Exploration génétique et moléculaire de défauts post-méiotiques sévères de la spermatogenèse entrainant une infertilité masculine

Zine-Eddine Kherraf

Communauté
UNIVERSITÉ Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : BIS - Biotechnologie, instrumentation, signal et imagerie pour la biologie, la médecine et l'environnement

Arrêté ministériel : 25 mai 2016

Présentée par

ZINE-EDDINE KHERRAF

Thèse dirigée par **Pierre RAY (EDISCE)**, UGA et CHU de Grenoble

préparée au sein de l'**Institut pour l'Avancée des Biosciences** dans **l'École Doctorale Ingénierie pour la santé la Cognition et l'Environnement**

Exploration génétique et moléculaire de défauts post-méiotiques sévères de la spermatogenèse entrainant une infertilité masculine.

Thèse soutenue publiquement le **12 juillet 2018**,
devant le jury composé de :

Monsieur Saadi KHOCHBIN
Directeur de Recherche, CNRS Délégation Alpes, Président

Madame Catherine PATRAT
Professeur des Universités - Praticien Hospitalier, Université Paris 5, Rapporteur

Monsieur Florian-Jean-Louis GUILLOU
Directeur de Recherche, INRA Centre Val de Loire, Rapporteur

Madame Sophie CHRISTIN-MAITRE
Professeur des Universités - praticien hospitalier, Université Pierre et Marie Curie, Examinateur

Monsieur Pierre RAY
Professeur des universités - Praticien Hospitalier, Université Grenoble Alpes, Directeur de thèse

Résumé

Titre : Exploration génétique et moléculaire de défauts post-méiotiques sévères de la spermatogenèse entrainant une infertilité masculine.

L'infertilité est considérée actuellement par l'organisation mondiale de la santé (OMS) comme une préoccupation majeure de santé affectant plus de 50 millions de couples dans le monde. Dans les pays occidentaux, la majorité des couples infertiles ont recours aux techniques d'assistance médicale à la procréation (AMP) pour obtenir une grossesse. Malgré le succès de ces techniques, près de la moitié des couples qui ont recours à l'AMP sortent du parcours de soin sans enfant. Une partie de ces échecs est expliquée par l'altération de la gamétogenèse. Chez l'homme, la spermatogenèse fait interagir des centaines de gènes spécifiquement exprimés dans le testicule. L'abondance de ces gènes suggère que les troubles de la spermatogenèse présentent une forte composante génétique. Récemment, les avancées techniques ont favorisé l'identification de gènes responsables de ces anomalies mais la grande majorité des cas d'infertilité masculine reste classée comme idiopathique. L'objectif de la thèse est d'identifier de nouvelles causes génétiques responsables d'infertilité masculine et d'élucider les mécanismes physiopathologiques associés à ces anomalies.

Au cours de ma thèse j'ai participé avec l'équipe GETI (génétique, épigénétique et thérapies de l'infertilité) à l'exploration génétique et moléculaire de deux phénotypes distincts d'anomalies spermatiques liés à des défauts post-méiotiques de la spermatogenèse : une forme rare d'azoospermie non obstructive (ANO) et le phénotype d'anomalies morphologiques multiples du flagelle spermatique (AMMF). Enfin j'ai joué un rôle important dans la création et l'analyse de modèles murins pour caractériser la pathogénie de ces anomalies.

L'analyse génétique de deux frères infertiles nés de parents consanguins et présentant une ANO idiopathique associée à un arrêt post-méiotique de la spermatogenèse nous a permis d'identifier un variant homozygote délétère dans le gène *SPINK2* qui code pour un inhibiteur de sérine-protéases. L'étude des souris KO pour ce gène nous a permis d'observer que les souris mâles adultes sont infertiles et miment parfaitement les phénotypes spermatique et testiculaire observés chez nos patients. Nous avons montré que la protéine codée par ce gène est exprimée dans l'acrosome à partir du stade de spermatide ronde. En l'absence de Spink2, l'activité protéolytique non-neutralisée des protéases cibles qui transitent par le Golgi cause sa fragmentation et bloque la spermiogénèse au stade de spermatide ronde. Nous avons également pu observer que les spermatozoïdes provenant de patients et de souris hétérozygotes présentent un taux élevé d'anomalies morphologiques et une baisse de la mobilité progressive conduisant à une hypofertilité à expressivité variable. Ces résultats montrent pour la première fois que l'oligo-tératozoospermie et l'azoospermie peuvent constituer un continuum pathologique dû à une même pathogénie.

Nous avons également réalisé le séquençage exomique complet d'une cohorte de 78 individus AMMF non apparentés et avons identifié chez 49 sujets des mutations bi-alléliques délétères dans 11 gènes candidats dont *DNAH1*, *CFAP43*, *CFAP44*, *WDR66* et *FSIP2*, soit un rendement diagnostique de 63%. Ces résultats confirment l'hétérogénéité génétique du phénotype MMAF et l'efficacité diagnostique du séquençage haut débit dans son exploration. Nous avons également validé l'implication de certains gènes candidats (n=4) dans ce phénotype chez le modèle murin knock-out créé par la nouvelle technologie d'édition du génome, CRISPR/Cas9.

Dans son ensemble, ce travail montre l'intérêt et l'efficacité de la combinaison du séquençage exomique et de la technique de CRISPR/Cas9 pour étudier les troubles de la spermatogenèse et l'infertilité masculine.

Mots clés : Infertilité, spermatogenèse, azoospermie, flagelle spermatique, séquençage exomique, CRISPR/Cas9.

Abstract

Title: Genetic and molecular exploration of severe post-meiotic defects of spermatogenesis leading to male infertility.

Infertility is currently considered by the World Health Organization (WHO) as a major health concern affecting more than 50 million couples worldwide. In western countries, the majority of infertile couples seek assisted reproductive technologies (ART) to achieve a pregnancy. Despite the success of these techniques, almost half of these couples fail to obtain a child. Part of these failures are explained by the alteration of gametogenesis. In humans, spermatogenesis involves hundreds of genes specifically expressed in the testis. The abundance of these genes suggests that spermatogenic defects are associated with a strong genetic component. Recently, technical advances have led to the identification of numerous causative genes, but the vast majority of male infertility cases remain idiopathic. The aim of the present thesis is to identify new genetic causes responsible for male infertility and to elucidate the physiopathological mechanisms associated with these anomalies.

During my thesis, I participated with the team GETI (genetics, epigenetics and therapies of infertility) in the genetic exploration of two phenotypes of male infertility related to post-meiotic defects of spermatogenesis: a rare form of non-obstructive azoospermia and the phenotype of multiple morphological abnormalities of the sperm flagella (MMAF). I have also played a key role in creation and analysis of transgenic mice to better characterize the pathogeny of the identified genetic causes in Human.

Genetic analyses performed on two infertile brothers born form consanguineous parents and presenting an-idiopathic non-obstructive azoospermia associated with a post-meiotic arrest of spermatogenesis allowed us to identify a homozygous variant in the *SPINK2* gene that encodes a serine-protease inhibitor. Phenotypic analysis of *Spink2*$^{-/-}$ adult male mice showed that they are infertile and perfectly mimic the sperm and testicular phenotypes observed in our patients. We showed that Spink2 protein is expressed from the round spermatid stage and localized in the acrosome, a lysosomal-like vesicle rich in proteases that play a key role during fertilization. When Spink2 is absent, the deregulated proteolytic activity of the targeted proteases such as acrosin leads to the fragmentation of the Golgi apparatus and arrest of spermiogenesis at the round spermatid stage. We also showed that sperm from heterozygous human and mice present a high level of morphological abnormalities and a decrease of progressive motility leading to a variable subfertility. These results showed for the first time that oligo-teratozoospermia and azoospermia could present a pathological continuum due to the same pathogeny.

We also performed exome sequencing in a cohort of 78 non related MMAF subjects and identified in 49 cases deleterious bi-allelic mutations in a total of 11 candidate genes including *DNAH1*, *CFAP43*, *CFAP44*, *WDR66* and *FSIP2* giving a genetic diagnosis yield of 63%. These results confirm the genetic heterogeneity of MMAF and the efficiency of high throughput sequencing in genetic exploration of this phenotype. We also demonstrated the pathogenic implication of certain candidate genes (n=4) using knock-out mice created by the new technology of genome editing, CRISPR/Cas9.

Overall, this work demonstrates the interest and effectiveness of combining exome sequencing and CRISPR/Cas9 system to study spermatogenesis disorders and male infertility.

Keywords: Infertility, spermatogenesis, azoospermia, sperm flagellum, exome sequencing, CRISPR/Cas9.

Je dédie cette thèse à mes parents Mounira et l'Hbib et à mon épouse Ismahane

REMERCIEMENTS

Je remercie *Madame* le Professeur Catherine PATRAT et *Monsieur* le Docteur Florian GUILLOU pour le temps qu'ils ont accepté de me consacrer afin d'évaluer mon travail de thèse et pour le rapport qu'ils en ont fait.

Je remercie *Madame* le Professeur Sophie CHRISTIN-MAITRE et *Monsieur* le docteur Saadi KHOCHBIN d'avoir accepté de participer à mon jury et de juger ce travail.

Je tiens à exprimer toute ma reconnaissance envers le Professeur Pierre RAY pour m'avoir accueilli dans son laboratoire et dirigé avec enthousiasme ces travaux, pour ses nombreux conseils, pour sa bonne humeur et pour sa disponibilité. Merci de recevoir mes remerciements les plus chaleureux.

Je remercie naturellement toutes les personnes de l'équipe GETI. Je remercie particulièrement le Docteur Christophe ARNOULT, le Professeur Serge BOTTARI et le professeur Charles COUTTON pour le cœur mis dans chacune de nos discussions et pour tout le savoir qu'ils ont su me transmettre avec passion.

De tout mon cœur, je remercie *Madame* Anne GUERIN-DUGUE, la directrice de l'école doctorale et son assistante, *Madame* Orélie GARZENA pour leur disponibilité, compréhension et gentillesse.

Je tiens également à remercier la direction de l'INSERM pour m'avoir sélectionné pour un poste d'accueil pour hospitaliers. La qualité de ce travail dépend fortement de ce poste et j'en suis très reconnaissant.

A toutes les personnes de l'Institut pour l'avancée des biosciences et en particulier au Professeur Pierre HAINAUT et son assistante Amélie FAUCONNET. Merci de m'avoir aidé dans mes démarches administratives.

Au doyen de la faculté de médecine *Monsieur* le Professeur Patrice MORAND et tous les membres de la commission recherche du CHUGA.

Au professeur Patrice FAURE, pour son accueil et intégration au sein de son département à l'IBP et à toutes les personnes exerçant au sein des UM de BGM et de GI-DPI.

