



**University of Dundee**

**Mutations involving the SRY-related gene SOX8 are associated with a spectrum of human reproductive anomalies.**

Portnoi, Marie-France; Dumargne, Marie-Charlotte; Rojo, Sandra; Witchel, Selma F.; Duncan, Andrew J.; Eozenou, Caroline; Bignon-Topalovic, Joelle; Yatsenko, Sventlana A.; Rajkovic, Aleksandar; Reyes-Mugica, Miguel; Almstrup, Kristian; Fusee, Lelia; Srivastava, Yogesh; Chantot-Bastarud, Sandra; Hyon, Capucine; Louis-Sylvestre, Christine; Validire, Pierre; de Malleray Pichard, Caroline; Ravel, Celia; Christin-Maitre, Sophie; Brauner, Raja; Rossetti, Raffaella; Persani, Luca; Charreau, Eduardo H.; Dain, Liliana; Chiauzzi, Violeta A.; Mazen, Inas; Rouba, Hassan; Schluth-Bolard, Caroline; MacGowan, Stuart; McLean, W. H. Irwin; Patin, Etienne; Rajpert-De Meyts, Ewa; Jauch, Ralf; Achermann, John C.; Siffroi, Jean-Pierre; McElreavey, Ken; Bashamboo, Anu

*Published in:*

Human Molecular Genetics

*DOI:*

[10.1093/hmg/ddy037](https://doi.org/10.1093/hmg/ddy037)

*Publication date:*

2018

*Document Version*

Peer reviewed version

[Link to publication in Discovery Research Portal](#)

*Citation for published version (APA):*

Portnoi, M-F., Dumargne, M-C., Rojo, S., Witchel, S. F., Duncan, A. J., Eozenou, C., ... Bashamboo, A. (2018). Mutations involving the SRY-related gene SOX8 are associated with a spectrum of human reproductive anomalies. *Human Molecular Genetics*. DOI: [10.1093/hmg/ddy037](https://doi.org/10.1093/hmg/ddy037)

**Mutations involving the *SRY*-related gene *SOX8* are associated with a spectrum of  
human reproductive anomalies**

Marie-France Portnoi<sup>1,2¶</sup>, Marie-Charlotte Dumargne<sup>3¶</sup>, Sandra Rojo<sup>3</sup>, Selma F Witchel<sup>4</sup>, Andrew J Duncan<sup>5</sup>, Caroline Eozenou<sup>3</sup>, Joelle Bignon-Topalovic<sup>3</sup>, Sventlana A Yatsenko<sup>6,7,8</sup>, Aleksandar Rajkovic<sup>6,7,8</sup>, Miguel Reyes-Mugica<sup>6,7,8</sup>, Kristian Almstrup<sup>9</sup>, Lelia Fusee<sup>3</sup>, Yogesh Srivastava<sup>10</sup>, Sandra Chantot-Bastarud<sup>1,2</sup>, Capucine Hyon<sup>1,2</sup>, Christine Louis-Sylvestre<sup>11</sup>, Pierre Validire<sup>12</sup>, Caroline de Malleray Pichard<sup>13</sup>, Celia Ravel<sup>14</sup>, Sophie Christin-Maitre<sup>2,15</sup>, Raja Brauner<sup>16</sup>, Raffaella Rossetti<sup>17</sup>, Luca Persani<sup>17</sup>, Eduardo H Charreau<sup>18</sup>, Liliana Dain<sup>18</sup>, Violeta A Chiauzzi<sup>18</sup>, Inas Mazen<sup>19</sup>, Hassan Rouba<sup>20</sup>, Caroline Schluth-Bolard<sup>21</sup>, Stuart MacGowan<sup>22</sup>, WH Irwin McLean<sup>22</sup>, Etienne Patin<sup>23</sup>, Ewa Rajpert-De Meyts<sup>9</sup>, Ralf Jauch<sup>10</sup>, John C Achermann<sup>5</sup>, Jean-Pierre Siffroi<sup>1,2</sup>, Ken McElreavey<sup>3&</sup>, Anu Bashamboo<sup>\*3&</sup>

1-APHP Département de Génétique médicale, Hôpital Armand Trousseau, Paris, 75012, France.

2-UPMC, University Paris 06, INSERM UMR\_S933, Hôpital Armand Trousseau, Paris, 75012, France.

3-Human Developmental Genetics, CNRS UMR3738, Institut Pasteur, Paris, France.

4-Division of Pediatric Endocrinology, Children's Hospital of Pittsburgh of UPMC, Pittsburgh, PA, USA

5-Genetics and Genomic Medicine, UCL Great Ormond Street Institute of Child Health, London WC1N 1EH, U.K.

6-Department of Obstetrics, Gynecology and Reproductive Sciences, Magee-Womens Research Institute, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

7-Department of Human Genetics, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

8-Department of Pathology, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

- 9-University Department of Growth and Reproduction, Rigshospitalet, DK-2100, Copenhagen, Denmark
- 10-<sup>[1]</sup>Genome Regulation Laboratory, Drug Discovery Pipeline, <sup>[2]</sup>Key Laboratory of Regenerative Biology, <sup>[3]</sup>Guangdong Provincial Key Laboratory of Stem Cell and Regenerative Medicine, South China Institute for Stem Cell Biology and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China.
- 11-Département Mère-Enfant, Institut Mutualiste Montsouris, Paris, France.
- 12-Département d'Anatomie Pathologique, Institut Mutualiste Montsouris, Paris, France.
- 13-Service d'Endocrinologie, Hôpital Cochin, Paris, France.
- 14-Biology of Reproduction, CHU Rennes, Rennes, France
- 15-Service d'Endocrinologie, diabétologie et endocrinologie de la reproduction, Hôpital Saint-Antoine, Paris, 75012, France.
- 16-Université Paris Descartes and Pediatric Endocrinology Unit, Fondation Ophtalmologique Adolphe de Rothschild, Paris, France.
- 17-Department of Clinical Sciences & Community Health, University of Milan, 20100 Milan, Italy. Laboratory of Endocrine & Metabolic Research and Division of Endocrine and Metabolic Diseases, IRCCS Istituto Auxologico Italiano, Piazzale Brescia 20, 20149 Milan, Italy.
- 18-Centro Nacional de Genética Médica, Administración Nacional de Laboratorios e Institutos de Salud (ANLIS), Dr. Carlos G. Malbrán, Buenos Aires, Argentina; Instituto de Biología y Medicina Experimental, Consejo Nacional de Investigaciones Científicas y Técnicas (IBYME-CONICET), Buenos Aires, Argentina.
- 19-National research Centre, Cairo, Egypt
- 20-Human Genetics Unit, Institut Pasteur of Morocco, Casablanca, 20100, Morocco.
- 21-Genetic Services, Hopital Femme, Mere, Enfant, 69677, Bron, France.
- 22-Centre for Dermatology and Genetic Medicine, School of Life Sciences, University of Dundee, Dundee, U.K
23. Human Evolutionary Genetics, Institut Pasteur, Paris, France.

**Corresponding author:** Dr Anu Bashamboo, Human Developmental Genetics, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cedex 15, France. Tel : 0033145688920, email : [anu.bashamboo@pasteur.fr](mailto:anu.bashamboo@pasteur.fr)

<sup>¶</sup>These authors contributed equally to this work.

<sup>&</sup>These authors also contributed equally to this work.

## Abstract

SOX8 is an HMG-box transcription factor closely related to SRY and SOX9. Deletion of the gene encoding Sox8 in mice causes reproductive dysfunction but the role of SOX8 in humans is unknown. Here, we show that SOX8 is expressed in the somatic cells of the early developing gonad in the human and influences human sex-determination. We identified two individuals with 46,XY disorders/differences in sex development (DSD) and chromosomal rearrangements encompassing the *SOX8* locus and a third individual with 46,XY DSD and a missense mutation in the HMG-box of SOX8. *In-vitro* functional assays indicate that this mutation alters the biological activity of the protein. As an emerging body of evidence suggests that DSDs and infertility can have common etiologies, we also analyzed *SOX8* in a cohort of infertile men (n=274) and two independent cohorts of women with primary ovarian insufficiency (POI; n=153 and n=104). *SOX8* mutations were found at increased frequency in oligozoospermic men (3.5%;  $p < 0.05$ ) and POI (5.06%;  $p = 4.5 \times 10^{-5}$ ) as compared to fertile/normospermic control populations (0.74%). The mutant proteins identified altered SOX8 biological activity as compared to the wild-type protein. These data demonstrate that SOX8 plays an important role in human reproduction and *SOX8* mutations contribute to a spectrum of phenotypes including 46,XY DSD, male infertility and 46,XX POI.

## Introduction

Human sex determination is a tightly controlled and highly complex process, where the bipotential gonad anlage develops as one of two mutually antagonistic fates – the ovary or a testis. *SRY* is well-established as the primary testis determining gene on the Y chromosome. In the XY gonad, *SRY* acts during embryonal development to upregulate the downstream effector *SOX9* beyond a critical threshold, which promotes testis development and in turn represses “ovarian pathways” in the male gonad through multiple mechanisms [1,2]. In mice, up-regulation of *Sox9* in males is by the synergistic action of *Sry* with *Nr5a1* (also known as Steroidogenic factor-1 (Sf1)), through binding to multiple elements within a testis-specific *Sox9* enhancer (*Tesco*) [1]. *Sox9* expression is required to establish Sertoli cell identity in the developing testis [2]. Once formed, foetal Sertoli cells produce anti-Müllerian hormone (AMH) and coordinate the cellular and morphogenetic events leading to primary sex determination [2].

