutant cohesin in premature ovarian failure. N Engl J Med 2014;370:943-9. DOI: 10.1056/NEJMoa1309635

SUPPLEMENTARY APPENDIX

A - Supplementary Material & Methods

I – Generation of transgenic mouse with a lentiviral insertional mutation in *Stag3*

1. Overview of the lentiviral insertional mutagenesis strategy
2. Methods
3. Founding of the OVE2312C line

II – Whole-exome sequencing and Sanger sequencing

1. Whole-exome capture and sequencing of genomic DNA
2. Sequence data analysis
3. Sanger sequencing confirmation

III – Characterization of *Stag3*$^{-/-}$ mouse

B - Supplementary Results

I – Whole-exome sequencing in MO1DA family

1. Details of sequencing results
2. Variants detected in the consanguineous MO1 family
3. Verification of the segregation of the STAG3 deletion in the family by Sanger sequencing
4. Genetic exclusion of *STAG3* locus in other families with infertility

II – Characterization of *Stag3*$^{-/-}$ mouse: supplemental data

1. Verification of the insertional mutant at the *Stag3* locus
2. Double strands breaks formation and Rad51 recruitment
3. Cohesin complexes in the absence of STAG3.
4. Lack of STAG3 causes premature loss of cohesion
5. Absence of Stag3 leads to male infertility by azoospermia

Supplementary Figure S1: Wild-type locus and insertional mutant at the Stag3 locus in the OVE2312C mouse

Supplementary Figure S2: Confirmation of the 1-bp deletion in the STAG3 gene in the MO1DA family

Supplementary Figure S3: Disruption of the Stag3 locus in the OVE2312C mouse

Supplementary Figure S4: Expression of the Stag3 gene in wild-type (+/+) and knockout (−/−) animals

Supplementary Figure S5: Double immunofluorescence assays for SYCP3 and STAG3 in mouse fetal oocytes

Supplementary Figure S6: Double Strand Break-associated proteins in Stag3$^{-/-}$ oocytes

Supplementary Figure S7: Cohesin complexes in the absence of STAG3.
Supplementary Figure S8: Lack of STAG3 causes premature loss of cohesion
Supplementary Figure S9: Ring model of the cohesin complex
Supplementary Figure S10: Azoospermia in the Stag3$^{-/-}$ male mouse

Supplementary Table S1: Primers for creation and characterization of transgenic mice
Supplementary Table S2: Primers used for PCR amplification and Sanger sequencing of variants
Supplementary Table S3: Quality statistics for the whole-exome sequencing of the two MO1DA sisters
Supplementary Table S4: Numbers of Single Nucleotide Variants identified during the bioinformatics filtering
Supplementary Table S5: Variants detected by whole-exome sequencing in the MO1DA POF patient

C - Supplementary References

A - Supplementary Material & Methods

I – Lentiviral insertional mutagenesis

1. Overview of the lentiviral insertional mutagenesis strategy:

Self-inactivating (SIN) lentiviruses have previously been shown to be efficient vectors for the generation of transgenic mice by infection of preimplantation embryos.[1,2] In order to test whether lentiviruses can be used for efficient insertional mutagenesis in live mice, we inserted a tyrosinase reporter gene to provide a visible reporter to keep track of the lentiviral integration sites. The tyrosinase minigene provides some unique advantages as a reporter gene. Expression of tyrosinase rescues albinism, so transgenic mice can be distinguished from non-transgenic mice by simple inspection for pigmentation.[3] The minigene is typically not expressed at wild type levels, but rather at levels that make it rate limiting for melanin production. [4,5] As a result, for most integration sites, homozygous transgenic mice can be identified by simple visual inspection since they have a darker coat color than their hemizygous siblings. Homozygosity can be confirmed by mating to an albino partner. Third, expression of the tyrosinase minigene shows chromosomal position effects. Different transgenic integration sites give different levels of tyrosinase production and different intensities of pigmentation. Therefore, it is relatively simple to identify founder mice with multiple integration sites, and it is straightforward to sort out the integration sites by simple visual inspection. Furthermore, when the tyrosinase gene is well expressed, there is a good probability that the lentivirus has integrated within an endogenous gene (data not shown).

Four different viruses were generated and used to infect 1-cell to 8-cell stage albino FVB embryos. Two of the vectors were modified to contain the arms of the pT2 Sleeping Beauty transposon.[6] We refer to these vectors as lentiposons. The lentiviruses and lentiposons provided efficient transgenesis as monitored by the rescue of albinism, and all four lentiviruses generated a collection of useful insertional mutations. After just 5 days of embryo infections, more than 1000 unique lentiviral insertion sites were generated and more than 80 novel lines of mutant mice were identified and molecularly defined (data not shown). The mouse lines are available at the JAX Repository, Jackson Laboratory (www.jax.org). This article describes the characterization of one of the lentiposon lines (OVE2312C).

2. Methods:

2a. Layout of the Lentiviral Vector

The SB-cHS4core-SB-Tyro-WPRE lentiposon (LV2229) was designed with inverted Sleeping Beauty (SB) arms (IR/DR repeats,[5,6]) flanking a tandem 0.25 kb chicken beta-globin core insulator element (cHS4 core,[7]), followed by a mouse tyrosinase minigene (Tyro), and a woodchuck hepatitis virus post-transcriptional regulatory element (WPRE) in the FUGW self-inactivating HIV-based lentiviral vector backbone.[1] The SB arms, insulator elements and Tyro minigene replaced the ubiquitin-c promoter and EGFP sequences originally found in the FUGW lentiviral vector. The Tyro minigene is composed of a 620 bp NsiI fragment containing the mouse Tyr enhancer,[8] linked to the mouse tyrosinase promoter (500 bp) and 1.7 kbp of tyrosinase cDNA sequence (including the stop codon). The minigene has a sense orientation relative to the lentiviral backbone.

There is no polyA site between the Tyro minigene and WPRE sequences. The WPRE sequences function to enhance transcript stability.[1,2] There are no introns or splice acceptor sites. The insulator sequences were added both to increase the likelihood of expression of the minigene and also to make the lentiposon more mutagenic. Mobilization of the transposon was not used for the experiments described in this manuscript.

2b. Lentiviral Packaging

Lentivirus were packaged and tittered in Gene Vector Core in Diabetes and Endocrinology Research Center at Baylor College of Medicine. Lenti transfer vectors were packaged using second-generation packaging plasmids, the psPAX2 and pMD.2 (provided by D. Trono). In brief, confluent 293T cells were split 1:6 and plated on a 15-cm dish (D-MEM, 10% FBS, 4.5 g/L glucose, 1 x antimicotic/antibiotics) 24 hours before transfection, and transfection was performed at 70-80% confluent. 1.5 ml of DNA mix was prepared as follows: 60 μg transfer plasmid, 45 μg psPAX2, 18 μg pMD2.G, 150 μl of 2.5 M $CaCl_2$ to 1.5 ml with transfection grade H_2O. The DNA mixture was added dropwise to1.5 ml 2 x HBS (42 mM HEPES, 290 mM NaCl, pH 7.1) while gently vortexing. After 10 min at room temperature, the entire mixture was overlayed on 293T cells. After 5 hours in a CO_2 incubator (5% CO_2), the medium was aspirated and 32 ml of fresh medium (D-MEM, 5% FBS, 4.5 g/l glucose, 1 x antimicotic/antibiotics) were added. After 48 hours, the supernatant was collected and filtered through 0.45 μm filter (Millipore). The supernatant was then centrifuge on top of 6 ml of D-MEM containing 20% sucrose in SW32 Ti rotor (Beckman) at 4°C at 25,000 rpm for 2 hours. The pellets containing the packaged lentiviruses were resuspended in 200 μl of plain D-MEM and stored at -80°C until use.

Transduction units were determined by an end point assay. In brief, 1 x 10^5 293 T cells were plated in a 24-well plate with 0.5 ml of D-MEM, 10% FBS, polybrene (8 μg/mL) and then 20 μl of serially diluted lentivirus were added. After 48 hours, lenti RNA genome integrated into chromosomal DNA was purified using DNA extraction kit (Qiagen). Lentiviral DNA was quantified by real time PCR using Stratagene Mx3005P and SYBRGreen Supermix (Quanta). Transducing units were calculated as lentiviral DNA copy #/diploid cell/μl.

2c. Embryo infections/transgenesis

FVB females were superovulated and mated to FVB males. Embryos were harvested in KSOM on day 1 (one-cell stage). Cumulus cells were removed by treatment with hyaluronidase. Individual embryos were microinjected under the zona pellucida using a micromanipulator. After washing 2X in KSOM, embryos were transferred to the oviducts of newly plugged pseudopregnant ICR females.