A tous mes amis qui m'ont soutenu ces dernières années, mais qui sont trop nombreux pour les citer tous ici…

J'exprime toute ma gratitude à mes parents L'Hbib et Mounira, mes frères Mohamed, Sif et Adam ainsi que tout le reste de ma grande famille…

Je conclurai en remerciant de tout mon cœur mon épouse Ismahane pour son aide et son soutien inconditionnels et mes enfants Kays, Soujoud et Yassine qui illuminent ma vie chaque jour.

Merci infiniment...

TABLE DES MATIERES

LISTE DES FIGURES *15*

LISTE DES TABLEAUX *16*

LISTE DES ABREVIATIONS *17*

PREFACE *19*

PARTIE 1: ETUDE BIBLIOGRAPHIQUE *21*

Chapitre I. Aspects physiologiques et moléculaires de la fertilité masculine **23**

1- Rappel : l'appareil de reproduction masculin 23

2- Rôle des principales glandes annexes 24

2.1- Les vésicules séminales 24

2.2- La prostate 24

3- Anatomie du testicule 25

4- L'organogenèse gonadique et la différenciation testiculaire 26

4.1- La formation des crêtes génitales 26

4.2- Le développement des cellules germinales primordiales 26

4.3- La différenciation masculine de la gonade bipotentielle 28

4.4- La descente des testicules 30

5- Description générale de la spermatogenèse 30

6- Les cellules somatiques de Sertoli 34

6.1.1- Description morphologique et fonctionnelle 34

6.1.2- La barrière hémato-testiculaire 35

7- Maturation et différenciation des cellules germinales 36

7.1- Les cellules germinales souches et les spermatogonies 36

7.2- La méiose ou division réductionnelle 38

7.2.1- Les spermatocytes primaires et la méiose I 38

7.2.2- Les spermatocytes secondaires et la méiose II 40

7.3- La spermiogénèse 40

7.3.1- Biogenèse de l'acrosome 40

7.3.2- Hypercompaction de la chromatine 45

7.3.3- Formation du flagelle 47

7.3.3.1- L'axonème 48

A. Les dynéines axonémales 48

B. Les ponts radiaires 50

C. Le CSC et le N-DRC 52

D. Le complexe central 54

7.3.3.2- Les structures para-axonémales 55

8- La spermiation 57

9- La maturation épididymaire des spermatozoïdes 59

10- La régulation hormonale de la spermatogenèse 60

Chapitre II. L'infertilité masculine **64**

1- Définition et données épidémiologiques 64

2- Les facteurs pré-testiculaires associés à l'infertilité masculine 65

2.1- L'hypogonadisme hypogonadotrope 65

2.2- L'hyperprolactinémie ... 65
2.3- Rôle des perturbateurs endocriniens ... 66
3- Les facteurs testiculaires associés à l'infertilité masculine ... 66
3.1- La varicocèle ... 66
3.2- Le cryptorchidisme ... 67
3.3- Les facteurs génétiques ... 67
3.3.1- L'azoospermie non-obstructive ... 67
3.3.1.1- Le syndrome de Klinefelter ... 68
3.3.1.2- Les microdélétions sur le bras long du chromosome Y ... 69
3.3.2- Les tératozoospermies monomorphes ... 70
3.3.2.1- La macrozoospermie : les spermatozoïdes macrocéphales multi-flagellés ... 70
3.3.2.2- La globozoospermie : les spermatozoïdes microcéphaliques ... 73
3.3.2.3- L'asthénozoospermie et le phénotype MMAF ... 76
4- Les infertilités obstructives ... 78
4.1- L'absence bilatérale des canaux déférents (ABCD) ... 78
4.1.1- Le gène CFTR (cystic fibrosis transmembrane regulator) ... 78
4.1.2- Le gène ADGRG2 (adhesion G protein-coupled receptor G2) ... 80
4.2- Le syndrome de Young ... 81
4.3- Les obstructions post-infectieuses et iatrogènes ... 81

Chapitre III. Les investigations cliniques et paracliniques de l'infertilité masculine ... 82
1- L'examen clinique ... 82
2- L'analyse du sperme ... 82
2.1- Le spermogramme ... 82
2.1.1- Le spermocytogramme ... 84
2.2- Le dosage des biomarqueurs séminaux ... 85
3- L'exploration hormonale de l'axe gonadotrope ... 85
4- Les investigations génétiques de routine ... 85
4.1- Identification de nouveaux gènes candidats par le séquençage exomique ... 86
4.1.1- Introduction ... 86
4.1.2- Principes du séquençage et étapes de l'analyse bio-informatique des données ... 87
4.1.2.1- Rappel : le séquençage conventionnel de Sanger ... 88
4.1.2.2- Le séquençage exomique ... 88
E. La préparation de l'ADN ... 88
F. L'enrichissement ... 89
G. L'amplification et la réaction de séquence ... 89
4.1.2.3- L'analyse bio-informatique ... 89
H. Objectifs et principes ... 89
I. Cas particulier : détection des variants de nombre de copies (CNVs, copy number variants) ... 91
4.2- Validation des gènes candidats chez le modèle murin CRISPR/Cas9 ... 92
4.2.1- Introduction ... 92
4.2.2- Historique de la transgénèse murine ... 92
4.2.3- Le système CRISPR/Cas9 ... 93
4.2.4- Article 1 : Creation of knock out and knock in mice by CRISPR/Cas9 to validate candidate genes for human male infertility, interest, difficulties and feasibility ... 95

Chapitre IV. Les techniques d'Assistance Médicale à la Procréation (AMP) ... 112
1.1- L'Insémination Artificielle (IA) ... 113

1.2- La Fécondation In Vitro (FIV) ... 113
1.2.1- La stimulation ovarienne ... 113
1.2.2- Le recueil des gamètes ... 113
1.2.3- La FIV proprement dite ... 114
1.2.4- Le transfert embryonnaire ... 114
1.3- La FIV avec micro-injection intracytoplasmique du spermatozoïde ... 115

PARTIE 2 : ETUDE EXPERIMENTALE ... ***118***

Objectifs ... **120**

Chapitre I. L'azoospermie non-obstructive ... **122**
1- Article 2 : SPINK2 deficiency causes infertility by inducing sperm defects in heterozygotes and azoospermia in homozygotes ... 122
2- Discussion ... 158

Chapitre II. Le phénotype MMAF ... **161**
1- Article 3 : A homozygous ancestral SVA-insertion mediated deletion in *WDR66* induces multiple morphological abnormalities of the sperm flagellum and male infertility ... 161
2- Article 4 : Mutations in *CFAP43* and *CFAP44* cause male infertility and flagellum defects in Trypanosoma and human ... 207
3- Article 5 : Whole-exome sequencing identifies mutations in *FSIP2* as a recurrent cause of multiple morphological abnormalities of the sperm flagella ... 256
4- Article 6 : Whole-exome sequencing of familial cases of multiple morphological abnormalities of the sperm flagella (MMAF) reveals new *DNAH1* mutations ... 293
5- Discussion ... 305
5.1- Exploration génétique du phénotype MMAF ... 305
5.2- Intérêt des modèles animaux dans l'étude du phénotype MMAF ... 308
5.3- Conclusion ... 311

PARTIE 3 : Discussion générale ... ***313***
1- Exploration génétique de l'infertilité par séquençage haut débit ... 315
2- Approches thérapeutiques ciblées dans l'infertilité masculine ... 316
3- Conclusion ... 319

Références ... ***321***

ANNEXES ... ***339***
4- Article Annexe 1 : PATL2 is a key actor of oocyte maturation whose invalidation causes infertility in women and mice ... 341
5- Article Annexe 2 : Bases moléculaires et physiopathologiques des tératozoospermies monomorphes ... 387

LISTE DES FIGURES

Figure 1. Anatomie de l'appareil de reproduction masculin. 23
Figure 2. Anatomie descriptive du testicule. 25
Figure 3. Formation des crêtes génitales chez la souris. 26
Figure 4. Les gonades à la sixième semaine du développement embryonnaire. 27
Figure 5. Régulation génétique du développement de la gonade masculine. 29
Figure 6. Différenciation gonadique masculine chez la souris. 30
Figure 7. Description générale de la spermatogenèse. 31
Figure 8. Le cycle de l'épithélium séminifère chez l'homme. 33
Figure 9. La cellule de Sertoli. 34
Figure 10. La barrière hémato-testiculaire (blood-testis barrier) 36
Figure 11. Développement des cellules germinales en post-natal chez l'humain. 37
Figure 12. La prophase I. 39
Figure 13. Assemblage et dissociation du complexe synaptonémal durant la prophase I. 40
Figure 14. Schéma de l'acrosome et de ces différentes membranes avant et après la réaction acrosomique 41
Figure 15. Présentation schématique des principales phases de la biogenèse de l'acrosome. 42
Figure 16. Organisation spatiale de la manchette dans la spermatide en élongation. 44
Figure 17. Remplacement des histones par les protamines et hypercompaction de la chromatine spermatique. 46
Figure 18. Formation du flagelle au cours de la spermiogenèse. 47
Figure 19. L'axonème du flagelle spermatique. 48
Figure 20. Architecture d'une chaîne lourde de dynéine 49
Figure 21. Les bras internes et externes de dynéines. 50
Figure 22. Localisation des ponts radiaires dans l'axonème de *Chlamydomonas*. 51
Figure 23. Impact du déficit en FAP61 et 251 sur la structure axonémale de Tétrahyména. 53
Figure 24. Localisation du N-DRC. 54
Figure 25. Organisation du complexe central 55
Figure 26. Présentation schématique de l'ultrastructure du flagelle spermatique. 56
Figure 27. Présentation schématique du processus de spermiation. 58
Figure 28. Présentation schématique de l'organisation de l'épididyme chez l'homme. 59
Figure 29. Régulation hormonale de la spermatogenèse. 61
Figure 30. Complexité de la régulation endocrine et paracrine de la spermatogenèse. 63
Figure 31. Anomalies chromosomiques dans le syndrome de Klinefelter 68
Figure 32. Microdélétions du chromosome Y. 69
Figure 33. La macrozoospermie : Rôle de l'AURKC durant la méiose et la cytokinèse. 72
Figure 34. La globozoospermie. 73
Figure 35. Physiopathologie de la globozoospermie. 75
Figure 36. Le phénotype MMAF. 77
Figure 37. La spermatogenèse dans l'azospermie. 78
Figure 38. Classification des mutations du gène *CFTR*. 80
Figure 39. Vue d'ensemble de la stratégie d'identification d'un variant candidat par séquençage exomique. 87
Figure 40. Procréation naturelle versus procréation médicalement assistée. 112
Figure 41. ICSI, intracytoplasmic sperm injection. 115

Figure 42. Comparaison des techniques d'extraction des spermatozoïdes épididymaires et testiculaires. 116
Figure 43. Synthèse sur la localisation intra-flagellaire de certaines protéines impliquées dans le phénotype MMAF chez nos patients. 312