Although the *SOX* genes *SRY* and *SOX9* play essential roles in driving early mammalian testis-determination, emerging evidence suggests the contribution of another *Sox* gene, *Sox8*, in murine testis-determination as well as in the maintenance of gonadal function. *Sox8* shows an overlapping expression pattern with *Sox9* in foetal and adult mouse gonad and functional redundancy between *Sox8* and *Sox9* occurs during testis development [3-6]. *Sox8* XY null mutants show normal testis development but develop postnatal progressive spermatogenic failure [7].

Development and maintenance of the mammalian gonad is regulated by a double repressive system, where an equilibrium of mutually antagonistic pathways must be attained for normal development of either the testis or ovaries. [2-6]. In humans, changes in this delicate balance

can lead to Disorders of (or Differences in) Sex Development (DSD), which are defined as congenital conditions with discordant development of chromosomal, gonadal or anatomical sex [8]. The incidence of DSD has been estimated at 1:2,500 to 1:5,000 births [9-10]. Excluding inborn errors in steroidogenesis, a molecular diagnosis will be reached in only 20% of all individuals with 46,XY DSD [9]. Around 40% of the cases with 46,XY complete gonadal dysgenesis (CGD) can be explained by mutations involving three genes *SRY*, *NR5A1* and *MAP3K1* [4,5]. Pathogenic variants in other sex-determining genes such as *CBX2*, *GATA4* or *FOG2/ZFPM2* are found in a very small percentage of cases [9, 10]. Hence, the aetiology in majority of the individuals with DSD remain poorly understood.

An emerging body of evidence suggests that DSD and infertility can have common aetiologies, For example, mutations involving NR5A1 (SF-1), a key player in many aspects of reproductive function including sex-determination, are associated not only with a spectrum of DSD such as 46,XY CGD, 46,XY undervirilised males with testes, or 46,XX (ovo)testicular DSD but also more prevalent forms of human infertility, i.e. 46,XY men with spermatogenic failure and 46,XX women with primary ovarian insufficiency (POI) [11-14]. Human infertility is a major human health issue; one in seven couples worldwide have problems conceiving and both men and women are affected equally [15,16]. POI is characterised by primary or secondary amenorrhea, high gonadotropin levels (FSH above 40iU/l on two occasions at least a month apart) and estrogen deficiency in women under the age of 40 years [17]. Male infertility includes azoospermia characterized by complete absence of spermatozoa in the ejaculate, whereas oligozoospermia is defined as sperm concentrations below the World Health Organization reference level of  $15 \times 10^6$  sperm/ml [18]. The underlying basis for either male or female infertility is complex, including both physiological and environmental factors as well as gene mutations. Although mouse models have revealed hundreds of genes that are

associated with fertility, only a few single-gene defects that cause male and/or female infertility have been identified in humans [12,13,19-24]. Here, for the first time we demonstrate that mutations involving the human *SOX8* gene are associated with a range of phenotypes including 46,XY DSD as well as male infertility and ovarian insufficiency in females.

## Results

**Rearrangements involving the *SOX8* locus and 46,XY DSD.** Patient 1 is a phenotypic female who presented at the age of 27 years with primary amenorrhea (Patient 1, Table 1). Chromosome analysis showed a 46,XY karyotype with a paracentric inversion (46,XY,inv(16)(p13.3p13.1); Fig. 1A, 1B), which was confirmed by FISH analysis (Fig. 1C). Her parents and unaffected sister were not available for analysis. Hence, it is unknown if this rearrangement is *de novo*, however this rearrangement has not been reported in control databases. Array-CGH did not reveal any chromosomal imbalances, including the chr16 short arm, associated with the phenotype. The centromeric breakpoint was mapped within the BAC clone RP11-609N14 (chr16:10428838-10600163) (Fig. 1D). The telomeric breakpoint was mapped within the BAC clone RP11-728H8 between chr16:814190-962809 (Fig. 1E). Whole genome sequencing confirmed these results and indicated a centromeric breakpoint at chr16:10522104-10522114 and a telomeric breakpoint at chr16:881929-881965. The transcription initiation site of the *SOX8* gene is at chr16:1031808. Therefore, the breakpoint is located ~150 Kb upstream of *SOX8* (Fig. 1A). A bilateral gonadectomy revealed two small gonads, which on histological examination showed two streak-like gonads with no germ cells or differentiated tissue (Fig. 2A).

Patient 2 is a 46,XY girl with mild post-natal anaemia, sleepiness, poor feeding and tachycardia (Patient 2, Table 1 and Supplemental information). Physical examination showed



a prominent forehead and ophthalmological evaluation revealed small bilateral chorioretinal colobomas. At birth the genitalia consisted of a midline clitorophallic structure with bilateral labioscrotal folds, which were fused at the midline. On pelvic ultrasound, there was no uterus or vagina noted and gonads and epididymides were identified in the labia bilaterally. Pituitary hormone levels were within normal limits. The child was assigned male and at 10 months of age underwent the first stage of surgery for hypospadias, chordee repair and posterior urethral mobilization, and scrotoplasty. At 2 yrs 10 months the patient underwent bilateral orchidopexy, bilateral inguinal hernia repair, and bilateral gonadal biopsy. Gonadal histology showed seminiferous cords, variable germ cells and reduced Leydig cell numbers by Calretinin staining (Fig. 2B-E). aCGH detected a two copy DNA gain in the 16p13.3 region (arr[h19] 16p13.3(137893-992302)x4, log<sub>2</sub>=+1), spanning at least 854 kb (Fig 1F), resulting in tetrasomy of that DNA segment. This region comprises at least 27 genes, including the  $\alpha$ -globin gene cluster (S1 Table). FISH analyses using RP11-598I20 BAC probe on cultured peripheral blood samples from the patient and both parents indicated that the tetrasomy was a *de novo* event (Fig. 1G, 1H). The centromeric breakpoint of this triplication was located within the intron 2 of the *LMFI* (lipase maturation factor 1) gene, ~39.5 kb upstream of the *SOX8* gene. Homozygous, loss-of-function *LMFI* gene mutations are responsible for hypertriglyceridemia and decreased lipase activity and hence this gene was not considered to be associated with the phenotype [25]. Interestingly, *SOX8*-specific enhancer elements are included in the triplication [26]. Together these data suggest that rearrangements at the *SOX8* resulting in dysregulation of *SOX8* expression could negatively impact testis-determination and may result in 46,XY DSD including disorders of testis-determination. Consistent with this postulation, we observed that *SOX8* is co-expressed with *NR5A1* and *SOX9* in the early stages of human testis-determination in Sertoli cells and Leydig cells as well as in Sertoli and Leydig cells in adult men (Fig. 3A, S1 Fig). Single cell RNA sequencing on XY mouse

gonads during sex determination has also demonstrated that *Sox8* is co-expressed with other genes involved in sex-determination including *Nr5a1*, *Dmrt1* and *Gata4* [27].

**A missense mutation in the HMG-box domain of SOX8 is associated with a lack of testis-determination.** Based on the above observations, we extended the *SOX8* mutation screen to a large cohort of 46,XY DSD individuals (Table 2). Three heterozygous missense *SOX8* variants were identified in a screen of 204 cases of unexplained 46,XY DSD. Two of these changes were observed in the control cohort and their contribution to the phenotype is considered unlikely. The third mutation, a heterozygous c.468G>C p.Glu156Asp amino acid substitution was identified in a 46,XY phenotypic female who presented at 16 years of age with primary amenorrhea (Patient 3 Table 1, Fig 4A). The Glu156 residue is located in the HMG-box and is evolutionarily conserved not only within the SOX8 protein but also in other members of groups E and F SOX protein family (Fig. 4B). The mode of inheritance of this mutation is unknown as the parents were unavailable for study. This mutation was not observed in our control cohort, but it is present as a rare variant in the ExAC database (2:121,228 alleles, rs764098477). The SOX8 p.Glu156Asp mutation is not expected to disrupt DNA binding *per se* because this residue does not map to the DNA interaction interface (Fig 4C. Furthermore, exchanging the equivalent residues in Sox2 (Sox2Lys to Glu) and Sox17 (Sox17Glu to Lys) are reported not to show any effect on DNA binding by the SOX HMG box in the absence of partner proteins [28]. However, this mutation could affect how SOX8 dimerizes with its partner factors on DNA, since changing the same residue in Sox2 and Sox17 affected the ability of these proteins to interact with OCT4 (Fig. 4C [28].