2d. Genotyping for the tyrosinase minigene and breeding

Tail snips were obtained from mice at 20-25 days of age and were incubated overnight at 60 degrees in PCR buffer supplemented with proteinase K and NP40. After heating to 95 degrees for 20 minutes to inactivate the proteinase K, samples were cleared by centrifugation in a microfuge. The presence of the tyrosinase transgene was assessed by PCR using the tyroA and tyroB primers (See Supplementary Table 1). All pigmented founder

mice were positive by PCR. Founder mice were bred to FVB partners. Pigmented F1 mice were again mated to FVB partners. When these matings produced litters where a) approximately 50% of the offspring were pigmented, b) only the pigmented mice were positive for the tyrosinase minigene, and c) all pigmented offspring exhibited the same coat color, the offspring were inbred to generate homozygotes. If the F1 matings produced offspring with more than one coat color, then the F2 mice were bred to FVB partners to identify mice with a single integration site. Similar matings were done with the F3 mice if needed. Putative homozygous mice were identified by simple visual inspection for a darker coat color. The putative homozygotes were then bred to FVB partners to test for fertility and to confirm homozygosity. For 7 families, the homozygous males or females were found to be infertile or to have significantly reduced fertility (data not shown). For the OVE2312C insertion site, homozygous males and females were both infertile.

2e. Identification of integration site.

* Inverse PCR: To amplify the junction between the 3'-LTR and the adjacent genomic sequences, 2 µg of tail genomic DNA was digested separately with Alw44I, Pst I and ApoI. The digested DNAs were diluted to 250 µl and ligated overnight at room temperature with T4 DNA ligase to circularize the genomic DNA fragments. The ligated DNAs were precipitated and resuspended in 25 µl of 10 mM Tris. Nested PCR amplifications were performed using Nef(F)#1 and WPRE(AS)#1 primers for the first round, then using primers Nef(F)#2 and WPRE(AS)#2 for a second round of amplification (See Supplementary Table 1). PCR products were purified by agarose gel electrophoresis and sequenced with the 3'-LTR(S) primer (See Supplementary Table 1). Sequences were analyzed by a BLAT search on the UCSC Genome Bioinformatics Website (genome.ucsc.edu). The BLAT results were assessed for the following: a) a clean junction between the end of the lentiviral LTR and the adjacent mouse genomic sequences, b) 99% or better identity to the mouse genome, and c) the presence within the genomic sequences of a recognition site for the enzyme that was used for the genomic digestion.

* Validation by direct PCR and changes in gene expression

In order to confirm each lentiviral integration site, primers were designed based on the mouse genomic sequences located both upstream and downstream from the putative insertion site. These primers were used in conjunction with primers from the 5'-LTR and the 3'-LTR for direct PCR amplification from genomic DNA (See Table S1). PCR products were first assessed by gel electrophoresis for the presence of a band of the predicted size, then sequenced to verify the junction site.

3. Founding of the OVE2312C line:

One-cell stage FVB embryos were infected by perivitelline injection of packaged lentivirus LV2229. 42 pigmented F0 transgenic mice were obtained. The founder mouse for family OVE2312 was a mottled female. After mating to an FVB male, she produced offspring with 3 different coat colors, initially labeled as 2312A, 2312B and 2312C. Matings revealed that the 2312B mouse had two integration sites, which corresponded to the 2312A and 2312C sites. The 2312A mice showed dark grey pigmentation. The 2312C mice were mottled. Mice

that were homozygous for site A were found to be viable and fertile. Mice that were homozygous for site C were sterile, both males and females. The 2312C integration site was amplified by inverse PCR (see above), and found to be located on mouse chromosome 5 in intron 8 of the Stag3 gene. Hemizygous transgenic mice display no overt phenotype. Hemizygotes were crossed to produce homozygous mice that were analyzed at fetal, perinatal and adult stages. The schematic representation of the mutated *Stag3* locus is presented in Figure S1.

II – Whole-exome sequencing and Sanger sequencing

1. Whole-exome capture and sequencing of genomic DNA.

For the capture of the genes we used Agilent SureSelect XT 50Mb Exon Capture. Genomic DNA libraries were created for patients IV-1 and IV-3 from MO1DA family, following manufacturer's protocol (Illumina protocol Preparing Samples for Sequencing Genomic DNA, p/n 11251892 Rev. A). After the capture, the array was washed and the captured DNA was eluted, amplified and diluted to 10nM final concentration based on both the Qubit and Agilent Bioanalyzer measures. One flowcell lane of paired-end sequencing was performed at the Broad Stem Cell Research Center (BSCRC) High Throughput Sequencing Core on the Illumina HiSeq 2000 for 50x cycles following manufacturer's protocol. The base-calling was performed by the real-time analysis (RTA) software provided by Illumina.

2. Sequence data analysis :

The Illumina output files (.qseq format) were aligned to the human reference genome (hg19, February 2009, build 37.5) using Novoalign V2.07.13 from the Novocraft Short Read Alignment Package (http://www.novocraft.com/index.html). The output format was set to SAM and default settings were used for all options. Using SAMtools (http://samtools.sourceforge.net/), the SAM file of each sample was converted to a BAM file, sorted and merged. PCR duplicates were removed using Picard (http://picard.sourceforge.net/). Files were further processed using Genome Analysis Tool Kit from the Broad Institute (http://www.broadinstitute.org/gatk/). All BAM files were locally realigned using indel realigner. Both single nucleotide variants (SNVs) and small INDELs (insertions and deletions) within the captured coding exonic intervals were called using GATK's Unified Genotyper. All high quality variants were reported and annotated using dbSNP 132, build 37 (http://www.ncbi.nlm.nih.gov/SNP). Filtering of the variants was performed using data from dbSNP and HGMD databases (http://www.hgmd.cf.ac.uk/ac/index.php).

3. Sanger sequencing confirmation:

The exons encompassing the 8 identified mutations were confirmed by Sanger sequencing in both unaffected parents, the remaining affected sisters and the remaining unaffected siblings. In addition, an independent Sanger sequencing verified the segregation of the strongest candidate variant in the *STAG3* gene in all members, including the two fertile brothers.

The primers used for PCR amplification and Sanger sequencing are shown in Table S2.

III – Characterization of *Stag3*$^{-/-}$ mouse:

IF was performed following a standard protocol. Briefly, slides were blocked in a PBS solution containing 10% BSA for 30 min. Incubation with primary antibodies was performed at 4°C overnight or at room temperature for 3 h. The following primary antibodies were used: mouse polyclonal antibody against REC8, rabbit polyclonal antibodies against REC8 (K675 and K775), STAG3 (K403), SMC1β (K976), SMC3 (K987), RAD51 and SCP3. After several washes, fluorescein isothiocyanate-, Cy3- and Cy5-conjugated antibodies were used (all from Jackson ImmunoResearch) for the detection of primary antibodies.

B - Supplementary Results

I – Whole-exome sequencing in MO1DA family

1. Details of sequencing results

The whole-exome quality statistics of the two members of the MO1DA family are shown in Table S3.

The numbers of Single Nucleotide Variants identified during the bioinformatic analysis are provided in Table S4.

2. Variants detected in the consanguineous MO1 family

No coherent variant was identified in the 7p linkage region.

The location and molecular consequences of the 8 variants detected on the 7q linkage region are shown in the Table S5. Variants are listed in the chromosomal order. Chromosomal positions are for the GRCh37.4 (hg19) build.

We verified the function and expression data (when known) of the genes that harbor these variants. Details are provided below. (Unigene Database: http://www.ncbi.nlm.nih.gov/unigene; GEOprofiles on Gene Expression Omnibus: http://www.ncbi.nlm.nih.gov/geo/; EST: Expressed Sequenced Tag)

- Variant 1: 7_99 022 873_C/T

PTCD1 and the read-through gene *ATP5J2-PTCD1* share a non-synonymous coding variant predicted as benign. *PTCD1* (pentatricopeptide repeat domain 1) encodes a mitochondrial matrix protein which associates with leucine tRNAs and precursor RNAs that contain leucine tRNAs.[9] It has a ubiquitous expression, and is not restricted to a developmental stage either (Unigene ID:2139207).

The read-through transcription between the *ATP5J2* (ATP synthase, H+ transporting, mitochondrial Fo complex, subunit F2) and the *PTCD1* genes encodes a fusion protein that shares sequence identity with each individual gene product, which was not functionally studied. The 3 known specific ESTs are from cerebellum, thalamus and fetal brain (ConjoinG database, http://metasystems.riken.jp/conjoing).