LISTE DES TABLEAUX

Tableau 1. Caractéristiques des protéines co-précipitées avec la calmoduline. 52
Tableau 2. Limites inférieures de référence des paramètres du sperme humain (OMS, 2010). 83
Tableau 3. Comparaison des techniques d'extraction des spermatozoïdes épididymaires et testiculaires. 117
Tableau 4. Liste des gènes identifiés par WES chez une cohorte de 78 sujets MMAF. 306

LISTE DES ABREVIATIONS

ABCD : absence bilatérale des canaux déférents
ABP : androgen-binding protein
ADGRG2 : adhesion G protein-coupled receptor G2
ADN : acide désoxyribonucléique
AMH : anti-Müllerian hormone
AMP : assistance médicale à la procréation
ANO : azoospermie non obstructive
AO : azoospermie obstructive
ARN : acide ribonucléique
AZF : Azoospermia factor
BHT : Barrière hémato-testiculaire
bp : base pair
Cas9 : crispr associated protein 9
CASA : Computer-Assisted Sperm Analysis
CFTR : Cystic fibrosis transmembrane conductance regulator
CRISPR : clustred regulatory interspaced short palindromic repeats
crRNA : crispr RNA
CSC : calmodulin and spoke associated complex
ddNTP : didésoxyribonucléotide
dNTP : désoxyribonucléotide
FIV : fécondation in vitro
FSH : Follicle-stimulating hormone
GnRH : Gonadotropin Releasing Hormone
HCG : human Chorionic Gonadotropin
HDR : homology directed repair
HTS : high-throughput sequencing
IA : Insémination artificielle
ICSI : intra cytoplasmic sperm injection
KI : knock-in
KO : knock-out
LH : luteinising hormone
MESA : microsurgical epididymal sperm
Micro-TESE : microsurgical testicular sperm extraction
MMAF : multiple morphological abnormalities of the sperm flagella
N-DRC : nexin-dynein regulatory complex
NHEJ : non-homologous end-joining
NMD : nonsense-mediated decay
OATS : oligo-asthéno-térato-spermie
OMS : organisation mondiale de la santé
OS : oligozoospermie sévère
PAM : protospacer adjacent motif
pb : paire de bases
PESA : percutaneous epididymal sperm aspiration
PLCZ : phospholipase C zeta
PRM : Protamine
sgRNA : Single guide RNA
SRY : sex-determining region of Y chromosome
ssDNA : single-stranded DNA
TALEN : transcription activator-like effector nuclease
TESA : percutaneous testicular sperm aspiration
TESE : conventional testicular sperm extraction
TNP : transition protein
tracrRNA : trans-activating crRNA
WES : Whole exome sequencing
WGS : Whole genome sequencing
ZFN : zinc finger nuclease

PREFACE

Le travail présenté dans cette thèse a été réalisé dans l'institut pour l'avancée des biosciences (IAB, INSERM 1209, CNRS UMR 5309) à Grenoble au sein de l'équipe GETI (génétique, épigénétique et thérapies de l'infertilité), sous la supervision du professeur Pierre Ray (PU-PH). La première année de ma thèse s'est déroulée en parallèle de mon activité hospitalière au sein de l'UM de génétique de l'infertilité et du diagnostic préimplantatoire moléculaire (GI-DPI) dirigée par le Pr. Ray, puis j'ai bénéficié d'un poste d'accueil INSERM pour hospitaliers qui m'a permis de me consacrer à mes travaux de recherche à plein temps pendant les deux dernières années (2016-2018).

L'objectif global de cette thèse est l'amélioration de nos connaissances sur les causes génétiques des défauts spermatogéniques sévères entrainent une infertilité masculine. Trois parties sont incluses dans ce manuscrit :

La première partie intitulée « étude bibliographique », présente les principaux aspects physiologiques et moléculaires de la fertilité masculine et décrit les principes de la prise en charge clinique et paraclinique des hommes infertiles. Elle introduit également notre stratégie d'exploration génétique des cas idiopathiques par le séquençage exomique et rapporte à travers un article de revue (Article 1), l'efficacité de la validation des variants et des gènes candidats associés par l'étude de souris transgéniques créées par la technique CRISPR/Cas9.

La seconde partie expose les études expérimentales réalisées au cours de cette thèse. Elle est organisée sous forme d'une série d'articles associés à deux chapitres :
Chapitre 1 : Caractérisation moléculaire et physiopathologique d'une nouvelle cause génétique de l'azoospermie non-obstructive (Article 2).
Chapitre 2 : Caractérisation génétique et moléculaire du phénotype MMAF (Articles 3, 4, 5 et 6).
La dernière partie présente une discussion générale de l'intérêt du séquençage haut débit dans l'exploration génétique de l'infertilité masculine et des approches thérapeutiques ciblées heurtées à la sphère éthique et réglementaire actuelle.

Finalement, en annexe, est présenté un article original portant sur l'identification et la caractérisation moléculaire et physiopathologique d'une nouvelle cause génétique de l'infertilité féminine (article annexe 1) et un article de revue sur les bases moléculaires et physiopathologiques des tératozoospermies monomorphes (article annexe 2).

PARTIE 1: ETUDE BIBLIOGRAPHIQUE

Chapitre I. Aspects physiologiques et moléculaires de la fertilité masculine

1- Rappel : l'appareil de reproduction masculin

L'appareil de reproduction masculin assure la production de spermatozoïdes, leur nutrition, stockage, transport et expulsion. Il assure aussi une fonction endocrine par la synthèse et la sécrétion de stéroïdes androgènes. Il est composé de deux testicules, de voies spermatiques excrétrices (épididymes, canaux déférents et urètres), de glandes annexes (vésicules séminales, prostate et glandes de cowper) et du pénis **(Figure 1)**.

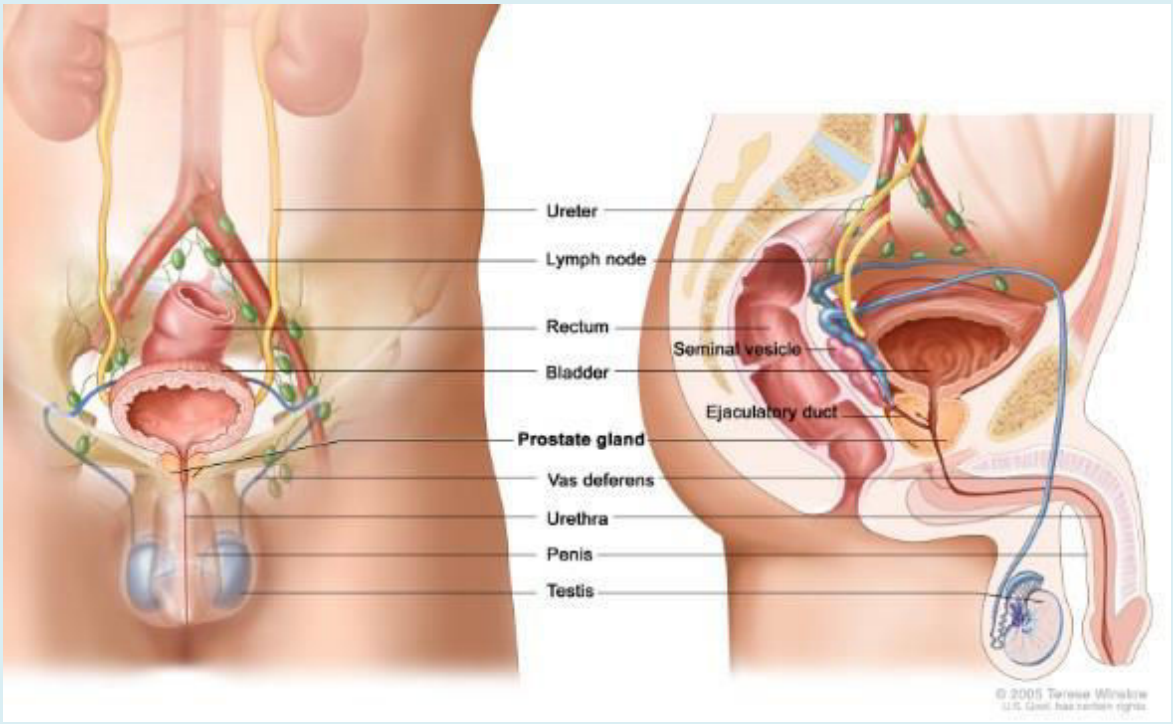


Figure 1. Anatomie de l'appareil de reproduction masculin.
Depuis : https://www.ncbi.nlm.nih.gov.

2- Rôle des principales glandes annexes

2.1- Les vésicules séminales

Les vésicules séminales sont des glandes exocrines au nombre de 2 situées derrière le col vésical au-dessus de la prostate. Chez l'homme, ces vésicules mesurent environ 15 cm de long et pèsent près de 5 g chacune. Elles élaborent et sécrètent un liquide de pH légèrement alcalin riche en fructose et en prostaglandines. Le volume de ces sécrétions constitue environ deux tiers du volume spermatique. La formation et la croissance de ces glandes sont très dépendantes des androgènes. L'élévation des taux plasmatiques de testostérone augmente l'activité sécrétoire de ces glandes (Gonzales, 2001). L'activité sécrétoire des vésicules séminales est également modulée par les systèmes nerveux sympathiques et parasympathiques via les récepteurs adrénergiques et cholinergiques respectivement (Hsieh et al., 2014).

Les sécrétions séminales représentent les fractions tardives de l'éjaculat. Ces sécrétions permettent de coaguler le sperme dès son éjaculation. Cette étape est importante pour mettre les spermatozoïdes en contact avec les composants de ces secrétions et promouvoir leur mobilité et la stabilité de leur chromatine (Du Plessis et al., 2013). La coagulation du sperme est induite par la fixation des ions de zinc d'origine prostatique sur la séménogéline (SgI et SgII), une glycoprotéine majeure dans les sécrétions séminales. Sa dégradation graduelle par l'activité protéolytique de la PSA (prostate specific antigen) contenue dans les sécrétions prostatiques permet la liquéfaction du sperme et la libération des spermatozoïdes (Yoshida et al., 2008). Les sécrétions séminales jouent également un rôle important dans le tractus génital féminin par la suppression de l'activité immunitaire dirigée contre les spermatozoïdes qui pourrait conduire à leur rejet (Pakravan et al., 2015).

2.2- La prostate

Chez l'homme, la prostate est une glande de la taille d'une châtaigne, située à la base de la vessie autour de l'urètre. La prostate est la cible directe de nombreuses pathologies bénignes et malignes qui sont potentiellement associées à une infertilité (Verze et al., 2016).