A series of *in-vitro* analyses were performed to assess the effect of this mutant on the biological activity of the SOX8 protein. Transient transfection assays were performed since a

number of *SOXE* group-responsive genomic elements have been identified that, when placed upstream of a reporter gene, can be used to quantitate transcriptional regulatory function of protein variants [29]. The function of *SOX* genes involves a complex interaction with many other transcriptional co-regulators, including other SOX proteins [29] and therefore we considered that the consequence of the *SOX8* mutation may be promoter or cell context dependent. Our data show that the mutant SOX8 protein exhibits a context-specific loss-of-function activity. The WT-and mutant SOX8 proteins can activate a series of gonadal promoters (Fig. 5A). However, the SOX8p.Glu156Asp specifically fails to synergize with NR5A1 to transactivate the *Sox9 Tesco* enhancer element (Fig. 5A), yet both WT-SOX8 and SOX8p.Glu156Asp proteins physically interact with the NR5A1 protein (Fig. 6, S2 Fig). The mutation also impacted on the functional interaction between SOX8 and SOX9. Both SOXE proteins have the ability to heterodimerize and combinatorially regulate their target gene expression [30, 31]. We show that the WT-SOX8 protein binds to SOX9, however the SOX8p.Glu156Asp protein cannot (Fig. 6, S2 Fig). The SOX8p.Glu156Asp mutant also exerts a repressive effect by negatively affecting the synergistic activation of the *Tesco* enhancer by NR5A1 and SOX9 (Fig. 5B). Finally, the SOX8p.Glu156Asp protein also has a dominant negative effect on the WT-SOX8 protein since it reduced, in a dose-dependent manner, the synergistic activation of the *Tesco* reporter by WT-SOX8 and NR5A1 (S3 Fig). All these results are consistent with the hypothesis that the mutant SOX8 protein has an altered biological activity impairs testis-determination

***SOX8* mutations associated with human infertility.** To investigate whether mutations in *SOX8*, like *NR5A1*, could also contribute to a wider spectrum of male and female infertility, we sequenced the coding region of *SOX8* in 427 men and women with infertility. The incidence of mutations in the control cohort of fertile and/or normospermic men was 0.74%.

In contrast, rare or novel *SOX8* variants were observed in 1.47% of azoospermic men, 3.5% of oligozoospermic men ( $p=0.0151$ , Fishers t-test, 2-tailed) and 9/153 women with POI in an initial cohort (5.9%;  $p=0.000107$ ) and in 4/104 (3.85%;  $p=0.0191$ ) of women with POI in a replication cohort (Tables 2 and 3). Combining the data from both cohorts, we found a significant association of mutations in *SOX8* with POI (5.06%;  $p=4.5 \times 10^{-5}$ ).

The majority of mutations associated with infertility flank the HMG-box and fall within one of the two transactivation domains. Although many of these missense mutations are predicted by computational methods to have a deleterious effect on protein function, the interpretation of the significance of mutations in human disease remains very challenging. We therefore sought to test the consequence of the *SOX8* mutations associated with infertility in a series of functional assays. As expected, the transient transfection assays using the *AMH*, *Dmrt1* and *NR5A1* promoters as reporters in HEK 293-T and mES cells indicated biological differences between mutated wild-type proteins that, in some cases, were promoter specific, e.g. the p.Asp382Asn mutation shows a specific loss-of-function with the *AMH* promoter (S4 Fig). Furthermore, some mutant proteins show changes in cellular localization (S5 Fig).

Although, a contribution of Sox8 to male infertility in the mouse has been established for some time, the finding of an association between human POI and SOX8 variants is a novel finding since *Sox8*<sup>-/-</sup> female mice are fertile [7]. We therefore sought to examine the profile of SOX8 expression in the human ovary to see if this is consistent with a role for mutations in the protein contributing to female infertility. We observed that the SOX8 protein is expressed in the 40 week foetal ovary and in the adult ovary (Fig. 3B and S1 Fig). The SOX8 protein is highly expressed in granulosa cells at both stages.

## Discussion

Our data provide the first evidence that *SOX8* plays a role in human sex-determination and in the function of the adult gonad. Mouse models also suggest that *Sox8* plays a key role in early testis-determination as well as maintaining male fertility. We identified three patients with 46,XY DSD who carried rearrangements/mutations involving the *SOX8* gene.

One patient had a paracentric inversion with a breakpoint ~150 Kb upstream of the *SOX8* gene. The only other gene within the region is *LMFI*. Since, homozygous, loss-of-function mutations in the *LMFI* gene are associated with hypertriglyceridemia and decreased lipase activity, this gene was not considered to be responsible for the phenotype [25].

A second patient carried four copies of an 854 Kb region immediately 5' to the *SOX8* gene including *SOX8* enhancer elements as well as the  $\alpha$ -globin gene cluster [26]. Duplication of  $\alpha$ -globin genes is a rare cause of anaemia that may lead to imbalances of  $\alpha$ - and  $\beta$ -chains in haemoglobin tetramer, especially in  $\beta$ -thalassemia carriers [32, 33]. Alternatively, the 16p13.3 chromosomal rearrangement may result in disruption of the long-range regulation leading to a partial silencing of genes expression consistent with prenatal and postnatal anaemia observed in patient 2 [34]. The other somatic anomalies seen in this patient may be a result of gene dosage of one or more of the other genes located in the duplicated region (Table S1). Indeed, a patient with 46,XY gonadal dysgenesis, skeletal and cardiac anomalies and developmental delay was reported to carry a similar 560 kb duplication located approximately 18 kb upstream of *SOX8* [35]. These rearrangements at the *SOX8* locus may cause dysregulation of *SOX8* expression. This hypothesis is consistent with the multiple reports of 46,XY gonadal dysgenesis associated with chromosomal rearrangements even up to several Mb upstream of genes involved in sex-determination including *SOX9* and *NR0B1* [36-39]. Furthermore, the

centromeric breakpoint in patient 2 falls between the conserved E1 and E2 *SOX8* enhancer elements, which are required for murine *Sox8* gene expression [26]. Dysregulation of *SOX8* expression could negatively impact on male gonadal development and lead to various degrees of 46,XY DSD. Indeed, 46,XY individuals carrying deletions that include *SOX8* (ATR16 syndrome) occasionally present with mild anomalies such as hypospadias or cryptorchidism and Patients 1 and 2 may represent part of this broad DSD spectrum [40].

The third patient presenting with 46,XY gonadal dysgenesis had a missense mutation involving a highly conserved glutamic acid residue within helix 3 of the *SOX8* HMG-box. The severe gonadal phenotype in this patient, which is due to lack of appropriate testis-determination, may be due to the specific dominant negative activity of the SOX8p.Glu156Asp protein on the WT-SOX8 protein and/or the repressive effect of this mutant protein on SOX9/NR5A1 synergy at the *Sox9 Tesco* enhancer that we observed using *in-vitro* assays (Fig. 7). These data, which provide evidence for a role of *SOX8* in early human testis development are supported by murine studies. The absence of both *Sox9* (pro-testis) and *Rspo1* (pro-ovary) in XX fetuses results in the development of ovotestes and hypoplastic testis [4]. *SoxE* group genes, *Sox10* and *Sox8* that are normally repressed by *Rspo1*, are activated in the double knockout gonads, suggesting that *Sox8* may serve as a driver of testis formation in the absence of *Sox9* [4]. Recent data suggest that *Sox8* may also be involved in repressing the ovarian pathway, via the repression of *Foxl2* expression during early testis formation [41].

In this study we have also identified mutations in *SOX8* associated with cases of male and female infertility. Although the mechanism(s) leading to infertility in either sex is unclear, all mutations associated with infertility showed differences in biological activity compared to the

WT protein and all mutations flank the HMG-box. Transient transfection assays using as a target the promoters of genes important for gonad function showed a wide range of effects including promoter specific loss-of-function or alterations in the cellular localization of the mutated protein. Inferring precisely how these mutations cause infertility is difficult since the gene regulatory pathways downstream of SOX8 have not been defined in either the XY or XX gonad. In mice *Sox8* is essential for the maintenance of male fertility beyond the first wave of spermatogenesis [10]. At 5 months of age *Sox8*<sup>-/-</sup> mice show a progressive degeneration of the seminiferous epithelium through perturbed physical interactions between Sertoli cells and the developing germ cells [10]. Here, we identified rare or novel *SOX8* mutations in 3.5% ( $p < 0.05$ ) of men with unexplained reduced sperm counts as compared to a frequency of 0.78% in ancestry-matched normospermic/fertile control cohorts. It is important to compare our data with a cohort of individuals who are known to be fertile and/or normospermic since infertility is a common phenotype and public databases, such as ExAC, are likely to contain a background pool of rare variants that can cause infertility and/or DSD. These cases of male infertility may represent mild forms of testicular dysgenesis or, similar to the mouse model the SOX8 protein may also be a regulator of Sertoli–germ cell adhesion independently of its role in primary testis-determination [10].

Our analyses also revealed mutations in *SOX8* in 5.06% of all women with POI from two replication studies ( $p = 4.5 \times 10^{-5}$ ). This suggests that mutations in human *SOX8* may have a greater impact on ovarian rather than testis function. Although *Sox8*<sup>-/-</sup> female mice are fertile, *Sox8* expression has been reported in preantral follicles, preovulatory follicles, cumulus granulosa cells and at high levels in mural granulosa cells, which line the wall of the follicle and are critical for steroidogenesis and ovulation [7, 42]. We have shown that SOX8 protein is expressed specifically in the granulosa cell of the developing and adult ovary in the human.

The observation that *Sox8* expression is higher in the mural cells that line the follicle wall than in the cumulus cells that surround the oocyte suggests that SOX8 may play a role in granulosa cell differentiation [42]. We can hypothesize that in female, in the absence of SOX9, SOX8 may be an important regulator of AMH expression (required for maintenance of the germ cell pool) in the adult ovary and, therefore, mutations in *SOX8* may result in POI.

The results presented in this study provide novel insight into the genetic mechanisms of human gonadal development and function and provide further evidence that a spectrum of reproductive phenotypes from DSD to infertility can be associated with variations in a single genetic factor.