- Variant 2: 7_99,786,485_C/-

In the exon 7 of the *STAG3* gene, and inducing a frameshift in the coding sequence (Details in the main text and Figure 1). In addition to the longer transcript shown in Figure 1B (main text), a shorter *STAG3* isoform of 1167 amino acids (reference Ensembl transcript ENST00000394018) is predicted to be truncated to 194 amino acids.

- Variant 3: 7_100 371 477_-/G

ZAN (zonadhesin) encodes a sperm membrane protein localized to the anterior part of the sperm head that acts as a receptor to the zona pellucida matrix of the egg in a species-specific manner. Like other male reproductive proteins located in the acrosome or sperm head and proposed to be involved in species-specific binding to eggs, ZAN possesses high levels of divergence between closely related species. [10,11]

GEOprofiles confirm an absence of expression in the fetal or adult ovary (i.e. GDS2203/1421614_at or GDS3368/1421614_at). Therefore, although this gene is involved in reproduction, its function is clearly not relevant for an ovarian phenotype.

- Variants 4,5 and 6: 7_100 637 008_C/A, 7_100 637 035_A/C, 7_100 637 041_C/A

MUC12 (mucin 12, cell surface associated) encodes a large cell surface glycoprotein of the non-secreted mucin family, involved in epithelial cell protection, adhesion modulation, and signaling.[12] This gene displays a wide range of gene expression profiles and is expressed 20x more in the intestine than in the ovary. In the family analysed here, it contains 3 closely-spaced, non-synonymous variants in the portion of the gene coding for the outer domain of the protein, encoded in a very large exon common to the two isoforms of the protein. The mutated threonine is predicted to be O-glycosylated. A low expression of MUC12 was identified as a marker for metastasis in colorectal cancer.[13] The functional impact of 3 closely-spaced, non-synonymous variants, could be important. But if this combination of three variants was functionally relevant in the ovary, we could expect a major phenotype in the digestive epithelium too.

- Variant 7: 7_100 839 291_T/A

MOGAT3 (monoacylglycerol O-acyltransferase 3) encodes the monoacylglycerol O-acyltransferase 3, that catalyzes the synthesis of diacylglycerol from 2-monoacylglycerol and fatty acyl-CoA. This enzymatic reaction is an essential and rate-limiting step for the absorption of fat in the small intestine.

In humans, expression of *MOGAT3* is normally restricted to the gastrointestinal tract with the highest level found in the ileum.[14] Intrahepatic lipid accumulation is associated with its pathological expression in obese patients.[15]

- Variant 8: 7_116870061_G/C

ST7 (suppression of tumorigenicity 7) encodes a low-density lipoprotein receptor-related protein with a cytoplasmic tail that interacts with three proteins involved in signal transduction and/or endocytosis: receptor for activated protein C kinase 1 (RACK1), muscle integrin binding protein (MIBP), and SMAD anchor for receptor activation (SARA).[16] This suggests that ST7, like other proteins in the LDLR superfamily, functions in

signal transduction and/or endocytosis. The exact function of this gene product has not been determined. It was identified and studied as a novel putative tumor suppressor gene.[17-22] Its expression varies depending on the cycling status of cancer cells, and is regulated by the protein arginine methyltransferase PRMT5 during its modulation of cell growth and proliferation.[23,24] Its tumor-suppressing effect could be based on remodeling of the extra-cellular matrix.[25]

ST7 presents a rather large and weak expression, the highest in cervix, present in the ovary but higher in testis, intestine and kidney. Its expression is enhanced in germ cell tumors (Unigene ID:197340). GeoProfiles show an absence of expression in the murine fetal ovary up to birth (GDS663/160591_at), and no sex dimorphic expression at E13 (GDS2223/1418967_a_at), in the developing gonad (GDS2719/1418967_a_at), or in the adult (GDS565/1418967_a_at). In human, its highest expression can be found in fetal or adult brain, in salivary glands, in pancreas and in mammary gland or in fat. Ovaries present a modest expression in all profiles. There might be an upregulation in response to AMH in rat P4 ovaries (GDS3199/1392925_at).

Therefore, although its role is not clear, it does not seem to be restricted to an ovarian function. Thus, the fact that the two fertile brothers are homozygous for the detected variant is not in support for its implication in human infertility and POF.

3. Verification of the segregation of the STAG3 deletion in the family by Sanger sequencing

The sequence of exons 6-8 of the *STAG3* gene was verified by Sanger sequencing in the genomic DNA of all available individuals in the family (Figure S2).

4. Genetic exclusion of *STAG3* locus in other families with infertility

Genetic analyses of four additional families with either POF or POF and male infertility excluded linkage at the *STAG3* locus.

Family 1 is a highly consanguineous family from the Middle East, with four patients affected with primary POF and nine unaffected siblings. Linkage analysis showed a LOD-score of -3.6 or below for all recessive models at the *STAG3* locus.

Family 2 is a consanguineous family from the Middle East, with distantly related parents. Three sisters are affected with secondary POF, and two sisters are fertile. Linkage at the *STAG3* locus indicated a LOD-score of -7.7 or below for all recessive models.

Family 3 is a consanguineous family from the Middle East, with two sisters affected with primary POF and one fertile sister. Due to its small size, this family is not informative for linkage analysis, but homozygosity mapping indicated that no coherent runs of homozygosity (i.e. genomic segments shared at the homozygous state between affected patients but not shared with their non-affected sister) were identified at the *STAG3* locus.

Family 4 is from the La Reunion Island in France. One sister is affected with secondary POF, one is affected with hypofertility without POF, and a brother was sterile and had a metastatic melanoma. There are no unaffected siblings. In this family, linkage analysis at the *STAG3* locus showed a lod-score of -6.8 for recessive model.

III – Characterization of *Stag3*$^{-/-}$ mouse: supplemental results

1. Verification of the insertional mutant at the *Stag3* locus

The insertion was verified at the genomic level (Figure S3). The absence of expression of a Stag3 transcript in *Stag3*$^{-/-}$ mice was verified by Western Blot (Figure S4), and the absence of protein by immunofluorescence assays (Figure S5)

2. Double strands breaks and RAD51 recruitment

We sought to determine whether the deficiency of *Stag3* inhibits the generation of programmed double strand breaks (DSBs) by the nuclease SPO11 at the leptotene stage.[26] Thus, we analyzed the presence of γ-H2AX histone variant, which is phosphorylated during early prophase I in response to SPO11-induced DSBs in an ATM-dependent manner.[27] As seen in Figure S6A, leptotene-like arrested oocytes from *Stag3*$^{-/-}$ mice showed a positive staining that was similar to that observed in the wild type mice (110,44±30 vs 115,94±28 arbitrary units of fluorescence, p=0.57). This suggests that the formation of programmed DSBs is not affected in the mutant mice, and also that DSBs are not repaired in the arrested oocytes.

We further analyzed the kinetics of proteins involved in this DSB-induced signaling cascade. After DSBs are generated, the recombinase RAD51 is recruited to promote homologue searching by strand invasion.[28] In wild type oocytes, RAD51 labeling first appears at leptotene stage and disappears towards pachynema.[29] As shown in Figure S6B, both wild type and *Stag3*-deficient oocytes showed a similar number of RAD51 foci at leptotene. However, it seems likely that these foci are maintained in the mutant oocytes because the meiotic progression stops due to the asynapsis.

3. Cohesin complexes in the absence of STAG3

We further analyzed the presence of other subunits of the cohesin complex. To achieve this, we immunolocalized the cohesin subunits SMC3, SMC1ß, RAD21, RAD21L, and REC8 in wild-type and *Stag3*-depleted oocytes from ovaries at gestational day 15.5. In the mutant oocytes, RAD21 protein co-localized with SYCP3, and to a very limited extent the SMC3, SMC1ß and RAD21l proteins. We did not detect any REC8 labeling (Figure S7).

4. Lack of STAG3 causes premature loss of cohesion

In addition to its role in the formation of the synaptonemal complex,[30,31] the cohesin ring is essential for sister chromatids cohesion. Thus, we analyzed chromosomal cohesion by immunolocalization of centromeres in chromosome spreads using anti-centromeric antibodies (ACA). The wild type females showed a single signal per each chromosome in all their leptonema oocytes (100%; n=40; Figure S8A). Strikingly, in the *Stag3*$^{-/-}$ females most of the oocytes (96%, n=31) showed several chromosomes with doublet signals (11±7,9; n=31) corresponding to unjoined chromatids (Figure S8A). This premature loss of cohesion was not observed in arrested oocytes lacking REC8 cohesin (Figure S8B), which is the cohesin involved in maintaining bivalent and dyad together,[32] and highlights STAG3-containing cohesin complexes as the most relevant in mammals in the maintenance of sister chromatid cohesion in oocytes.