Cette glande produit des sécrétions légèrement acides contenant de l'acide citrique, des phosphatases, la PSA et le zinc (Mawhinney Michael and Mariotti Angelo, 2012). Ces sécrétions contiennent également des prostasomes qui sont de petites vésicules

d'une taille comprise entre 30 et 500 nm sécrétées par les cellules épithéliales. Ces vésicules sont entourées d'une bicouche lipoprotéique et ont la capacité de véhiculer des molécules comme les lipides et les protéines aux spermatozoïdes durant leur transit dans le tractus génital féminin et pourraient jouer à ce titre un rôle important dans la fécondation (Saez and Sullivan, 2016).

3- Anatomie du testicule

Les testicules sont deux organes symétriques de forme ovoïde. Ils sont logés dans le scrotum et sont entourés par une enveloppe rigide nommée l'albuginée ou tunica albuginea. Le parenchyme est divisé en 200-300 lobules séparés par des invaginations de la tunica albuginea. La partie supérieure de cette tunique présente un épaississement triangulaire (corps de Highmore) qui contient les canaux du rete testis. C'est au niveau de cette dernière structure que converge l'ensemble des tubules séminifères du parenchyme testiculaire **(Figure 2)**.

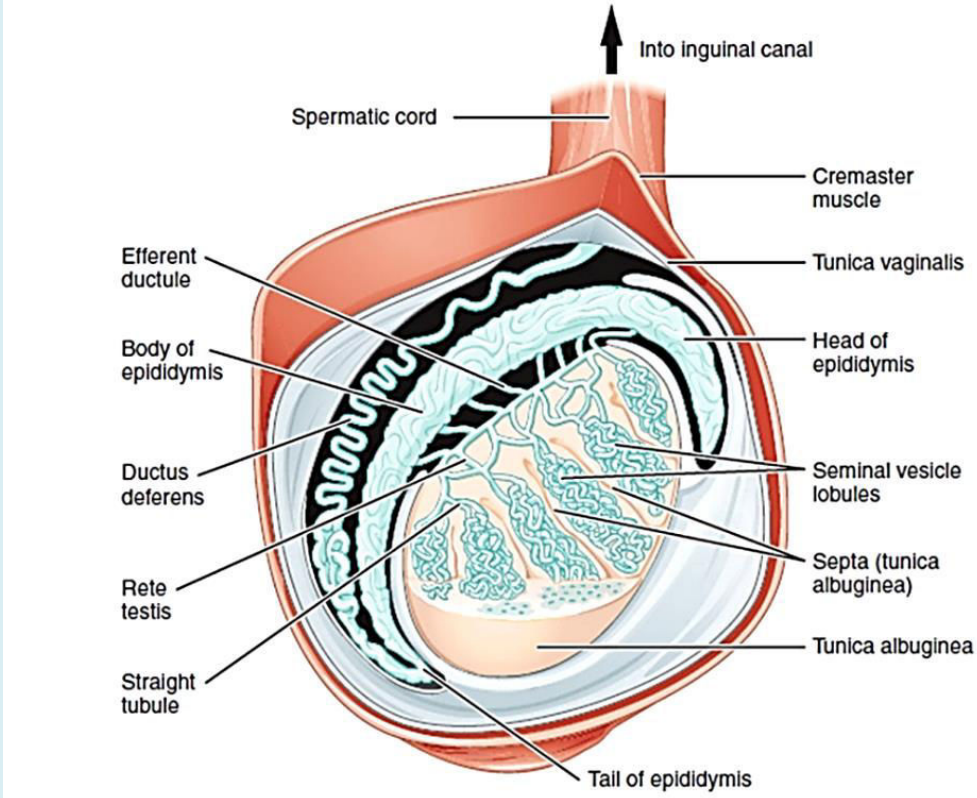


Figure 2. Anatomie descriptive du testicule.
Depuis : http://cnx.org.

4- L'organogenèse gonadique et la différenciation testiculaire

4.1- La formation des crêtes génitales

Le développement des gonades et leur différenciation est un processus moléculaire très complexe qui fait intervenir et interagir un grand nombre de gènes de manière synergique ou antagoniste (Bashamboo et al., 2017).

Au cours de la quatrième semaine du développement embryonnaire, les crêtes génitales se mettent en place sur les côtés médiaux du mésonéphros (rein embryonnaire transitoire) et s'étendent de la région cardiaque à la région cloacale **(Figure 3)**. Au niveau de la région moyenne de la crête, le mésenchyme croît par prolifération et recrutement des cellules du mésonéphros en plus de l'épaississement de l'épithélium cœlomique pour former l'ébauche génitale qui sera envahie rapidement par les cellules germinales primordiales.

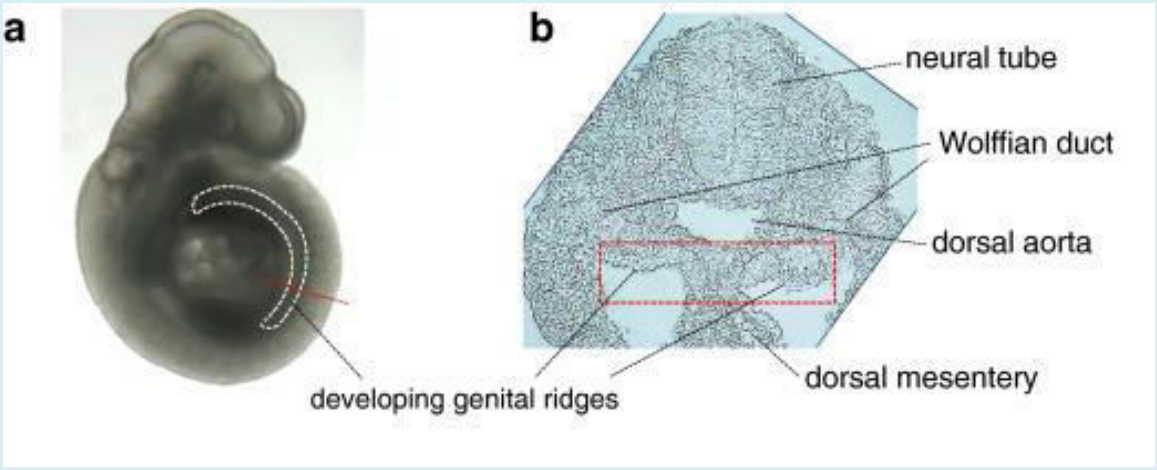


Figure 3. Formation des crêtes génitales chez la souris.
(a) Embryon de 10 jours. La ligne blanche discontinue indique la localisation des crêtes génitales en développement et (b) représente une coupe transversale au niveau de ces formations (d'après Tanaka and Nishinakamura, 2014).

4.2- Le développement des cellules germinales primordiales

Les cellules germinales primordiales (CGPs) se développent à partir de l'épiblaste postérieur proximal dès la troisième semaine de développement embryonnaire. Ce processus est principalement contrôlé par le facteur de croissance BMP4 qui est sécrété par l'ectoderme extra-embryonnaire (Dolci et al., 2015). Les CGPs subissent une migration extra-embryonnaire dans la paroi postérieure du sac vitellin puis migrent de nouveau dans la paroi

du tube digestif pour finalement coloniser les ébauches génitales. Pendant leurs vagues de migration entre la quatrième et sixième semaine, les CGPs continuent à proliférer par mitoses.

Jusqu'à la sixième semaine de développement, les ébauches génitales ne sont pas différenciées chez les deux sexes **(Figure 4)**. Plusieurs gènes ont été décrits comme étant critiques pour le développement des gonades bipotentielles chez la souris comme *Lhx9*, *Emx2*, *Wt1*, *Cbx2*, *Nr5a1*, *Gata4* et *Six1/4* (Lucas-Herald and Bashamboo, 2014). Chez l'Homme, en raison des restrictions qui contraignent la recherche sur les embryons, les processus moléculaires qui contrôlent la formation des gonades bipotentielles sont peu connus. En effet, bien que ces processus semblent être très similaires chez les mammifères, des différences existent. A ce titre, il a été montré que l'expression de certains facteurs de transcription est complétement divergente chez l'Homme et la souris. Par exemple, le gène *Sox2* est exprimé préférentiellement chez la souris à l'inverse de *SOX17* qui est préférentiellement exprimé chez l'Homme (Irie et al., 2015).

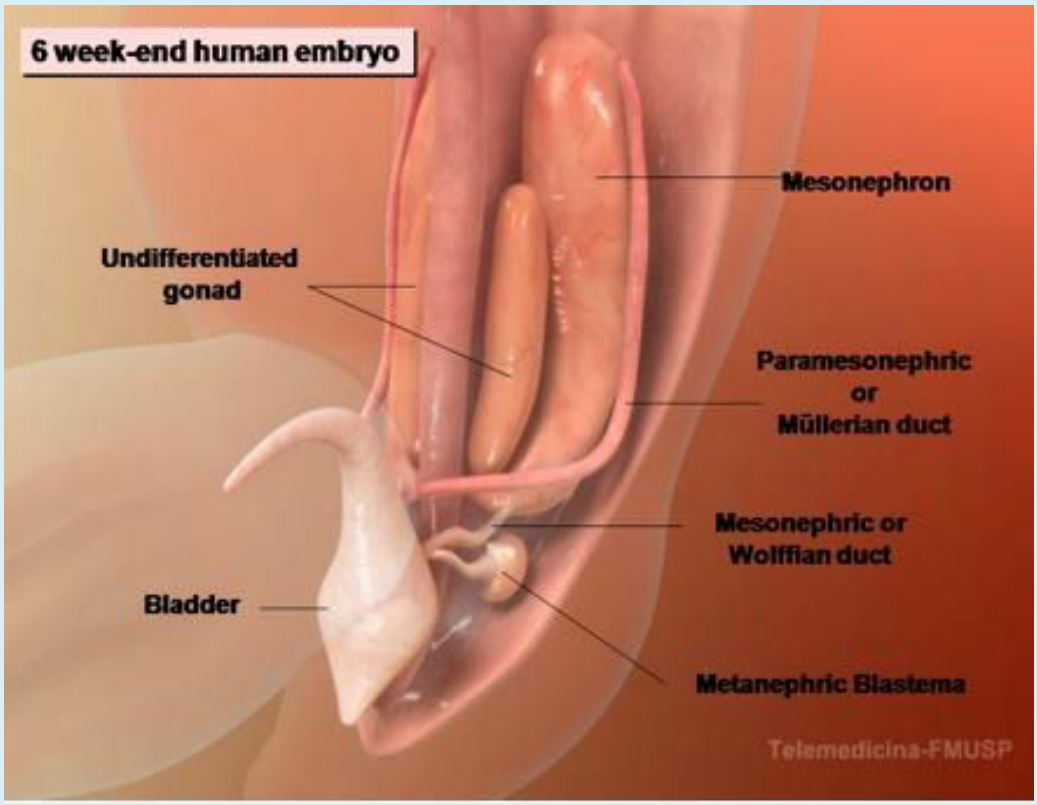


Figure 4. Les gonades à la sixième semaine du développement embryonnaire.
(D'après Domenice et al., 2017)

4.3- La différenciation masculine de la gonade bipotentielle

La différenciation sexuelle des gonades bipotentielles commence vers la septième semaine. La différenciation des cellules de Sertoli constitue le premier évènement observé de l'organogenèse testiculaire. L'expression du gène *SRY (sex-determining region of Y chromosome)* par certaines cellules somatiques dérivées de l'épithélium cœlomique de l'ébauche génitale est responsable de l'induction de cette différenciation.