## **Materials and Methods**

### **Detailed study methods are provided in SI Methods**

**Subjects and Samples.** All patients with 46,XY DSD met the revised criteria of the Pediatric Endocrine Society (LWPES)/European Society for Paediatric Endocrinology (ESPE). This study was approved by the local French ethical committee (2014/18NICB – registration number IRB00003835) and consent to genetic testing was obtained from adult probands or from the parents when the patient was under 18 years. Patient ancestry was determined by self reporting, based on responses to a personal questionnaire, which asked questions pertaining to the birthplace, languages and self-reported ethnicity of the participants, their parents and grandparents. Genes known to be involved in 46,XY DSD were screened for mutations in the XY DSD cohort and high resolution aCGH was performed on all cases and indicated normal ploidy in all cases. An extended description of patient 2 is provided in SI Methods. The infertile male population consisted of 274 men of European or North African ancestry. For each man ejaculates were obtained by masturbation after 2–7 days of sexual



abstinence. All underwent an andrological work-up, which included medical history, physical examination, hormonal evaluation (FSH, LH, and testosterone) and semen analysis. Men with known clinical (cryptorchidism, infections, varicocele) or genetic (karyotype anomalies, Y chromosome microdeletions) causes of infertility were excluded. Oligozoospermia was defined as having less than  $15 \times 10^6$  sperm/ml. Primary amenorrhea was diagnosed when pubertal development was absent despite the patient being of pubertal age (greater than 13 years) with increased basal plasma FSH concentration ( $>9$  IU/L). Cases of secondary amenorrhea (no menstruation after 6 cycles) and premature ovarian failure (amenorrhea, hypoestrogenism, and elevated serum gonadotropin levels in women younger than 40 years of age) were included.

The control panel consisted of 280 unrelated normospermic 46,XY males of French ethnic origin with known fertility and no history of testicular anomalies (determined by self-reporting) from the Biobank for Research on Human Reproduction (GERMETHEQUE). An additional in-house 180 normospermic men and 130 fertile men of Arab/North African ancestry were included as controls. European ancestry-matched in-house controls of known fertility status were also included ( $n=20$ ). Other control populations included fathers of European ancestry from the 1000 genomes project (Iberian and Northern Europeans from Utah,  $n=103$ ), the Danish Genome Project ( $n=50$ ) and from the University of Dundee ( $n=50$ ).

**Genomic analysis.** Chromosome analysis, ArrayCGH, genomic, exome and Sanger sequencing are all described in SI Materials and Methods.

**Immunohistochemistry.** Human fetal testis tissue (9 weeks post conception) was provided with approval from the Human Developmental Biology Resource (HDBR, [www.hdbr.org](http://www.hdbr.org)).

Term fetal (40 weeks) ovary was obtained from Abcam (#ab4412) and adult ovary tissue (19 years) was obtained from Tissue Solutions. Tissue sections (12  $\mu$ m) were fixed briefly in 4% PFA in TBS, rinsed in TBS and blocked in 1% BSA in TBS-Tween (0.5% Tween) for 1 hour before incubating overnight with rabbit monoclonal SOX8 antibody (Sigma HPA41640; 1:200 dilution) and, where relevant, mouse monoclonal NR5A1 (SF1) antibody (Thermo Fisher #434200; 1:200 dilution). Sections were washed in TBS-Tween and incubated for 1 hour with the relevant secondary antibodies: Alexa555 goat anti-rabbit (Invitrogen, A21429; 1:400) and Alexa488 goat anti-mouse (Invitrogen, A11001; 1:400). Nuclei were counterstained with DAPI (Sigma). Slides were washed and mounted using ProLong Gold Antifade Mountant (Lifetech, #P36930). Images were collected on a Zeiss LSM 710 confocal microscope (Carl Zeiss) and analyzed using Zeiss Zen 2009 and Image J.

**Details of the plasmids, cells lines, cellular localization assays, structural modelling and co-immunoprecipitation** are provided in the SI Methods

**Statistics.** The results of the luciferase assays were compared for their statistical significance by calculating two-tailed Student's t Test using GraphPad Prism software. 95% confidence interval was calculated using the mean difference between the two groups being tested. The data for each of the groups tested and their t values, along with degrees of freedom (df), standard error of difference, P-values and predicted statistical significance are summarized in S2 Table. Based on the P-values the GraphPad Prism software calculates the difference as not significant, not likely to be significant, significant, very significant and extremely significant.

## **Acknowledgements**

A.B is funded in part by the program Actions Concertees Interpasteuriennes (ACIP) and a research grant from the European Society of Pediatric Endocrinology. A. B and K. McE are funded by a research grant from the EuroDSD in the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement n° 201444 as well as grant N°295097 as part of the EU call FP7-INCO-2011-6. The work is also funded by a Franco-Egyptian AIRD-STDF grant (A.B., K.M., I.M.). Y.S. thanks the Chinese Government Scholarship and University of the Chinese Academy of Science (UCAS) for financial and infrastructure support. R.J. is supported by a 2013 MOST China-EU Science and Technology Cooperation Program, Grant No. 2013DFE33080, by the National Natural Science Foundation of China (Grant No. 31471238) and a 100 talent award of the Chinese Academy of Sciences. K.A. is supported by the Innovation Fund Denmark (grant # 14-2013-4). The human embryonic and fetal material was provided by the Joint MRC/Wellcome Trust (grant # 099175/Z/12/Z) Human Developmental Biology Resource ([www.hdbr.org](http://www.hdbr.org)). J.C.A. is a Wellcome Trust Senior Research Fellow in Clinical Science (098513/Z/12/Z) and received support from the National Institute for Health Research Biomedical Research Centre at Great Ormond Street Hospital for Children NHS Foundation Trust and University College London. R.R. is a fellow supported by the Italian Ministry of Health, Rome, Italy (grant # GR-2011-02351636). This work is supported by the COST Action DSDnet BM1303.

## **Conflict of Interest Statement**

The authors declare no competing financial interests

## **Supplemental Data**

Supplemental Data include 5 figures and 3 tables and can be found with this article online.

## References

1. Sekido R, Lovell-Badge R. (2008) Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. *Nature*. 453, 930-934.
2. Lin YT, Capel B. (2015) Cell fate commitment during mammalian sex determination. *Curr. Opin. Genet. Dev.* 32, 144-152.
3. Chaboissier MC, Kobayashi A, Vidal VI, Lützkendorf S, van de Kant HJ, Wegner M, de Rooij DG, Behringer RR, Schedl A. (2004) Functional analysis of Sox8 and Sox9 during sex determination in the mouse. *Development.*, 131, 1891-1901.
4. Lavery R, Chassot AA, Pauper E, Gregoire EP, Klopfenstein M, de Rooij DG, Mark M, Schedl A, Ghyselinck NB, Chaboissier MC. (2012) Testicular differentiation occurs in the absence of R-spondin1 and Sox9 in mouse sex reversals. *PLoS Genet.*, 8, e1003170.
5. Barrionuevo F, Georg I, Scherthan H, Lécureuil C, Guillou F, Wegner M, Scherer G. (2009) Testis cord differentiation after the sex determination stage is independent of Sox9 but fails in the combined absence of Sox9 and Sox8. *Dev. Biol.*, 327, 301-312.
6. Nicol B, Yao HH. (2015) Gonadal Identity in the Absence of Pro-Testis Factor SOX9 and Pro-Ovary Factor Beta-Catenin in Mice. *Biol. Reprod.* 93, 35.
7. O'Bryan MK, Takada S, Kennedy CL, Scott G, Harada S, Ray MK, Dai Q, Wilhelm D, de Kretser DM et al. (2008) Sox8 is a critical regulator of adult Sertoli cell function and male fertility. *Dev. Biol.*, 316, 359-370.

8. Lee PA, Nordenström A, Houk CP, Ahmed SF, Auchus R, Baratz A, Baratz Dalke K, Liao LM, Lin-Su K, Looijenga LH et al. (2016) Global Disorders of Sex Development Update since 2006: Perceptions, Approach and Care. *Horm. Res. Paediatr.*, 85, 158-180.
9. Ahmed SF, Bashamboo A, Lucas-Herald A, McElreavey K. (2013) Understanding the genetic aetiology in patients with XY DSD. *Br. Med. Bull.*, 106, 67-89.
10. Bashamboo A, McElreavey K. (2013) Gene mutations associated with anomalies of human gonad formation. *Sex. Dev.*, 7, 126-146.
11. Suntharalingham JP, Buonocore F, Duncan AJ, Achermann JC. (2015) DAX-1 (NR0B1) and steroidogenic factor-1 (SF-1, NR5A1) in human disease. *Best Pract. Res. Clin. Endocrinol. Metab.*, 29, 607-619.
12. Bashamboo A, Ferraz-de-Souza B, Lourenço D, Lin L, Sebire NJ, Montjean D, Bignon-Topalovic J, Mandelbaum J, Siffroi JP, Christin-Maitre S et al. (2010) Human male infertility associated with mutations in NR5A1 encoding steroidogenic factor 1. *Am. J. Hum. Genet.*, 87, 505-512.
13. Lourenço D, Brauner R, Lin L, De Perdigo A, Weryha G, Muresan M, Boudjenah R, Guerra-Junior G, Maciel-Guerra AT, Achermann JC et al. (2009) Mutations in NR5A1 associated with ovarian insufficiency. *N. Engl. J. Med.*, 360, 1200-1210.