5. Absence of Stag3 leads to male infertility by azoospermia

Male homozygous for the insertional mutation in *Stag3* are completely infertile. *Stag3*$^{-/-}$ testis were analysed at 8 weeks of age, and displayed a complete azoospermia and empty epididymides (Figure 10). Histological analysis by hematoxilin and eosin showed that spermatogenesis block corresponds to an epithelial stage iv arrest (data not shown).

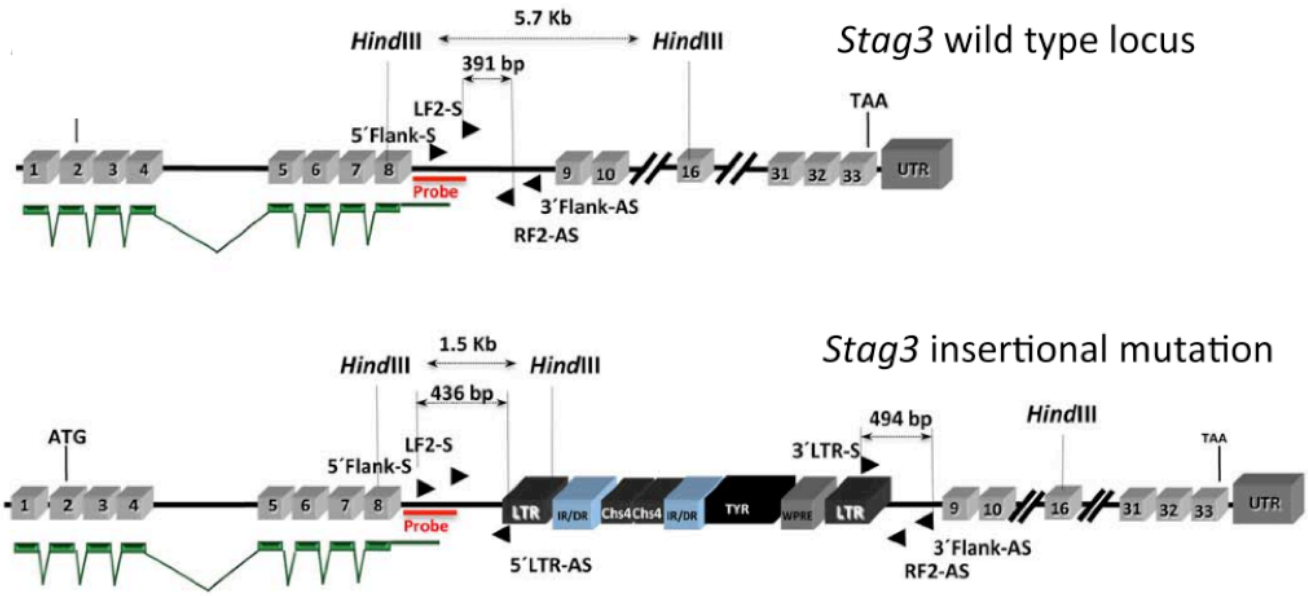

Figure S1: Wild-type locus and insertional mutant at the *Stag3* locus in the OVE2312C mouse

The schematic representation of the wild-type and mutant locus shows the insertion site, the corresponding coding exons (light gray) and noncoding exons (open boxes), and the inserted lentiposon (dark gray). Thin (noncoding) and thick (coding sequences) lines under exons represent the expected transcript derived from wild-type and *Stag3* mutated allele. (ATG) Initiation codon; (TAA) stop codon; (LTR and WPRE) integrated lentiviral sequences; (IR/DR) left and right arms from pT2/Sleeping Beauty; (Chs4) chicken globin HS4 insulator; and (Tyr) tyrosinase minigene.

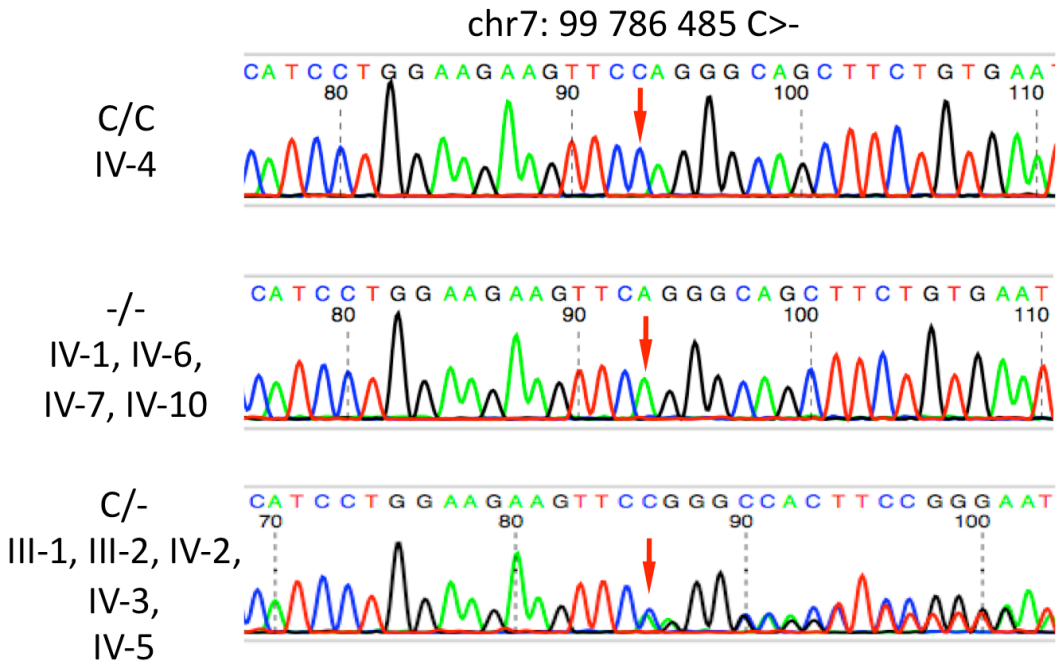

Figure S2: Confirmation of the 1-bp deletion in the *STAG3* gene in the MO1DA family

Sequences obtained by Sanger sequencing of genomic DNA of, from top to bottom, normal individuals (homozygous wild-type), affected individuals (homozygous for the mutation) and healthy carriers (heterozygous for the mutation). The 1-bp deletion in the sequence (position indicated by the red arrow) induces a frameshift in the coding sequence, visible as a superposition of two sequences in the heterozygous DNA.

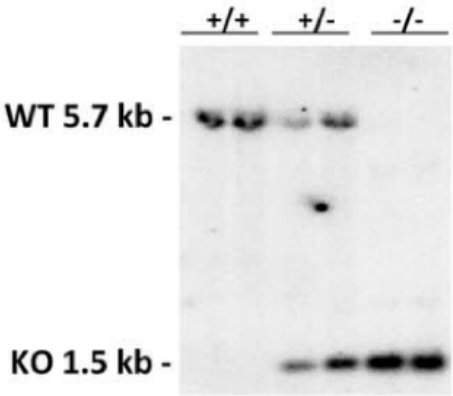

Figure S3: Disruption of the Stag3 locus in the OVE2312C mouse

The Southern blot analysis of genomic DNA from three littermate progeny of *Stag3* heterozygote crosses. HindIII-digested DNA revealed 5.7 kb and 1.5 kb fragments for wild-type and disrupted alleles, respectively. (+/+), (+/−), and (−/−) designate wild-type, heterozygous and homozygous transgenic animals, respectively.

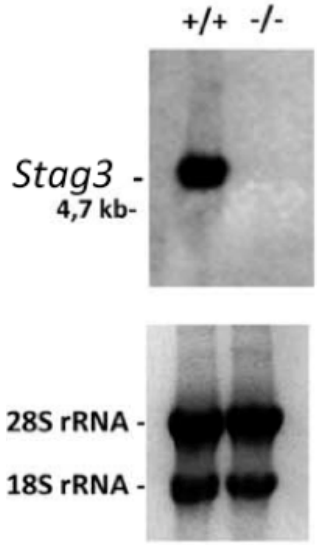

Figure S4: Expression of the *Stag3* gene in wild-type (+/+) and knockout (−/−) animals

Total RNA extract from testis were analyzed by Northern blot of RNA extracts using a full-length *Stag3* cDNA probe. The corresponding ethidium bromide stained 18S and 28S ribosomal RNA used as a loading control is also shown below.