Le gène *SRY* est localisé sur le bras court du chromosome Y (Yp11.31). Il a été observé que les sujets de sexe génétique féminin (46,XX) mais portant une copie du gène *SRY* développent une gonade masculine (Nakashima et al., 2014; Wu et al., 2014). Des observations similaires ont été également observées chez la souris transgénique. Alors que les souris XX portant le gène *Sry* développent une gonade masculine, les souris XY délétées pour le gène *Sry* développent plutôt une gonade féminine (Gubbay et al., 1992; Hawkins et al., 1992; Koopman et al., 1990).

L'expression du gène *SRY* active le gène autosomique *SOX9* et induit la différenciation des cellules de soutien de Sertoli. Plusieurs boucles de régulation se mettent alors en place pour maintenir l'expression du gène *SOX9* à un niveau élevé et contrôler le développement testiculaire **(Figure 5)** (Larney et al., 2014). Actuellement, le mécanisme moléculaire exact de la différenciation gonadique n'est pas encore connu.

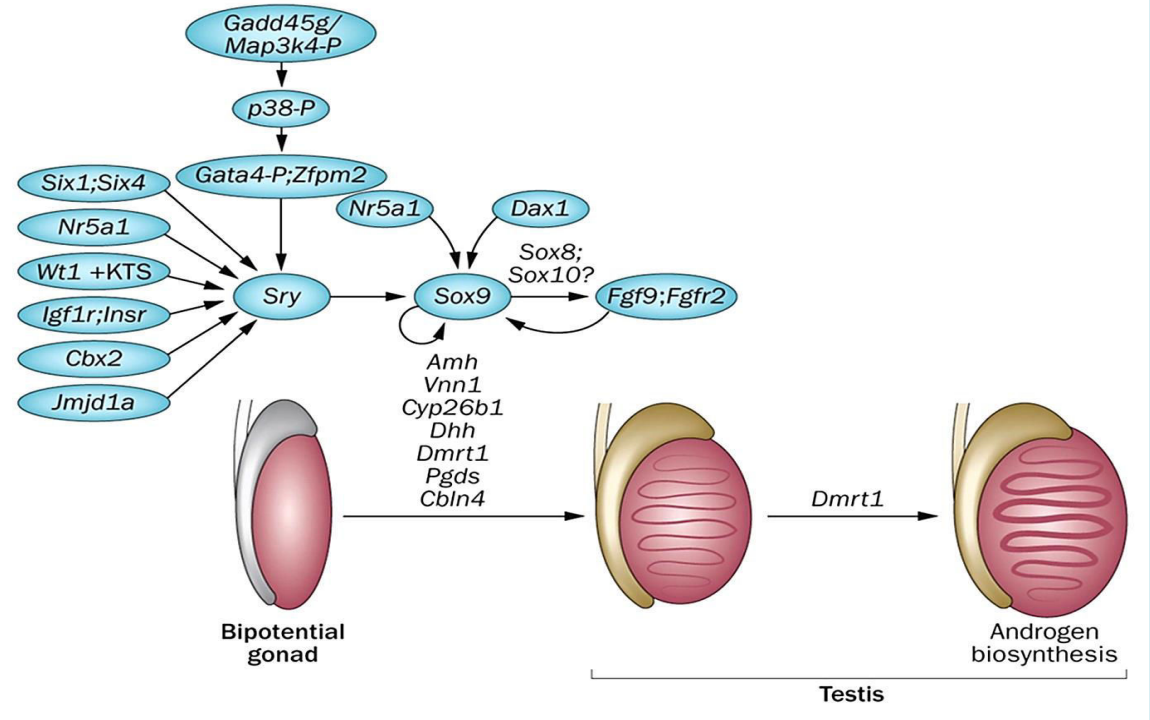


Figure 5. Régulation génétique du développement de la gonade masculine.
(D'après Eggers et al., 2014).

Une fois les cellules de Sertoli différenciées, elles établissent entre elles des jonctions cellulaires et envoient des prolongements cytoplasmiques autour des cellules germinales pour les englober et former des cordons sexuels, futurs tubules séminifères droits et contournés. Des cellules somatiques d'origine mésonéphrotiques participent à la formation externe de ces cordons et se différencient en cellules myoïdes péritubulaires. Des cellules du mésenchyme vont se différencier en cellules de Leydig. Le mésenchyme séparant les cordons sexuels se différencie en tissu conjonctif dense et s'organise pour subdiviser le parenchyme testiculaire en lobules et former la tunique albuginée à la surface. L'épithélium cœlomique se transforme en cavité séreuse autour du testicule. Finalement, le rete testis qui fait suite aux tubules séminifères est connecté au cours du troisième mois aux canaux efférents d'origine mésonéphrotiques qui se sont mis en place pendant la neuvième semaine **(Figure 6)**.

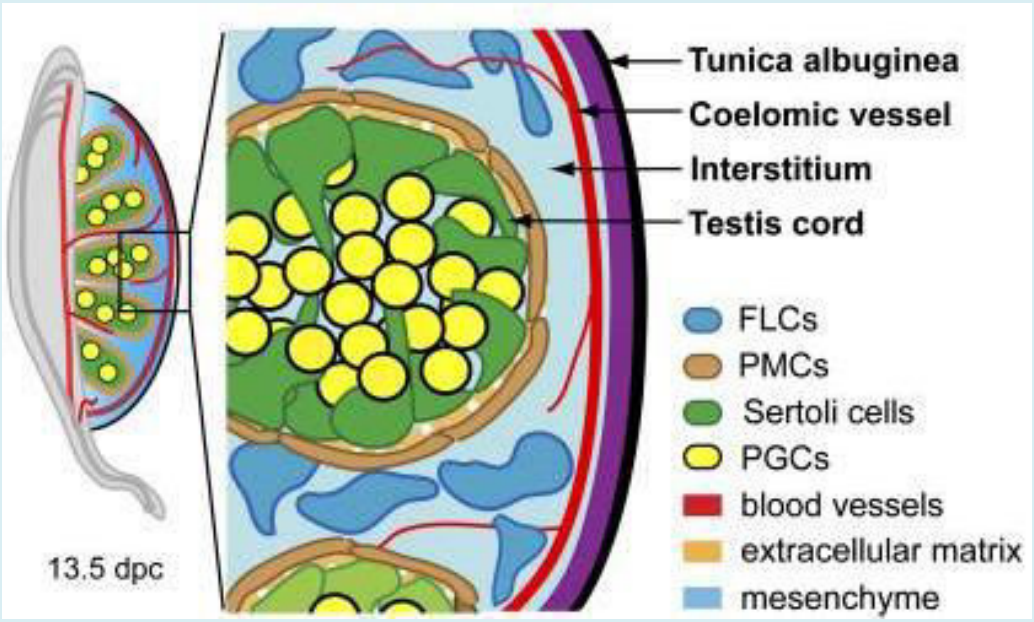


Figure 6. Différenciation gonadique masculine chez la souris.
FLCc: fetal Leydig cells. PMCs: peritubular myoid cells. PGCs: primordial germ cells (d'après Svingen and Koopman, 2013).

4.4- La descente des testicules

La descente des testicules vers la loge scrotale se fait de manière progressive jusqu'au neuvième mois. Cette décente est soutenue par la rétraction du gubernaculum, un ligament qui connecte la base du testicule au scrotum. Le raccourcissement de ce ligament tire vers le bas le testicule et le fait passer par le canal inguinal pour se retrouver dans sa loge scrotale. En passant par le canal inguinal, le testicule entraine avec lui une partie du péritoine appelée processus vaginal.

5- Description générale de la spermatogenèse

La spermatogenèse est un processus biologique complexe évoluant dans les tubules séminifères du testicule. Ce processus produit de manière continue les gamètes males haploïdes (les spermatozoïdes) à partir des cellules souches germinales diploïdes (les spermatogonies) (Griswold, 2016) à travers trois étapes successives : 1) la spermocytogenèse : prolifération des spermatogonies et différenciation en spermatocytes, 2) la méiose: division méiotique des spermatocytes et production des spermatides et 3) la spermiogénèse : différenciation des spermatides ronds en spermatozoïdes **(Figure 7)**. L'ensemble de ce processus se déroule en 35 jours chez la souris et 64 jours chez l'homme (Schlatt and

Ehmcke, 2014). Ce processus permet la production massive et continue de spermatozoïdes au cours de la vie reproductive de l'individu. On estime que chaque testicule produit environ 45 millions de spermatozoïdes par jour soit près de 1000 spermatozoïdes par seconde (Griswold, 2016).