14. Bashamboo A, Bashamboo A, Donohoue PA, Vilain E, Rojo S, Calvel P, Seneviratne SN, Buonocore F, Barseghyan H, Bingham N, Rosenfeld JA et al. (2016) A recurrent p.Arg92Trp variant in steroidogenic factor-1 (NR5A1) can act as a molecular switch in human sex development. *Hum. Mol. Genet.*, 25, 3446-3453.
15. Skakkebaek NE, Rajpert-De Meyts E, Buck Louis GM, Toppari J, Andersson AM, Eisenberg ML, Jensen TK, Jørgensen N, Swan SH, Sapra KJ et al. (2016) Male Reproductive Disorders and Fertility Trends: Influences of Environment and Genetic Susceptibility. *Physiol. Rev.*, 96, 55-97.
16. Ray A, Shah A, Gudi A, Homburg R. (2012) Unexplained infertility: an update and review of practice. *Reprod. Biomed. Online.*, 24, 591-602.
17. Nelson LM. (2009) Clinical practice. Primary ovarian insufficiency. *N. Engl. J. Med.*, 360, 606-614.
18. Cooper TG, Noonan E, von Eckardstein S, Auger J, Baker HW, Behre HM, Haugen TB, Kruger T, Wang C, Mbizvo MT et al. (2010) World Health Organization reference values for human semen characteristics. *Hum. Reprod. Update*, 16, 231-245.
19. Jamsai D, O'Bryan MK. (2011) Mouse models in male fertility research. *Asian J. Androl.*, 13, 139-151.
20. Matzuk MM, Lamb DJ. (2008) The biology of infertility: research advances and clinical

challenges. *Nat. Med.*, 14, 1197-1213.

21. Ben Khelifa M, Coutton C, Zouari R, Karaouzène T, Rendu J, Bidart M, Yassine S, Pierre V, Delaroche J, Hennebicq S, et al. (2014) Mutations in DNAH1, which encodes an inner arm heavy chain dynein, lead to male infertility from multiple morphological abnormalities of the sperm flagella. *Am. J. Hum. Genet.*, 94, 95-104.

22. Wang J, Zhang W, Jiang H, Wu BL. (2014) Primary Ovarian Insufficiency Collaboration. Mutations in HFM1 in recessive primary ovarian insufficiency. *N. Engl. J. Med.*, 370, 972-974.

23. Caburet S, Arboleda VA, Llano E, Overbeek PA, Barbero JL, Oka K, Harrison W, Vaiman D, Ben-Neriah Z, García-Tuñón I, et al. (2014) Mutant cohesin in premature ovarian failure. *N. Engl. J. Med.*, 370, 943-949.

24. Yatsenko AN, Georgiadis AP, Röpke A, Berman AJ, Jaffe T, Olszewska M, Westernströer B, Sanfilippo J, Kurpisz M, Rajkovic A, et al. (2015) X-linked TEX11 mutations, meiotic arrest, and azoospermia in infertile men. *N. Engl. J. Med.*, 372, 2097-2107.

25. Péterfy M, Ben-Zeev O, Mao HZ, Weissglas-Volkov D, Aouizerat BE, Pullinger CR, Frost PH, Kane JP, Malloy MJ, Reue K et al. (2007) Mutations in LMF1 cause combined lipase deficiency and severe hypertriglyceridemia. *Nat. Genet.*, 39, 1483-1487.

26. Guth SI, Bösl MR, Sock E, Wegner M. (2010) Evolutionary conserved sequence elements with embryonic enhancer activity in the vicinity of the mammalian Sox8 gene. *Int. J. Biochem. Cell. Biol.*, 42, 465-471.
27. Neirijnck Y, Borel C, Escoffier J, Smith LB, Antonarakis SE, Dermitzakis ET, Nef S (2017) Deciphering cell lineage specification during male sex determination with single-cell RNA sequencing. *bioRxiv* 190264; doi: <https://doi.org/10.1101/190264>
28. Ng CK, Li NX, Chee S, Prabhakar S, Kolatkar PR, Jauch R. (2012) Deciphering the Sox-Oct partner code by quantitative cooperativity measurements. *Nucl. Acids Res.*, 40, 4933-4941.
29. Schepers G, Wilson M, Wilhelm D, Koopman P. (2003) SOX8 is expressed during testis differentiation in mice and synergizes with SF1 to activate the Amh promoter in vitro. *J. Biol. Chem.*, 278, 28101-28108.
30. Huang YH, Jankowski A, Cheah KS, Prabhakar S, Jauch R. (2015) SOXE transcription factors form selective dimers on non-compact DNA motifs through multifaceted interactions between dimerization and high-mobility group domains. *Sci. Rep.*, 5, 10398.
31. Stolt CC, Lommes P, Friedrich RP, Wegner M. (2004) Transcription factors Sox8 and Sox10 perform non-equivalent roles during oligodendrocyte development despite functional redundancy. *Development.*, 131, 2349-2358.



32. Origa R, Moi P. Alpha-Thalassemia. 2005 Nov 1 [updated 2016 Dec 29]. In: Pagon RA et al., Editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2017.
33. Liu S, Jiang H, Wu MY, Zhang YL, Li DZ. (2015) Thalassemia Intermedia Caused by 16p13.3 Sectional Duplication in a  $\beta$ -Thalassemia Heterozygous Child. *Pediatr. Hematol. Oncol.*, 32, 349-353.
34. Hughes JR, Lower KM, Dunham I, Taylor S, De Gobbi M, Sloane-Stanley JA, McGowan S, Ragoussis J, Vernimmen D, Gibbons RJ et al. (2013) High-resolution analysis of cis-acting regulatory networks at the  $\alpha$ -globin locus. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 368, 20120361.
35. Erickson RP, Yatsenko SA, Larson K, Cheung SW. (2011) A Case of Agonadism, Skeletal Malformations, Bicuspid Aortic Valve, and Delayed Development with a 16p13.3 Duplication Including GNG13 and SOX8 Upstream Enhancers: Are Either, Both or Neither Involved in the Phenotype? *Mol. Syndromol.*, 1, 185-191.
36. Bhagavath B, Layman LC, Ullmann R, Shen Y, Ha K, Rehman K, Looney S, McDonough PG, Kim HG, Carr BR. (2014) Familial 46,XY sex reversal without campomelic dysplasia caused by a deletion upstream of the SOX9 gene. *Mol. Cell. Endocrinol.*, 393, 1-7.
37. Sutton E, Hughes J, White S, Sekido R, Tan J, Arboleda V, Rogers N, Knower K, Rowley L, Eyre H et al. (2011) Identification of SOX3 as an XX male sex reversal gene in mice and humans. *J. Clin. Invest.*, 121, 328-341.

38. Jakubiczka S, Schröder C, Ullmann R, Volleth M, Ledig S, Gilberg E, Kroisel P, Wieacker P. (2010) Translocation and deletion around SOX9 in a patient with acampomelic campomelic dysplasia and sex reversal. *Sex. Dev.*, 4, 143-149.
39. Smyk M, Berg JS, Pursley A, Curtis FK, Fernandez BA, Bien-Willner GA, Lupski JR, Cheung SW, Stankiewicz P. (2007) Male-to-female sex reversal associated with an approximately 250 kb deletion upstream of NR0B1 (DAX1). *Hum. Genet.*, 122, 63-70.
40. Hartevelde CL, Kriek M, Bijlsma EK, Erjavec Z, Balak D, Phylipsen M, Voskamp A, di Capua E, White SJ, Giordano PC. (2007) Refinement of the genetic cause of ATR-16. *Hum. Genet.*, 122, 283-292.
41. Gonen N, Quinn A, O'Neill HC, Koopman P, Lovell-Badge R. (2017) Normal Levels of Sox9 Expression in the Developing Mouse Testis Depend on the TES/TESCO Enhancer, but This Does Not Act Alone. *PLoS Genet.*, 13, e1006520.
42. Salmon NA, Handyside AH, Joyce IM. (2005) Expression of Sox8, Sf1, Gata4, Wt1, Dax1, and Fog2 in the mouse ovarian follicle: implications for the regulation of Amh expression. *Mol. Reprod. Dev.*, 70, 271-277.

## Figure Legends

### Figure 1 Chromosomal rearrangements at the *SOX8* locus associated with 46,XY DSD.

(A) Schematic representation of the 16p inversion in patient 1 indicating the genes and clones of the region and the position of the telomeric breakpoint (TB) and centromeric breakpoints (CB). The position of the transcription initiation site of the *SOX8* gene is indicated (chr16:1031808), the *Sox8* enhancer elements (E1-E4) and the position of the BACs used for FISH analysis. Patient 1 with (B) GTG banding showing the rearranged chromosome 16 (arrow). (C) FISH on metaphases with BAC RP11-161M6 (green; chr16:991,108-1,141,991-UCSC Genome Browser; GRCh37/hg19) and RP11-297M9 (red; chr16:9,727,655-9,920,250) confirming the paracentric inversion (arrow). (D) The CB is located within the clone, RP11-609N14 on 16p13.13. (E) The TB mapped within RP11-728H8, on 16p13.3 (arrows). Patient 2 (F) Chromosome 16p rearrangement. The aCGH profile indicated an 854 kb triplication in the 16p13.3 region. Affected region is indicated by a blue bar. Red rectangles outline the triplication breakpoints. A magnified view of the triplication breakpoints shows, on the left, a CB located in the intron 1 of the *NPRL3* gene. The 16p13.3 triplication starts at chr16:137893 position and spans proximally including the  $\alpha$ -globin genes (*HBZ*, *HBM*, *HBA1*, *HBA2*, and *HBQ1*) and their upstream cis-regulatory elements R1-R4 (red bars). On the right, the TB of the triplication (chr16:992,302) is located ~39.5 kb upstream of the *SOX8* gene. The *SOX8*-specific enhancers are mapped within the triplicated segment. Genes are indicated by black arrows with vertical lines specifying exons. Arrows correspond to a direction of gene transcription. (G) Interphase FISH analysis using RP11-598I20 (16p13.3, labeled red) and a control RP11-121O8 (16p13.1, labeled green) probe showing a triplication (white arrow) in the patient 20. (H) Representative FISH image using the same probes shows a normal hybridization pattern observed in both parents

**Figure 2. Gonad histology of Patients 1 and 2.** Patient 1, (A) Ovarian-like stromal cells with no evidence of testicular tissue consistent with 46,XY complete gonadal dysgenesis. Patient 2, (B) Seminiferous tubules (cords) are mildly contorted in places (black arrow) and some are misshapen. The interstitium space is expanded and mildly fibrotic (white arrow; H&E; 200X). (C) Immunohistochemistry for calretinin shows only focal Leydig cells (arrows; calretinin stain, blue; 200X). (D) Inhibin immunohistochemistry highlights Sertoli cells, making the contorted tubules more evident (inhibin; 200X). (E) Immunohistochemistry for SOX9 staining the nuclei of Sertoli cells within tortuous tubules (SOX9; 200X).