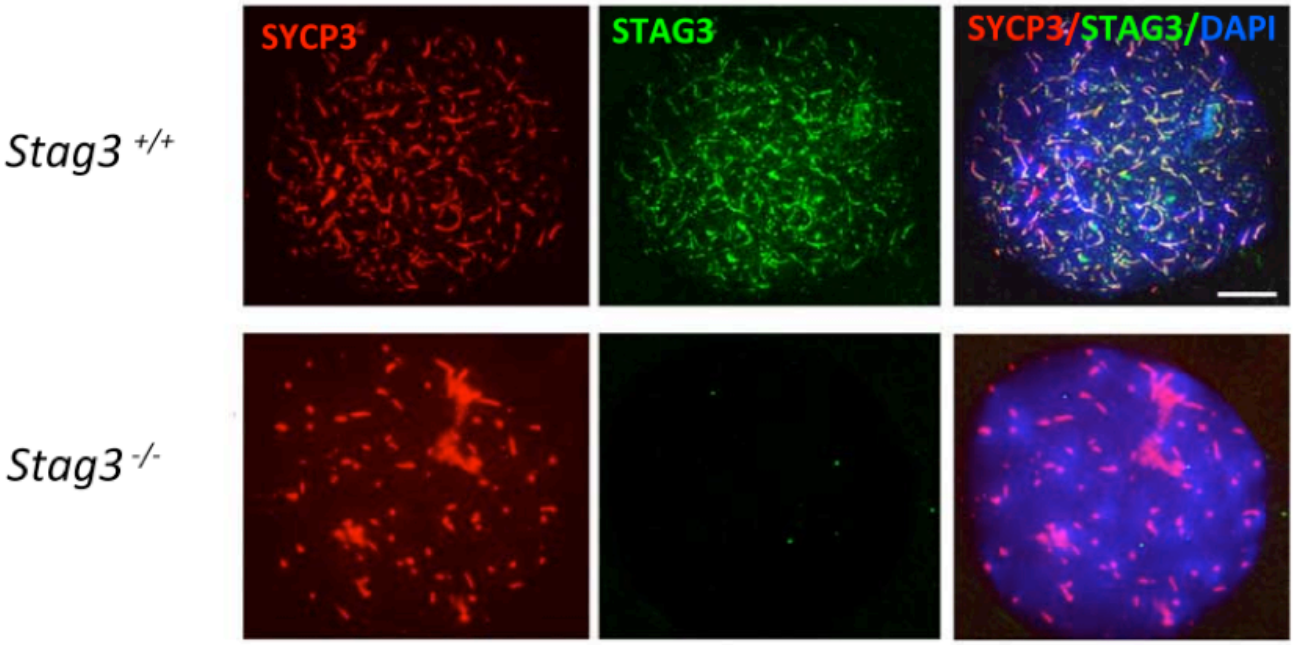

Figure S5: Double immunofluorescence assays for SYCP3 and STAG3 in mouse fetal oocytes

Fetal oocytes were labeled using anti-SYCP3 (detected in red) and anti-STAG3 (labeled in green) antibodies. The absence of STAG3 protein in *Stag3*$^{-/-}$ arrested oocytes implies that the mutant insertional allele is null.

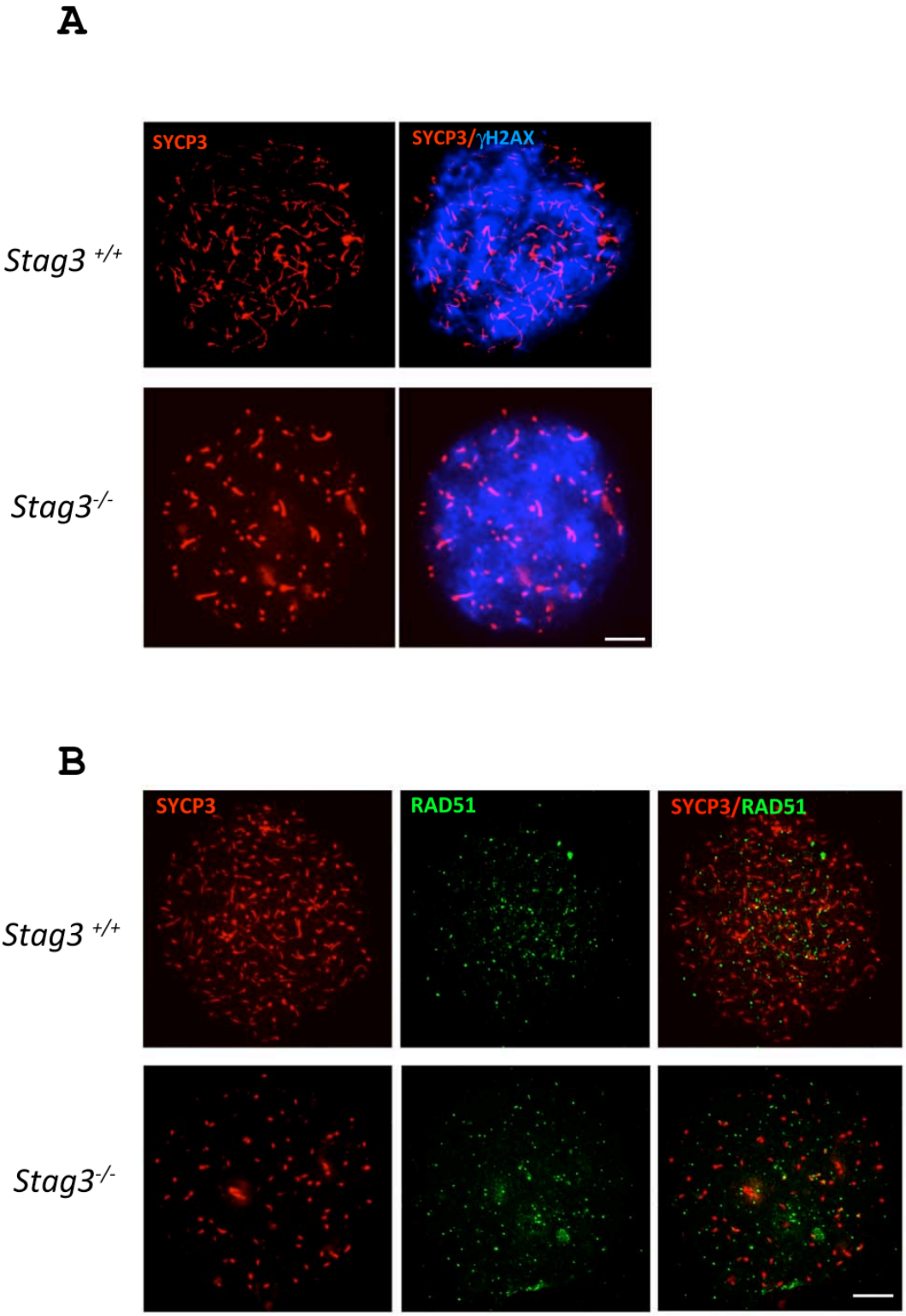

Figure S6: Double Strand Break-associated proteins in *Stag3*$^{-/-}$ oocytes

Double immunolabeling of SYCP3 (red) with γ-H2ax (blue, panel A) and RAD51 (green, panel B) in representative leptotene and leptotene-like oocytes from wild-type and *Stag3*$^{-/-}$ mice. The scale bar in (A) applies to all micrographs and represents 5 µm.