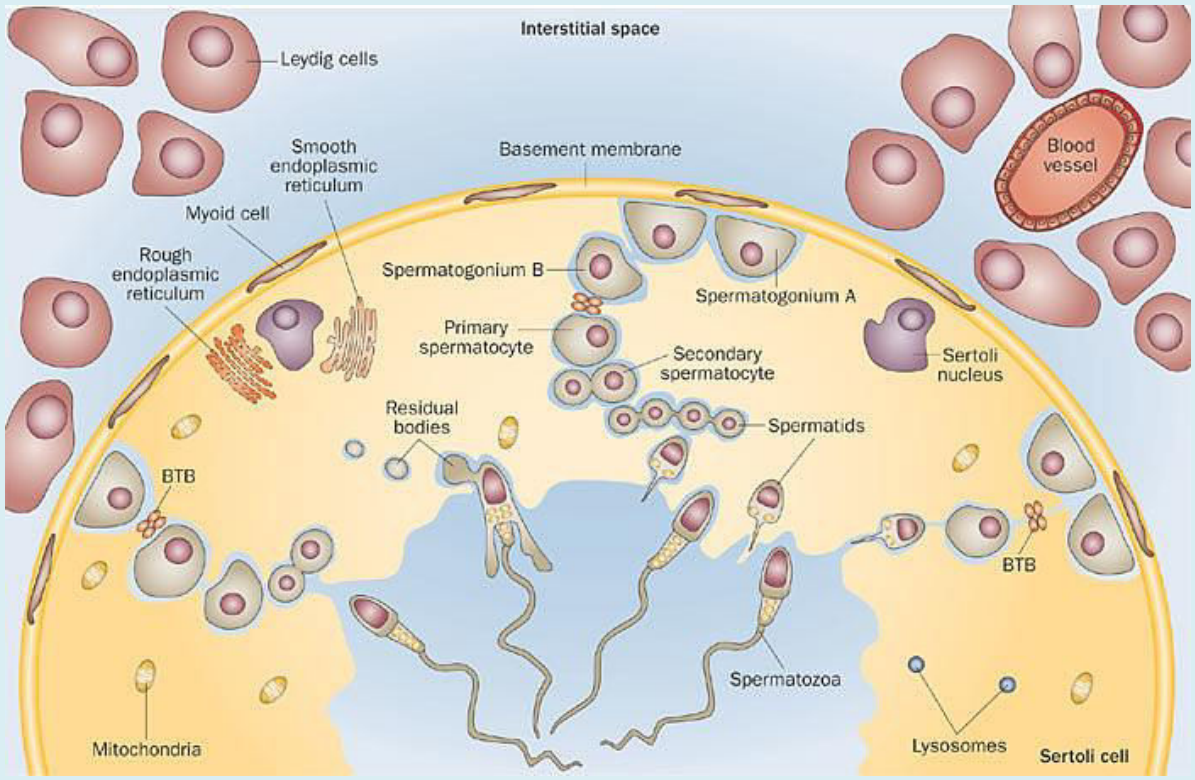


Figure 7. Description générale de la spermatogenèse.
Au sein de l'épithélium séminifère, on distingue deux types de populations cellulaires, les cellules somatiques de Sertoli et les cellules germinales en maturation depuis les spermatogonies jusqu'aux spermatozoïdes. Les cellules myoïdes sont localisées en périphérie des tubules séminifères. Entre ces tubules, est disposé l'espace interstitiel qui contient, entre autres, les cellules de Leydig, les macrophages et les vaisseaux sanguins. BTB = blood-testis barrier (d'après Rato et al., 2012).

La spermatogenèse représente un modèle d'organisation cellulaire spatio-temporelle très complexe. Outre l'existence d'une barrière hémato-testiculaire qui compartimentalise ces cellules selon leur mode de division (mitotique ou méiotique), la présence de vagues de progression cellulaire le long des tubules séminifères rythme également ce processus (Bittman, 2016). Ces vagues semblent être fortement dépendantes de l'action de la vitamine A qui agit via son métabolite ATRA (*all-trans* retinoic acid) (Teletin et al., 2017). Il a été montré que L'ATRA induit la différenciation des spermatogonies et

l'engagement des spermatocytes primaires en méiose de manière synchronisée. Cependant le mécanisme moléculaire en aval de l'exposition cellulaire à ce métabolite reste inconnu (Busada and Geyer, 2016).

L'induction de vagues spermatogéniques successives le long des tubules séminifères explique l'observation de plusieurs associations cellulaires possibles au niveau des sections transversales des tubules séminifères. Ces associations définissent ensemble le cycle de l'épithélium séminifère. La classification de ces associations est généralement basée sur l'observation de la morphologie de l'acrosome des spermatides après une coloration PAS (periodic acid-Shiff) (Schlatt and Ehmcke, 2014). Chez l'homme, six associations ont été définies de I à VI **(Figure 8)** (Amann, 2008).

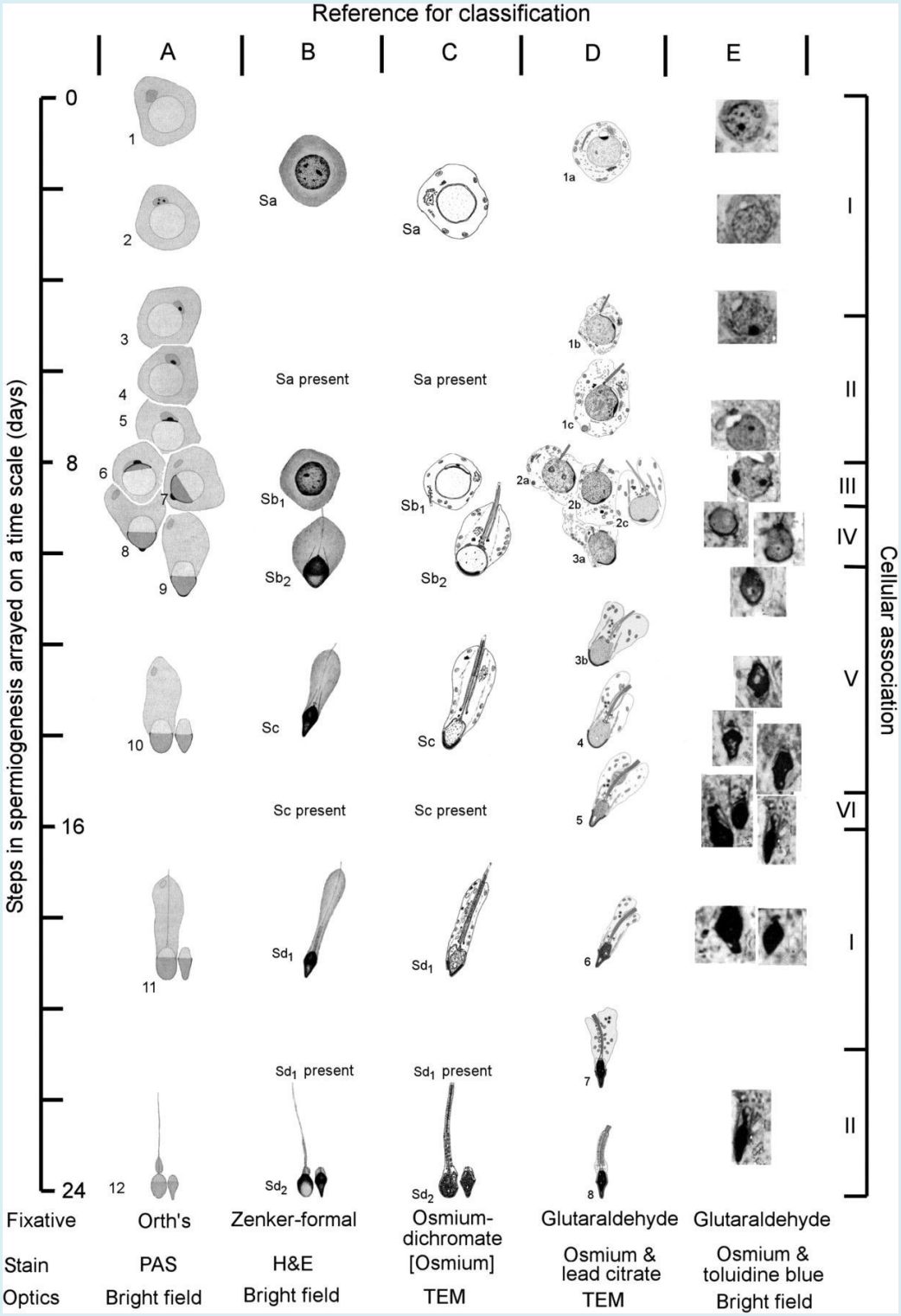


Figure 8. Le cycle de l'épithélium séminifère chez l'homme.
Les colonnes (A-E) représentent l'aspect des cellules germinales (spermatides rondes et allongées) en microscopie optique ou électronique suite à des colorations spécifiques dont la coloration PAS (en colonne A) qui permet d'observer la forme de l'acrosome au cours du cycle de l'épithélium séminifère (stades I-VI) (d'après Amann, 2008).

6- Les cellules somatiques de Sertoli

6.1.1- Description morphologique et fonctionnelle

Les cellules de Sertoli sont de grandes cellules reposant sur la lame basale et dotées de plusieurs prolongements cytoplasmiques qui s'insinuent entre les cellules germinales. Elles ont un alignement polarisé avec une base reposant sur la membrane basale et un apex orienté vers la lumière du tubule **(Figure 9)** (França et al., 2016). Les cellules de Sertoli jouent un rôle important durant la spermatogenèse. Elles assurent le support, la protection et la nutrition des cellules germinales (Dimitriadis et al., 2015). Une seule cellule de Sertoli supporte un nombre défini de cellules germinales à différents stades de développement. De ce fait, le nombre total de cellules de Sertoli est en corrélation positive avec le nombre de spermatozoïdes produits et participe à la définition du potentiel spermatogénique du testicule. Les cellules de Sertoli sont dotées d'un pouvoir de phagocytose. Ce processus leur permet d'éliminer les cellules germinales dégénérées et les résidus cytoplasmiques des spermatides (Iliadou et al., 2015).

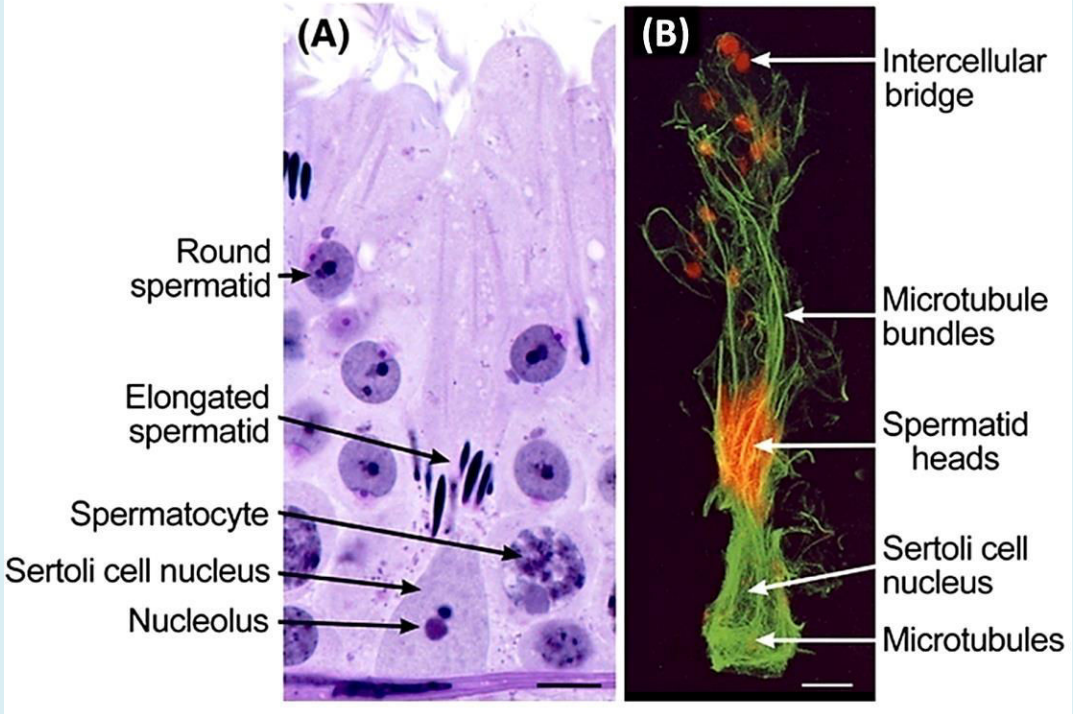


Figure 9. La cellule de Sertoli.
(A) En microscopie optique après coloration H&E. (B) En microscopie confocale après marquage de la tubuline (en vert) et l'actine en (rouge) (d'après França et al., 2016).