**Figure 3. Expression of the SOX8 protein in human gonad tissues.** (A) The SOX8 protein (red) is expressed in the human male Sertoli and Leydig cells together with NR5A1 (green) during early testis formation (image at 9 weeks post conception). Primitive seminiferous cords are indicated by the dashed lines (B) The SOX8 protein expression in the granulosa cells of the late fetal ovary (image at 40 weeks of gestation). (C) Immunohistochemistry showing SOX8 expression in granulosa cells (arrow) lining the follicles of the adult ovary (19 yrs old; scale bars are 50  $\mu$ m).

**Figure 4. Mutation in the HMG-box of SOX8 associated with 46,XY gonadal dysgenesis.** (A) Schematic representation showing important functional domains of the SOX8 protein. The amino acid sequence of the HMG domain including the three alpha-helices, the two nuclear localization signals (NLS1, NLS2), and the nuclear export sequence (NES) together with the position of SOX8p.Glu156Asp (p.E156D) mutation are indicated. The DNA-dependent dimerization domain (DIM), the DNA-binding HMG domain, and the transactivation domains (TA1, TA2) are shown. (B) Left, alignment of the C-terminal of the SOX8 HMG-box domain containing helix 3 from selected vertebrates. Right, the alignment of

the distal portion of the SOX8 HMG-box domain with other human SOXE and SOXF family members. The position of the p.Glu156Asp missense mutation is highlighted. (C) Structural models for ternary complexes of conjectured SOX8p.Glu156Asp -SOX8, SOX8p.Glu156Asp -SOX9 and SOX8p.Glu156Asp -OCT4 dimers on composite DNA elements. SOX8 is shown in dark red, SOX9 in orange and OCT4 in blue. The HMG domains of SOX E proteins and the POU domain of OCT4 are shown as cartoon with cylindrical alpha-helices and the DNA as gray van-der-Waals surface. The SOX8p.Glu156Asp site is drawn as ball-and-sticks. The DNA sequences used to generate the models with color-coded binding elements are depicted underneath the models. The SOXE HMG box is N-terminally extended by a 40 amino acid dimerization (DIM) domain of unknown structure that mediates highly cooperative DNA dependent dimerization presumably by interactions with the HMG box.

**Figure 5. SOX8p.Glu156Asp shows altered biological activity.** (A) The transcriptional activities of WT-SOX8 and SOX8p.Glu156Asp were studied using the human *AMH* and *NR5A1* promoters, mouse *Dmrt1* promoter and the *Sox9 Tesco* enhancer as reporters, following transfection in HEK293-T cells. The data shown represent the mean  $\pm$  SEM of minimum three independent experiments, each of which was performed at least in quadruplicate. The reporter constructs were transfected into HEK293-T cells with either the WT-SOX8 or SOX8p.Glu156Asp expression vector. The results are expressed as relative percentage of WT-SOX8 activity (100%). The SOX8p.Glu156Asp can activate the gonadal promoters in a similar fashion to WT-SOX8. However, although SOX8p.Glu156Asp can synergistically activate the *AMH* promoter with NR5A1, it fails to synergise with NR5A1 to activate the *Sox9 Tesco* enhancer. (B) HEK 293-T cells were co-transfected with *Tesco* reporter construct (10ng), WT-NR5A1 (1ng) and WT-SOX9 (1ng) and increasing amount of WT-SOX8 or SOX8p.Glu156Asp (0,1,2,5,10 ng). Results are expressed as Relative

Luminescence Units (RLU). The SOX8p.Glu156Asp mutant exerts a repressive effect by preventing synergistic activation of the *Sox9 Tesco* enhancer by NR5A1 and SOX9 even at three times lower concentrations.

**Figure 6. Protein-Protein interaction assayed using Proximity ligation assay.** Plasmids encoding SOX8-WT or SOX8p.Glu156Asp were transiently expressed with WT-NR5A1 or SOX9 for 48hrs in HEK293T cells. Protein-protein interaction of SOX8-WT and SOX8p.Glu156Asp with NR5A1 and SOX9 were analysed using the Duolink proximity ligation assay. Nuclei are stained with DAPI (blue) and Duolink signal representing interaction between the proteins is shown in green. Each green dot represents a single dimerization event. Both the SOX8-WT and SOX8p.Glu156Asp proteins can physically interact with NR5A1. Whereas, although SOX8-WT protein can interact with SOX9, the SOX8p.Glu156Asp cannot.

**Figure 7. Schematic representation of the results of the in vitro assays for the SOX8p.Glu156Asp protein.** (A) The mutant SOX8 protein has the ability to physically interact with the NR5A1 protein but in contrast to the WT-SOX8 protein does not synergize with NR5A1 to promote reporter gene activity using the *Tesco* enhancer as a target. (B) The mutant SOX8 protein does not physically interact with the SOX9 protein and shows a lack of synergy with SOX9/NR5A1 to promote reporter gene activity. (C) The SOX8 mutant protein also shows dominant negative activity by impairing promoter gene activity of the WT-SOX8 protein in synergy with NR5A1.



**Table 1. Phenotypes, genetics and clinical investigations of 46,XY DSD with mutations involving the *SOX8* locus (patients 1 and 2) or gene (patient 3).**

Patient	Phenotype	Clinical presentation (age)	Genetics	Hormonal profile (normal range)	Genitalia and gonads	Diagnosis
1	Female	Primary amenorrhea with female external genitalia	46,XY,inv(16)(p13.3p13.1)	FSH: 59.5 iU/l (2.0–15.0) LH: 23.6 iU/l (2.1–14.9) T: 0.7 ng/ml (2.70–9.00)	Bilateral streak gonads with fallopian tubes. Stromal-like tissue with no evidence for testicular tissue or follicles. Small immature uterus, 43x22x30 mm (27 yrs).	46,XY CGD
2	Female at birth then changed to male assignment	Ambiguous genitalia (birth), mild anemia, skeletal anomalies, mild developmental delay	46,XY arr[h19] 16p13.3(137,893-992,302)x3	FSH: 2.31 iU/l (0.16-4.1 mIU/ml) LH: 17.35 iU/l (0.02-7.0 mIU/ml) T: 1.38 ng/ml (0.6-4 ng/ml) DHT 36 ng/dL (12-85)	At birth clitorophallic structure (1x1cm) with urethral meatus at base. Fused labioscrotal folds with palpable gonads (R: 1.2x0.6x0.8cm, L: 1.1x0.4x0.7cm). No uterus or vagina. Testis histology showed seminiferous tubules, variable germ cells and reduced Leydig cell numbers.	46,XY severe undervirilization
3	Female	Primary amenorrhea (16 yrs)	46,XY <i>SOX8</i> p.Glu156Asp PolyPhen2 Score 0.999	Elevated FSH, LH: undetectable T levels	Bilateral streak gonads with absent uterus. Epididymis-like structure present.	46,XY PGD

CGD: complete gonadal dysgenesis; PGD: partial gonadal dysgenesis; FSH: Follicle stimulating hormone; LH: Luteinizing hormone. All mutations are heterozygous.



**Table 2. *SOX8* mutations in 46,XY DSD, male infertility, primary ovarian insufficiency and in the control cohort of ancestry-matched fertile/normospermic individuals**

46,XY DSD (n=204)	Azoospermia (n=131)	Oligozoospermia (n=143)	POI (n=153)	POI replication study (n=104)	Fertile and/or normospermic controls (n=813; European ancestry 503; North African/Arab ancestry 310)
p.Glu156Asp	p.Lys241Thr	p.Gly359Arg (x2)	p.Arg8_Ser9del deletion of 2 AA	p.Pro11Leu	p.Ala240Thr <sup>1</sup>
p.Pro336Leu	p.Asp382Asn	p.Ala373Ser	p.Ala32Val	p.Pro196Leu	p.Thr331Met <sup>2</sup>
p.Thr340Ser		p.Gly378Ser	p.Pro242Leu (x2, sisters)	p.Pro242Leu	p.Pro336Leu <sup>3</sup>
		p.Asp382Asn	p.Ser267Leu	p.Ala416Thr	p.Thr340Ser <sup>3</sup>
			p.Val282Ile		p.Ala373Thr <sup>1</sup>
			p.Asp382Asn (x3)		p.Ala428Thr <sup>3</sup>
			p.Ala416Thr		
<b>1.47%</b>	<b>1.53%</b>	<b>3.5%<sup>a</sup></b>	<b>5.9%<sup>b</sup></b>	<b>3.85%<sup>c</sup></b>	<b>0.74%</b>

<sup>a</sup>p=0.0151, <sup>b</sup>p=0.000107, <sup>c</sup>p=0.0191 compared to controls (Fishers t-test, 2 tailed). POI: Primary ovarian insufficiency; DSD: Disorder of Sex Development; AA: Amino acids. All mutations are heterozygous. Fertile fathers from <sup>1</sup>1000 genomes project (European ancestry), <sup>2</sup>Danish Genome Project and <sup>3</sup>in-house controls.