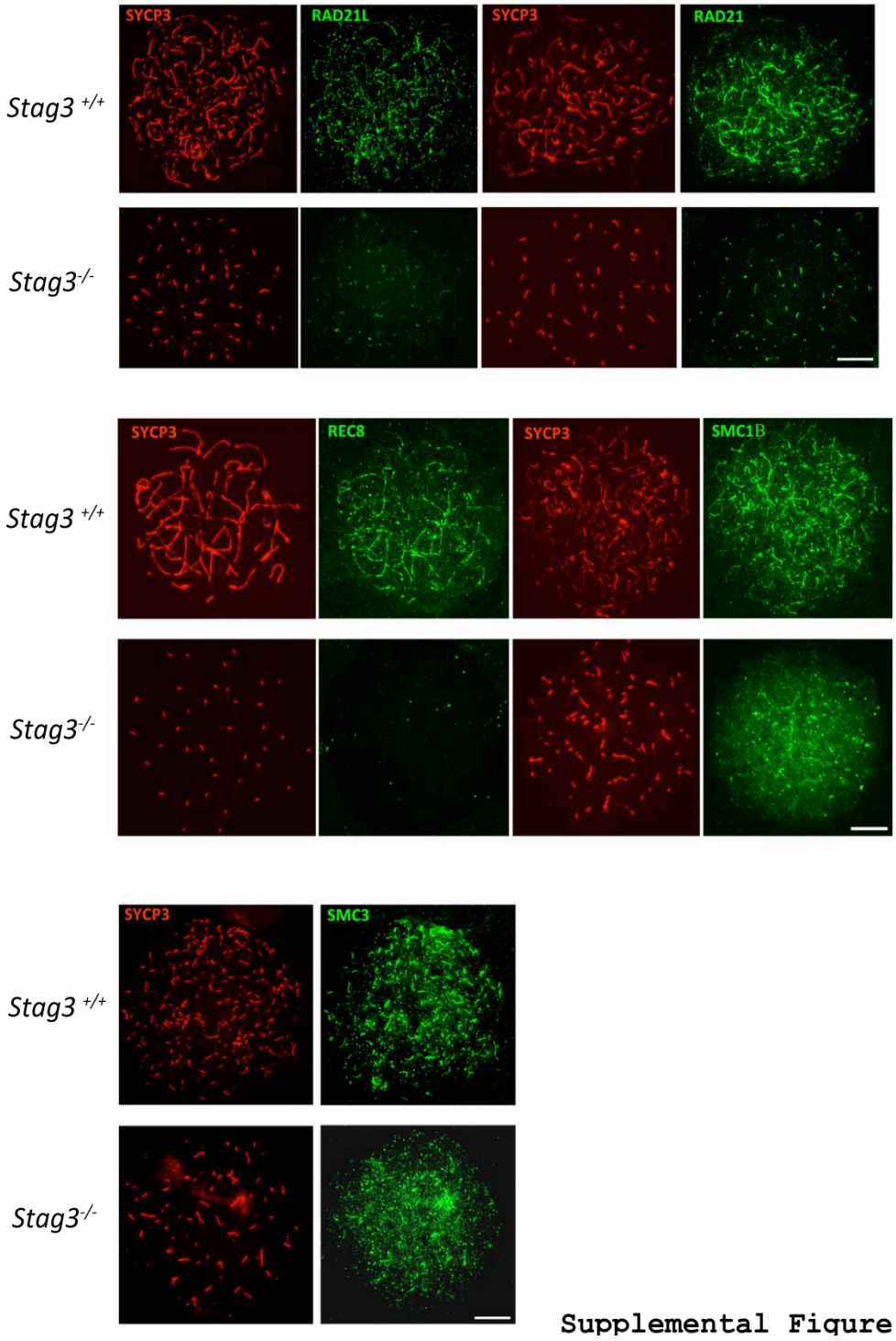

Figure S7: Cohesin complexes in the absence of STAG3

Double immunofluorescence of SYCP3 and either RAD21, REC8, RAD21l, SMC1β, or SMC3. In wild-type leptotene oocytes, the cohesins RAD21, REC8, RAD21L, SMC1β and SMC3, colocalize with SYCP3 along the axial elements on the chromosomes. In Stag3$^{-/-}$ oocytes at gestational day 15.5, arrested at leptonema, REC8 is not detected by immunofluorescence. RAD21L, SMC1β and SMC3 show very faint fluorescence signals, and immunolabeling for RAD21 renders a robust signal. The scale bar applies to all micrographs and represents 5 µm.

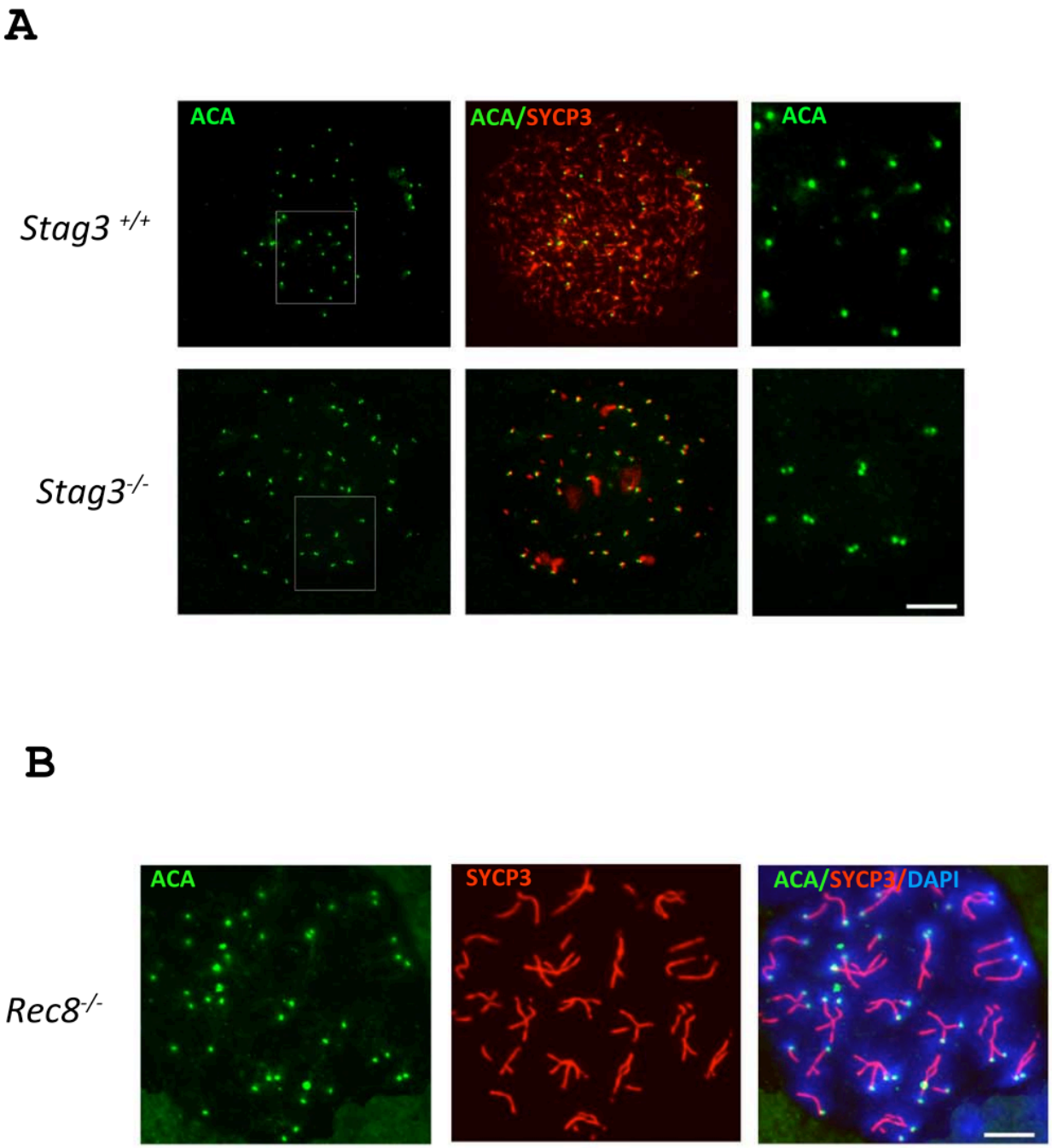

Figure S8: Lack of STAG3 causes premature loss of cohesion

Double immunofluorescence of SYCP3 (red) and anti centromeric antibodies (ACA, green). **A.** Wild-type leptonema oocytes show 40 single signals of ACA. In *Stag3*$^{-/-}$ leptotene-like-arrested oocytes, the ACA signal is observed as doublet in many chromosomes. **B.** Rec8$^{-/-}$ zygotene-like arrested oocytes show 40 single signals of ACA. The scale bar represents 5 µm and 2 µm in the magnified insert.