Au cours de la vie intra-utérine, la cellule de Sertoli primitive sécrète l'AMH qui est responsable de la régression des conduits mülleriens (Iliadou et al., 2015).

La cellules de Sertoli exprime le récepteur de la FSH et synthétise l'inhibine B, une glycoprotéine qui participe à la régulation de la production et la sécrétion de la FSH (follicle-stimulating hormone) via une boucle de rétrocontrôle négatif (Namwanje and Brown, 2016). Elle produit aussi l'ABP (androgen-binding protein) qui se lie avec une forte affinité à la testostérone et la dihydrotestostérone et permet de contrôler leur biodisponibilité dans les tubules séminifères. Contrairement aux cellules germinales, la cellules de Sertoli exprime le récepteur des androgènes ce qui suggère que l'effet de ces hormones au niveau des tubules séminifères est indirectement médié par la cellule de Sertoli (Petersen and Söder, 2006).

La cellule de Sertoli produit également des facteurs de croissance qui jouent un rôle important dans la modulation du développement des cellules germinales comme GDNF qui inhibe l'auto-renouvellement des cellules souches spermatogoniales, BMP4 qui stimule la prolifération et la différenciation de ces cellules et KL (KIT ligand) qui stimule un récepteur tyrosine kinase exprimé spécifiquement par les spermatogonies en différenciation (Rossi and Dolci, 2013).

6.1.2- La barrière hémato-testiculaire

Dans un contexte physiologique normal, le testicule assure une régulation étroite de l'environnement immunitaire notamment grâce à l'existence d'une barrière hémato-testiculaire (Mruk and Cheng, 2015a). En effet, Le testicule est considéré à l'instar du cerveau, l'œil et l'utérus gestationnel comme un site remarquable de privilège immun. Cette homéostasie immunitaire est nécessaire pour assurer une spermatogenèse fonctionnelle (Chen et al., 2016).

La barrière hémato-testiculaire est formée par les jonctions serrées étanches (zonula occludens) qu'établissent les cellules de Sertoli entre elles (Mruk and Cheng, 2015b). Cette barrière subdivise l'épithélium séminifère en deux compartiments : le compartiment basal et le compartiment ad-luminal **(Figure 10)**. Les cellules germinales sont disposées au sein de l'épithélium séminifère de manière très organisée avec les cellules immatures dans le compartiment basal et les cellules de plus en plus matures dans le compartiment ad-luminal (Gao et al., 2016).

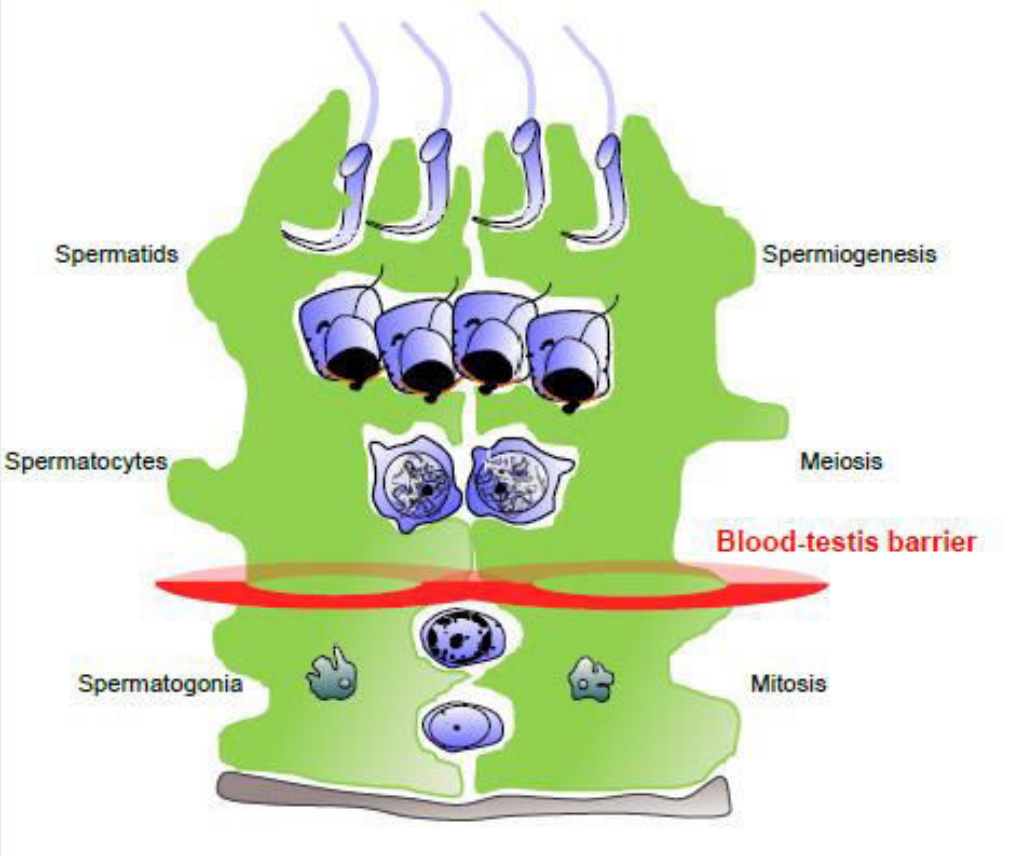


Figure 10. La barrière hémato-testiculaire (blood-testis barrier).
La barrière hémato-testiculaire, formée par les jonctions serrées entre les cellules de Sertoli adjacentes divise l'épithélium séminifère en deux compartiments, basal et ad-luminal. Cette compartimentalisation permet de créer un microenvironnement favorable à la méiose et la spermiogénèse au sein du compartiment ad-luminal (d'après O'Donnell et al., 2017).

7- Maturation et différenciation des cellules germinales

7.1- Les cellules germinales souches et les spermatogonies

Les cellules germinales sont des cellules capables d'entamer le processus de méiose. Elles permettent de véhiculer l'information génétique et épigénétique aux nouvelles générations. Au cours de l'organogenèse du testicule, les gonocytes, cellules dérivées des cellules germinales primordiales, entrent en phase G0 et restent quiescentes jusqu'à 6-12 mois après la naissance avant d'entamer leur différenciation en spermatogonies **(Figure 11)** (Mechlin and Kogan, 2014).

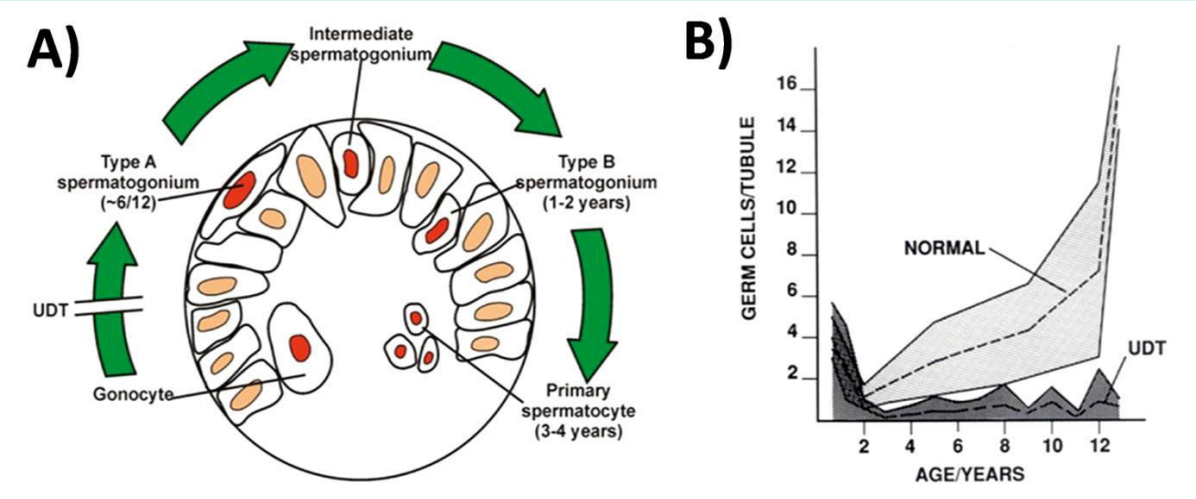


Figure 11. Développement des cellules germinales en post-natal chez l'humain.
A) Les gonocytes migrent du centre vers la périphérie des tubules séminifères autours du sixième mois après la naissance et se différencient en spermatogonies. Vers 3-4 ans, le centre des tubules est colonisé par les spermatocytes. UDT= Undescended testis. B) Evolution du nombre des cellules germinales dans les tubules des testicules normaux et les testicules non descendus (UDT) (d'après Hutson et al., 2013).

Les spermatogonies sont les cellules progénitrices de toutes les autres cellules germinales. Elles ont une double responsabilité : 1) la production des gamètes par méiose et 2) l'auto-renouvellement par mitoses pour assurer la production continue de spermatozoïdes tout au long de la vie reproductive de l'homme. Les spermatogonies sont localisées dans le compartiment basal et sont en contact étroit avec les cellules de Sertoli. Durant l'avancement des divisions mitotiques, les spermatogonies s'alignent en chaînes de différentes longueurs grâce aux ponts cytoplasmiques qui résultent des cytokinèses incomplètes (Boitani et al., 2016).

Ces cellules sont divisées en 3 sous-types en fonction de l'aspect de leur hétérochromatine : les spermatogonies Ad (dark), les spermatogonies Ap (pale) et les spermatogonies B (von Kopylow and Spiess, 2017). Les spermatogonies Ad représentent les cellules souches quiescentes alors que les spermatogonies Ap représentent la forme cellulaire active. Une des cellules filles Ap renouvelle le stock des spermatogonies et l'autre se transforme en spermatogonie B. C'est la phase de multiplication cellulaire. L'étape suivante est la phase d'accroissement durant laquelle les spermatogonies B se divisent par mitose puis augmentent de taille pour devenir des spermatocytes primaires.

7.2- La méiose ou division réductionnelle

La méiose est un processus spécialisé de division cellulaire. Ce processus est essentiel pour la reproduction sexuée chez les eucaryotes. Chez les organismes diploïdes, la méiose permet de produire des gamètes haploïdes et générer une diversité génétique au sein de la nouvelle génération. Cette diversité est secondaire à deux événements : la recombinaison homologue et la ségrégation aléatoire de ces chromosomes (Ohkura, 2015). Les défauts de ségrégation des chromosomes sont à l'origine d'aneuploïdies qui peuvent être responsables d'infertilité, de fausses couches ou de maladies congénitales (Hassold and Hunt, 2001). La méiose évolue en deux étapes : méiose I et méiose II. Ces deux étapes sont précédées par une seule phase de réplication de l'ADN. La méiose I est entamée par les spermatocytes primaires et permet la répartition des chromosomes homologues dans deux spermatocytes secondaires. Ces cellules vont directement rentrer en méiose II qui est analogue à une division mitotique et permet donc la ségrégation des chromatides sœurs.