**Table 3. Mutations in the *SOX8* gene associated with human male or female infertility.** Heterozygous missense mutations were identified in the *AIRE* and *NR5A1* genes in cases 18 and 20 respectively. In such cases the *SOX8* mutation may contribute to the severity of the phenotype. \*Novel variant. POI, primary ovarian insufficiency. All mutations are heterozygous.

Patient	Mutation	Variant identifier	Global minor allelic frequency	PolyPhen2 Score	Ancestry	Phenotype and comments
4	p.Lys241Thr	rs553836905	0.0002	0.762	European	Azoospermia
5	p.Gly359Arg	rs199719055	0.000014	0.961	Eastern European	Oligozoospermia $3.4 \times 10^6$ /ml
6	p.Gly359Arg	rs199719055	0.000014	0.961	French	Oligozoospermia $2 \times 10^6$ /ml
7	p.Ala373Ser	rs200828189	0.0036	0.000	French	Oligozoospermia $9 \times 10^6$ /ml
8	p.Gly378Ser	rs576173763	0.0064	0.004	Senegalese	Severe oligozoospermia $0.005 \times 10^6$ /ml
9	p.Asp382Asn	rs143203270	0.0057	0.986	Tunisian	Oligozoospermia. $6 \times 10^6$ /ml
10	p.Asp382Asn	rs143203270	0.0057	0.986	Algerian	Azoospermia
11	p.Arg8_Ser9del*	-	-	-	Italian	POI. Menarche at 13 yrs; one pregnancy; secondary amenorrhea and POI at 37 yrs. Mother had menopause at 35 yrs
12	p.Pro11Leu	rs200231150	0.010	0.636	European	POI. Primary amenorrhea
13	p.Ala32Val*	-	0.000009	0.919	Italian	POI. Menarche at 14 yrs; secondary amenorrhea and POI at 21 yrs
14	p.Pro196Leu	-	0.0000275	0.001	European	POI. Primary amenorrhea. Atrophic ovaries
15	p.Pro242Leu	rs567451602	0.0057	0.001	Paraguay	POI. Primary amenorrhea
16	p.Pro242Leu	rs567451602	0.0057	0.001	Paraguay	POI. Primary amenorrhea (sister of case 12)
17	p.Pro242Leu	rs567451602	0.0057	0.001	European	POI. Primary amenorrhea. Atrophic ovaries
18	p.Ser267Leu	rs368972027	0.000017	0.942	Italian	POI. Amenorrhea at 30. Hypoparathyroidism during childhood
19	p.Val282Ile*	-	-	0.990	Argentinian	POI. Secondary amenorrhea
20	p.Asp382Asn	rs143203270	0.0057	0.986	North African	POI. Menarche at 12 yrs; secondary amenorrhea at 14 yrs.
21	p.Asp382Asn	rs143203270	0.0057	0.986	European	POI. Primary amenorrhea at 16 yr. Absent pubertal development, ovaries not seen by ultrasound
22	p.Asp382Asn	rs143203270	0.0057	0.986	French	POI. Menarche at 11 yrs; secondary amenorrhea at 14 yrs
23	p.Ala416Thr	rs201602067	0.0004	0.005	Italian	POI. Menarche at 9 yrs; secondary amenorrhea and POI at 38 yrs
24	p.Ala416Thr	rs201602067	0.0004	0.005	European	POI. Primary amenorrhea



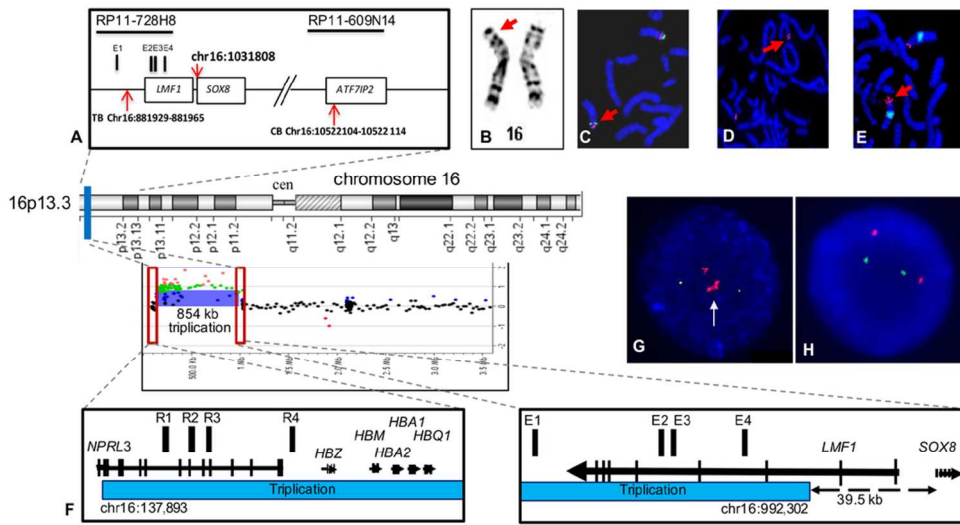


Figure 1

94x52mm (300 x 300 DPI)

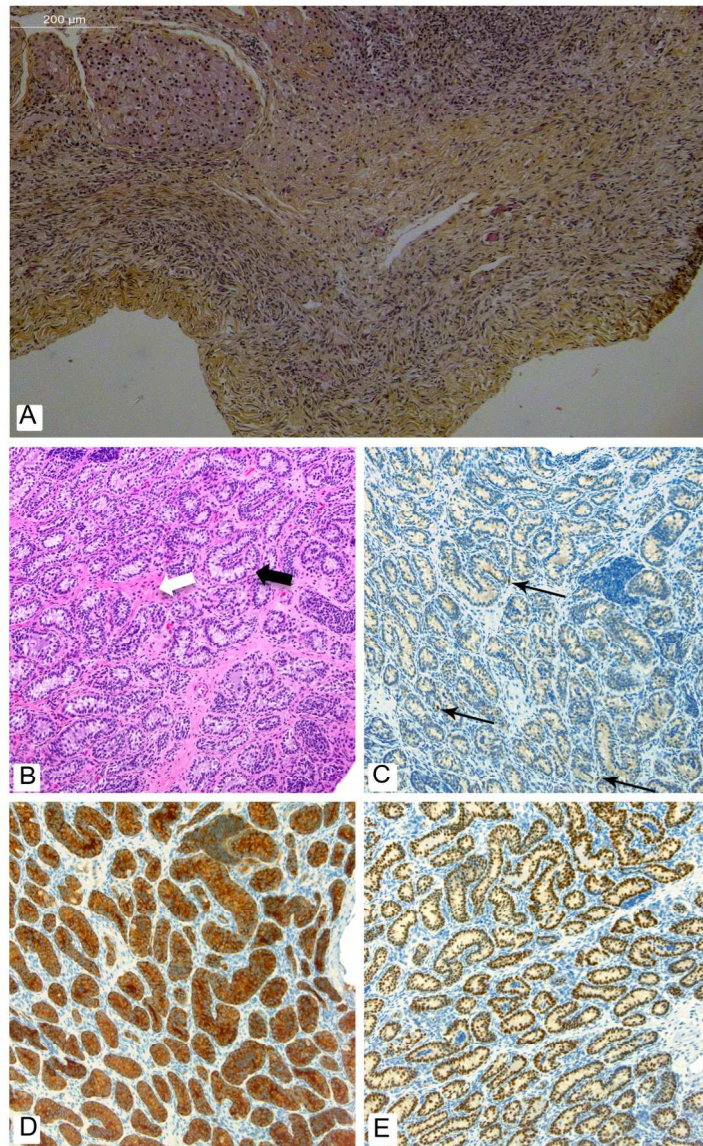


Figure 2

180x287mm (300 x 300 DPI)

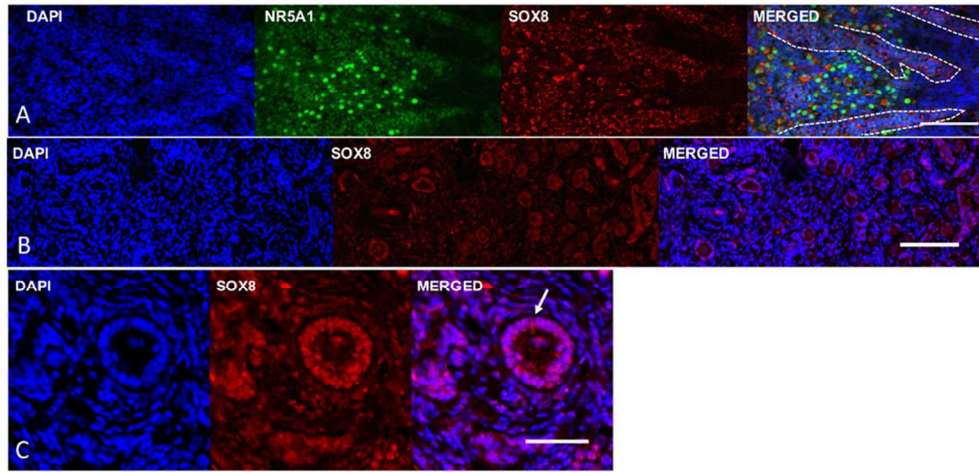


Figure 3

82x39mm (300 x 300 DPI)