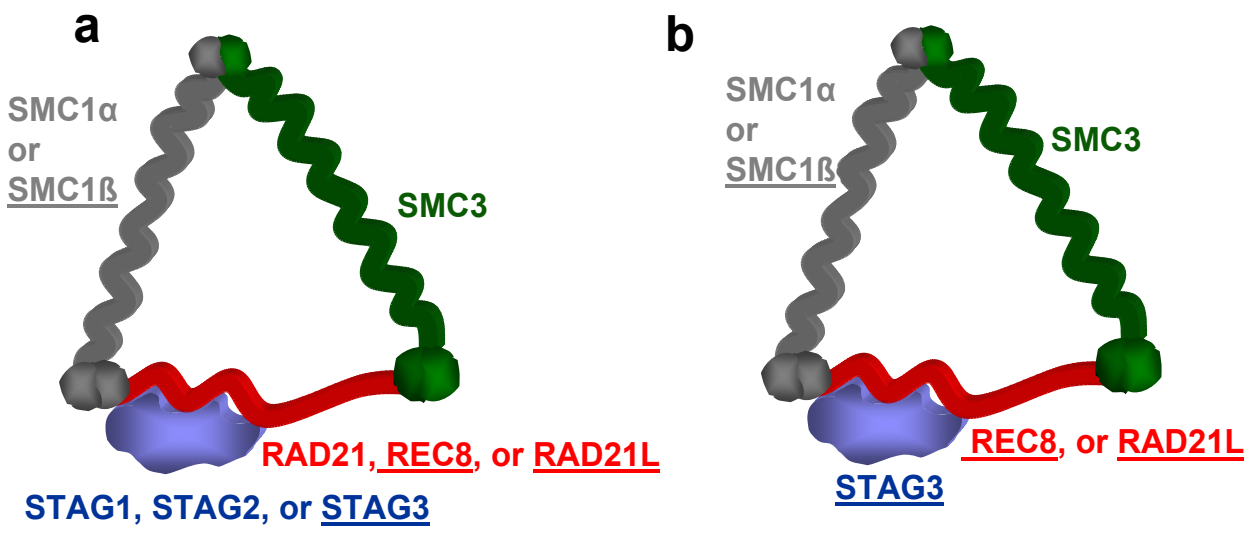

Figure S9: Ring model of the cohesin complex

A. The cohesin complex is constituted by four subunits: two members of the family of structural maintenance of chromosome proteins (SMC1α and SMC3), which heterodimerize, one kleisin subunit that closes the ring (Scc1/RAD21) and is the substrate of the separase protease (ESPL1), and a HEAT repeat domain protein (SA1/STAG1 or SA2/STAG2). In addition, there are meiosis-specific mammalian paralogues of the RAD21, SMC1α, and STAG1/2, which are REC8 and RAD21L, SMC1ß, and STAG3, respectively (underlined). **B.** Theoretically, the STAG3-containing cohesin complexes would be constituted by SMC1β or SMC1α, SMC3 and REC8 or RAD21L. This model is sustained by the analysis of defective meiosis in *Stag3*-deficient oocytes in mice (this study).

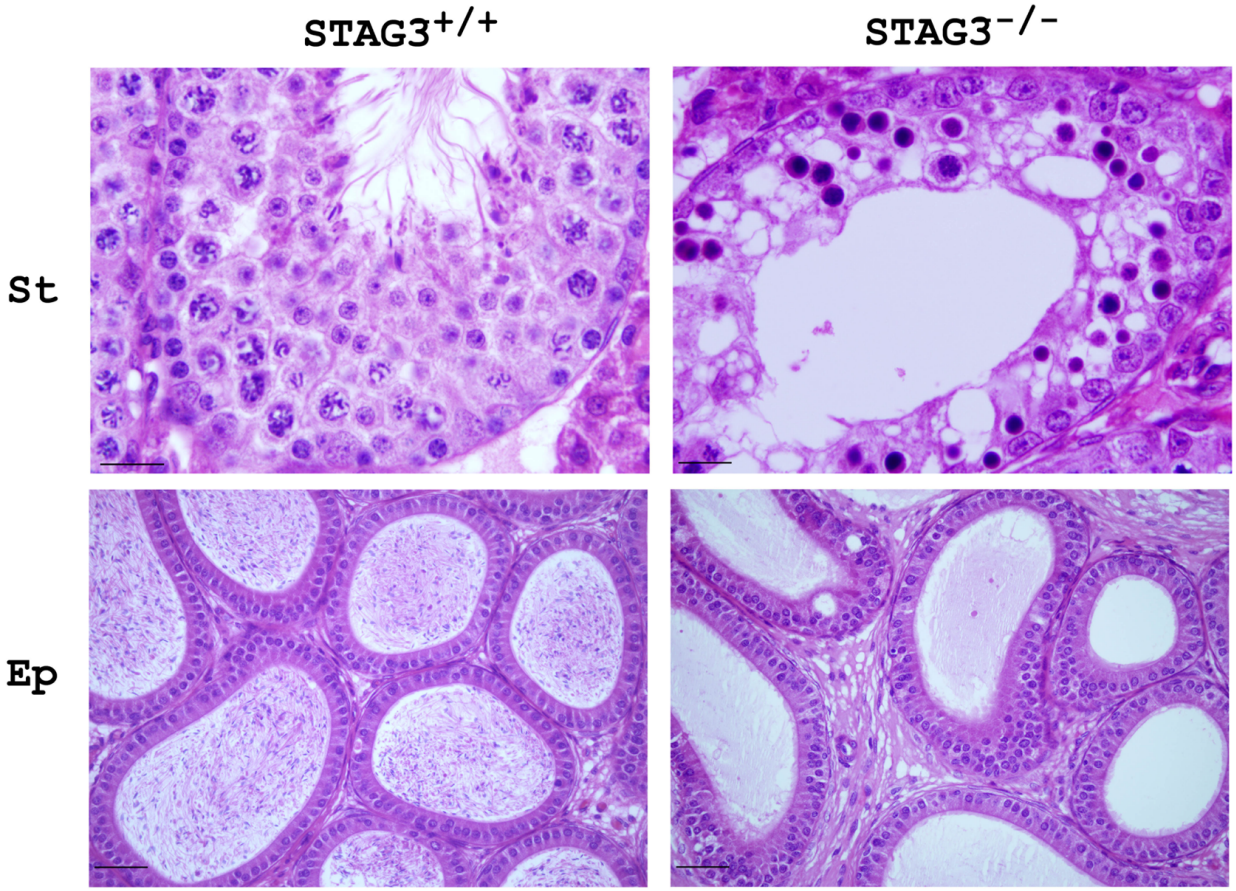

Figure S10: Azoospermia in the *Stag3*$^{-/-}$ male mouse.

The deficiency of STAG3 promotes a complete block of mouse spermatogenesis. The spermatogenic arrest leads to empty epididymides and azoospermia. The scale bar represents 10 µm in upper panels, and 25 µm in lower panels. St stands for Seminiferous tubules and Ep for Epididymides.

Table S1: Primers for creation and characterization of transgenic mice

Primer name		Sequence
tyroA	exon 1,S	5'-GGGGTTGCTGGAAAAGAAGTCTGTG
tyroB	exon 2, AS	5'-GATTACGTAATAGTGGTCCCTCAGG
Nef(F)#1		5'-GGTTTTCCAGTCACACCTCAGGTA
Nef(F)#2		5'-AAGACCAATGACTTACAAGGCAGC
WPRE(QPCR) F		5'- TTTATGAGGAGTTGTGGCCC
WPRE(QPCR) R		5'- CAACACCACGGAATTGTCAG
WPRE(AS)#1		5'-AATAGGGAGGGGGAAAGCGAAAGT
WPRE(AS)#2		5'-AAAGGAGCTGACAGGTGGTGGCAA
3'-LTR(S)		5'-CGTCTGTTGTGTGACTCTGGTAAC
5'-LTR(AS)		5'-CCTGGTGTGTAGCTTTGCCAATCA

Table S2: Primers used for PCR amplification and Sanger sequencing of variants

Region	Forward	Reverse
hSTAG3ex6-8	ATCAAAGCATCCTGTTATTTTCCTT	CTATGAAACTATTGCCAGACTCCA
hSTAG3ex6-8 (seq)	TTTTGACATCCAAGCCCCTA	GCCAAGATGCAGGTAGGAAA
TNRC18	GGAAGAGTCCTCGTCTGTGG	AGACAGAAGGGGAGGAGGAA
ATP5J2-PTCD1	GCCTCTACCTGGTGCTCATC	TGGAACCTTCTCAGGCACTT
MOGAT3	CCTTTTATTGGAGGCATGGA	ACCGAGGAGGAAGTCAATCA
ST7	CTCCTCTGTGCCAACTCCT	GGTTGGAAGACACAGGAGGA
ZAN	GCGCTGGTACACAGAGAACA	AGGGAGTCTGTTGGGGAGTT

Table S3: Quality statistics for the whole-exome sequencing of the two MO1DA sisters

			% of bases with	
	Total Mapped Reads	**Reads on Target**	**>10x coverage**	**>20x coverage**
MO1DA- IV-1	7 953 528 387	3 329 592 720	92,6	88,7
MO1DA- IV-3	7 983 284 423	3 318 917 676	92,6	88,6

Table S4: Numbers of Single Nucleotide Variants identified during the bioinformatics filtering

	Single Nucleotide Variants					
	Total	**Known**	**Novel**	**Shared***	**In the 7p region**	**In the 7q region**
MO1DA- IV-1	20435	19662	773	54	0	**8**
MO1DA- IV-3	20513	19702	811			

* Shared between the two sisters, but homozygous in IV-1 and heterozygous in IV-3.

Table S5: Variants detected by whole-exome sequencing in the MO1DA POF patient

HGNC	Chr	LOCATION	mutation	MO1DA IV-1	MO1DA IV-3	Consequence	Amino_acids	Peptide_Length	Codons
ATP5J2-PTCD1*	7	99,022,873	C>T	homozygous	heterozygous	non-synonymous_coding	A/T	749	Gca/Aca
PTCD1*	7	99,022,873	C>T	homozygous	heterozygous	non-synonymous_coding	A/T	700	Gca/Aca
STAG3	7	99,786,485	C>-	homozygous	heterozygous	Frameshift_coding	.	1225	.
ZAN*	7	100,371,476	->G	homozygous	heterozygous	Frameshift_coding	.	2601	.
MUC12*	7	100,637,008	C>T	homozygous	heterozygous	non-synonymous_coding	A/E	5478	gCa/gAa
MUC12*	7	100,637,035	A>C	homozygous	heterozygous	non-synonymous_coding	H/P	5478	cAc/cCc
MUC12*	7	100,637,041	C>A	homozygous	heterozygous	non-synonymous_coding	T/K	5478	aCa/aAa
MOGAT3*	7	100,839,291	T>A	homozygous	heterozygous	non-synonymous_coding	Q/L	341	cAg/cTg
ST7*	7	116,870,061	G>C	homozygous	heterozygous	non-synonymous_coding	V/L	561	Gtt/Ctt

* See paragraph B I-2 in this Appendix for details about function and/or expression for each gene.

C - Supplementary References

1. Lois C, Hong EJ, Pease S, Brown EJ, Baltimore D. Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. Science 2002;295:868-72.
2. Ikawa M, Tanaka N, Kao WW, Verma IM. Generation of transgenic mice using lentiviral vectors: a novel preclinical assessment of lentiviral vectors for gene therapy. Molecular therapy : the journal of the American Society of Gene Therapy 2003;8:666-73.
3. Overbeek PA, Aguilar-Cordova E, Hanten G, et al. Coinjection strategy for visual identification of transgenic mice. Transgenic research 1991;1:31-7.
4. Wang B, Harrison W, Overbeek PA, Zheng H. Transposon mutagenesis with coat color genotyping identifies an essential role for Skor2 in sonic hedgehog signaling and cerebellum development. Development 2011;138:4487-97.
5. Lu B, Geurts AM, Poirier C, et al. Generation of rat mutants using a coat color-tagged Sleeping Beauty transposon system. Mamm Genome 2007;18:338-46.
6. Geurts AM, Yang Y, Clark KJ, et al. Gene transfer into genomes of human cells by the sleeping beauty transposon system. Molecular therapy : the journal of the American Society of Gene Therapy 2003;8:108-17.
7. Yusufzai TM, Felsenfeld G. The 5'-HS4 chicken beta-globin insulator is a CTCF-dependent nuclear matrix-associated element. Proc Natl Acad Sci U S A 2004;101:8620-4.
8. Ganss R, Montoliu L, Monaghan AP, Schutz G. A cell-specific enhancer far upstream of the mouse tyrosinase gene confers high level and copy number-related expression in transgenic mice. EMBO J 1994;13:3083-93.
9. Rackham O, Davies SM, Shearwood AM, Hamilton KL, Whelan J, Filipovska A. Pentatricopeptide repeat domain protein 1 lowers the levels of mitochondrial leucine tRNAs in cells. Nucleic acids research 2009;37:5859-67.
10. Gao Z, Harumi T, Garbers DL. Chromosome localization of the mouse zonadhesin gene and the human zonadhesin gene (ZAN). Genomics 1997;41:119-22.
11. Gasper J, Swanson WJ. Molecular population genetics of the gene encoding the human fertilization protein zonadhesin reveals rapid adaptive evolution. Am J Hum Genet 2006;79:820-30.
12. Desseyn JL, Tetaert D, Gouyer V. Architecture of the large membrane-bound mucins. Gene 2008;410:215-22.
13. Matsuyama T, Ishikawa T, Mogushi K, et al. MUC12 mRNA expression is an independent marker of prognosis in stage II and stage III colorectal cancer. Int J Cancer 2010;127:2292-9.
14. Cheng D, Nelson TC, Chen J, et al. Identification of acyl coenzyme A:monoacylglycerol acyltransferase 3, an intestinal specific enzyme implicated in dietary fat absorption. J Biol Chem 2003;278:13611-4.
15. Hall AM, Kou K, Chen Z, et al. Evidence for regulated monoacylglycerol acyltransferase expression and activity in human liver. Journal of lipid research 2012;53:990-9.

16. Battle MA, Maher VM, McCormick JJ. ST7 is a novel low-density lipoprotein receptor-related protein (LRP) with a cytoplasmic tail that interacts with proteins related to signal transduction pathways. Biochemistry 2003;42:7270-82.

17. Thomas NA, Choong DY, Jokubaitis VJ, Neville PJ, Campbell IG. Mutation of the ST7 tumor suppressor gene on 7q31.1 is rare in breast, ovarian and colorectal cancers. Nat Genet 2001;29:379-80.

18. Zenklusen JC, Conti CJ, Green ED. Mutational and functional analyses reveal that ST7 is a highly conserved tumor-suppressor gene on human chromosome 7q31. Nat Genet 2001;27:392-8.

19. Dong SM, Sidransky D. Absence of ST7 gene alterations in human cancer. Clin Cancer Res 2002;8:2939-41.

20. Lu C, Xu HM, Ren Q, et al. Somatic mutation analysis of p53 and ST7 tumor suppressor genes in gastric carcinoma by DHPLC. World journal of gastroenterology : WJG 2003;9:2662-5.

21. Sivasundaram K, Suzuki H, Seto M, Hosokawa Y. Mutational analysis of the ST7 gene in human myeloid tumor cell lines. Oncol Rep 2003;10:1737-9.

22. Wang S, Mori Y, Sato F, et al. An LOH and mutational investigation of the ST7 gene locus in human esophageal carcinoma. Oncogene 2003;22:467-70.

23. Pal S, Vishwanath SN, Erdjument-Bromage H, Tempst P, Sif S. Human SWI/SNF-associated PRMT5 methylates histone H3 arginine 8 and negatively regulates expression of ST7 and NM23 tumor suppressor genes. Molecular and cellular biology 2004;24:9630-45.

24. Hooi CF, Blancher C, Qiu W, et al. ST7-mediated suppression of tumorigenicity of prostate cancer cells is characterized by remodeling of the extracellular matrix. Oncogene 2006;25:3924-33.

25. Charong N, Patmasiriwat P, Zenklusen JC. Localization and characterization of ST7 in cancer. Journal of cancer research and clinical oncology 2011;137:89-97.

26. Keeney S. Mechanism and control of meiotic recombination initiation. Current topics in developmental biology 2001;52:1-53.

27. Mahadevaiah SK, Turner JM, Baudat F, et al. Recombinational DNA double-strand breaks in mice precede synapsis. Nat Genet 2001;27:271-6.

28. Symington LS, Gautier J. Double-strand break end resection and repair pathway choice. Annual review of genetics 2011;45:247-71.

29. Tarsounas M, Morita T, Pearlman RE, Moens PB. RAD51 and DMC1 form mixed complexes associated with mouse meiotic chromosome cores and synaptonemal complexes. The Journal of cell biology 1999;147:207-20.

30. Herran Y, Gutierrez-Caballero C, Sanchez-Martin M, et al. The cohesin subunit RAD21L functions in meiotic synapsis and exhibits sexual dimorphism in fertility. EMBO J 2011;30:3091-105.

31. Llano E, Herran Y, Garcia-Tunon I, et al. Meiotic cohesin complexes are essential for the formation of the axial element in mice. The Journal of cell biology 2012;197:877-85.

32. Tachibana-Konwalski K, Godwin J, van der Weyden L, et al. Rec8-containing cohesin maintains bivalents without turnover during the growing phase of mouse oocytes. Genes & development 2010;24:2505-16