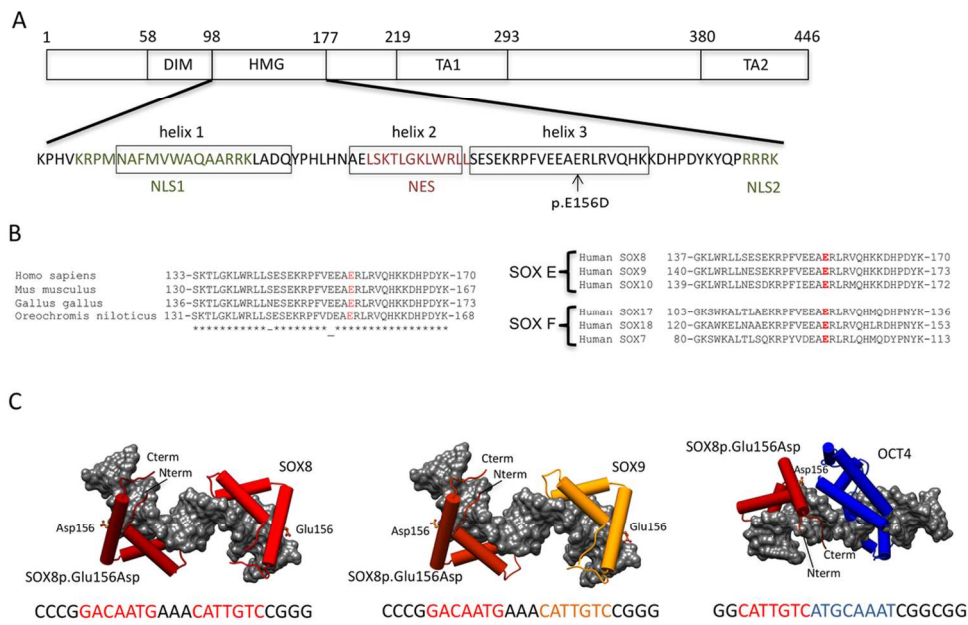


Figure 4

108x69mm (300 x 300 DPI)

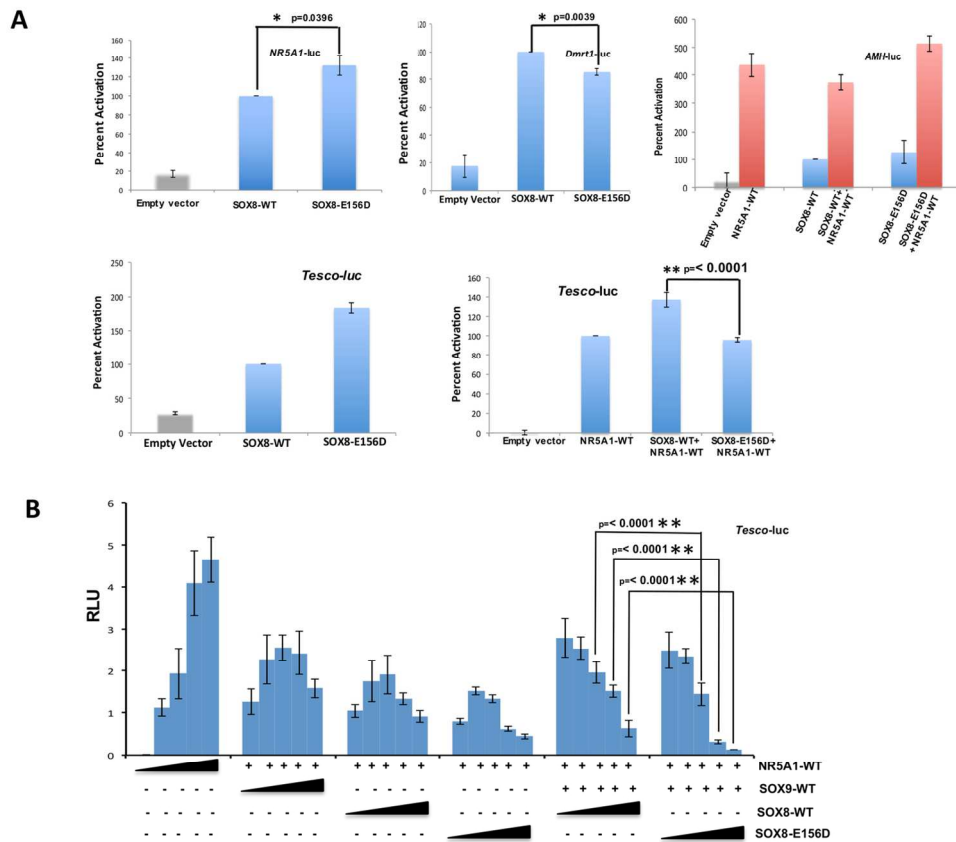


Figure 5

131x119mm (300 x 300 DPI)



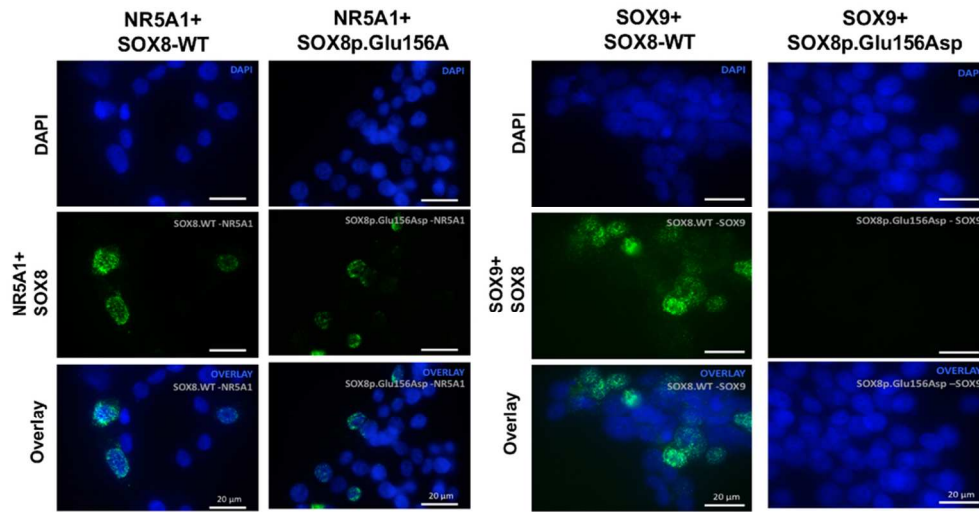


Figure 6

45x24mm (600 x 600 DPI)

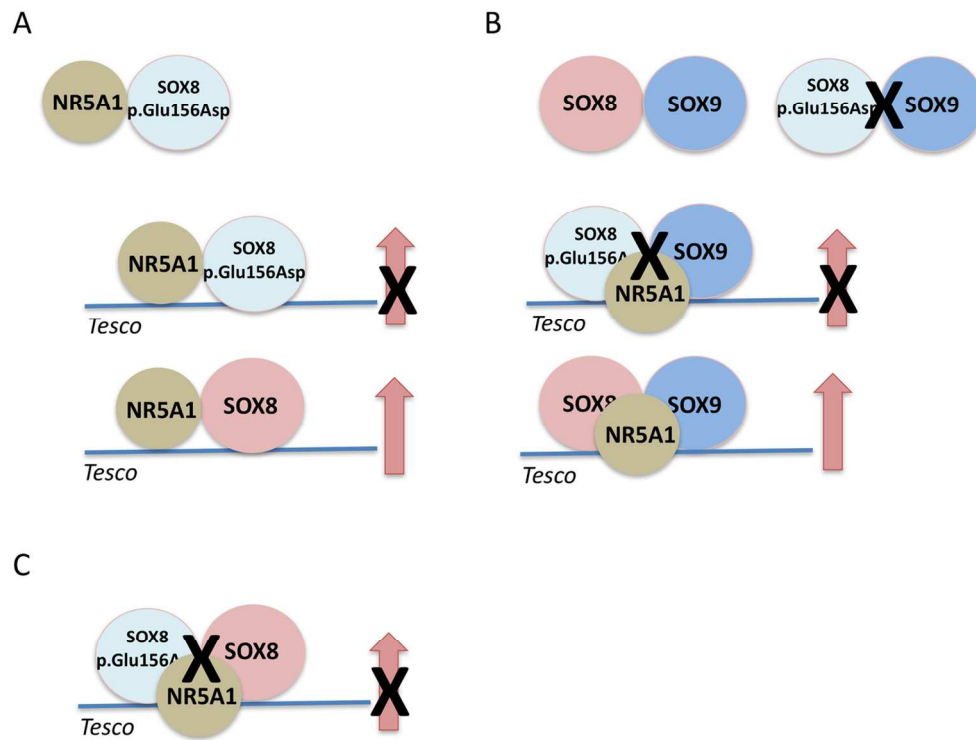


Figure 7

123x92mm (300 x 300 DPI)