7.2.1- Les spermatocytes primaires et la méiose I

Les spermatocytes primaires traversent la barrière hémato-testiculaire pour accéder au compartiment ad-luminal. Dans ce compartiment, ces cellules deviennent immunologiquement isolées et vont entamer la première division méiotique. Les spermatocytes primaires sont classés selon les stades de la prophase 1 (leptotène, zygotène, pachytène, diplotène et diacinèse) **(Figure 12)**.

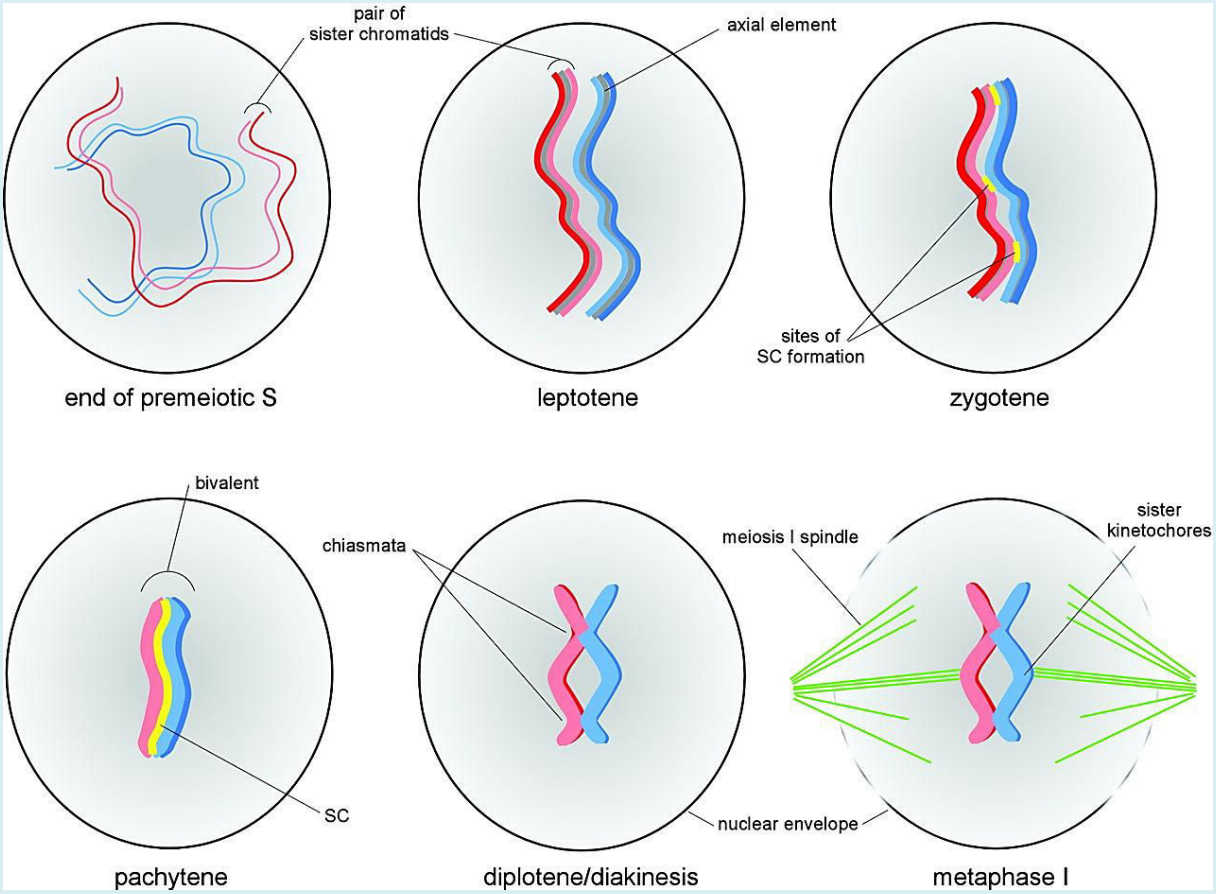


Figure 12. La prophase I.
Le début de la méiose-I est marqué par le stade de leptotène pendant lequel, la chromatine se condense pour former les chromosomes. Les chromosomes homologues s'apparient pendant le stade de zygotène grâce à la formation de complexes synaptonémaux. Les recombinaisons homologues ont lieu pendant le stade de pachytène. Durant les phases de diplotène et de diacinèse, on observe la dissociation du complexe synaptonémal et le début de disjonction des chromosomes homologues qui ne restent appariés qu'à certains points appelés chiasma. Après la rupture de l'enveloppe nucléaire et la formation du fuseau méiotique, les chromosomes homologues s'alignent sur la plaque équatoriale de la cellule (métaphase I) (d'après Page and Hawley, 2003).

Les chromosomes homologues s'apparient lors du stade zygotène grâce à un complexe multiprotéique : le complexe synaptonémal **(Figure 13)**. Le processus de recombinaison d'ADN (crossing-over) a lieu durant le stade de pachytènes (Lam and Keeney, 2014). Au stade diplotène, les complexes synaptonémaux commencent à se dissocier. Cependant, les chromosomes homologues conservent certains points d'appariement appelés chiasmas. La condensation maximale des chromosomes et la disparition de l'enveloppe nucléaire caractérisent la diacinèse qui marque la fin de la prophase I.

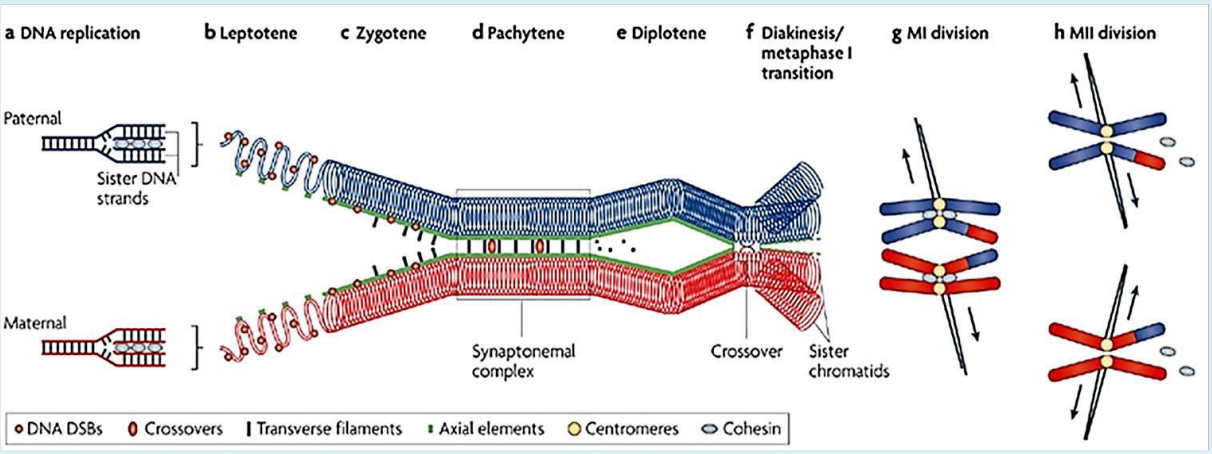


Figure 13. Assemblage et dissociation du complexe synaptonémal durant la prophase I.
(d'après Burgoyne et al., 2009).

7.2.2- Les spermatocytes secondaires et la méiose II

La division réductionnelle produit des cellules haploïdes appelées les spermatocytes secondaires. Après une très courte interphase sans réplication d'ADN, ces cellules se divisent pour produire des spermatides rondes. Ces cellules ne se divisent plus mais se différencient en spermatides allongées lors de la dernière étape appelée « spermiogénèse ».

7.3- La spermiogénèse

Au cours de cette phase, les spermatides rondes subissent 3 modifications majeures avant de se transformer en spermatides allongées : 1) la biogenèse de l'acrosome, 2) l'hypercompaction de la chromatine et 3) la formation du flagelle (Ramalho-Santos et al., 2002). Cette phase dure environ 23 jours chez l'homme.

7.3.1- Biogenèse de l'acrosome

L'acrosome est une vésicule apparentée à la famille des lysosomes. Il est localisé au niveau du pôle apical de la tête du spermatozoïde et recouvre jusqu'à deux tiers de la surface du noyau. L'acrosome a un pH acide et contient de nombreuses enzymes protéolytiques comme la proacrosine (Mao and Yang, 2013). Au cours de la réaction acrosomique, les enzymes acrosomales libérées permettent aux spermatozoïdes de pénétrer la

corona radiata et la zone pellucide entourant l'ovocyte pour permettre la fécondation **(Figure 14)** (Berruti, 2016).

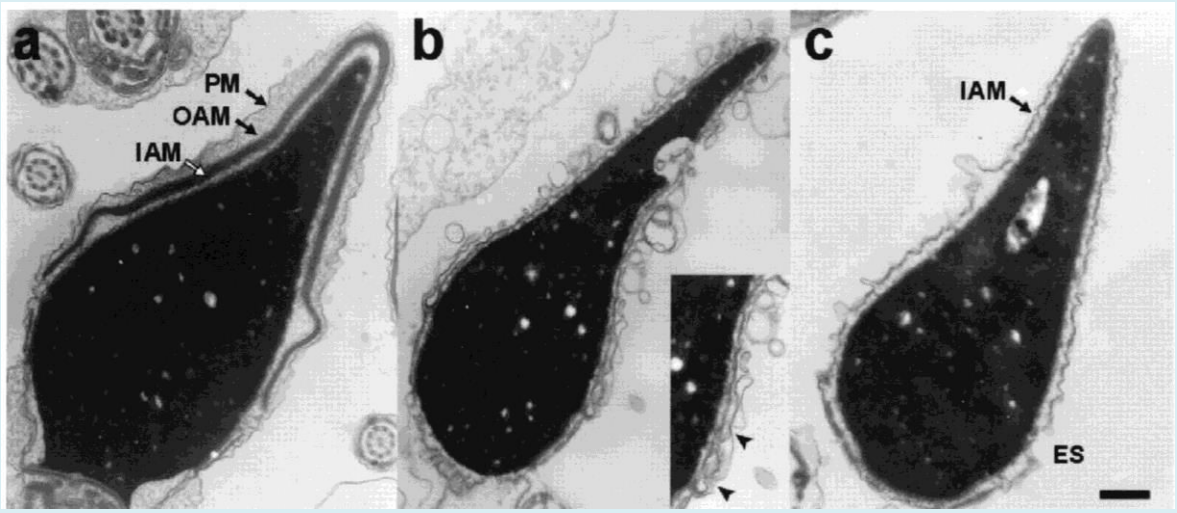


Figure 14. Schéma de l'acrosome et de ces différentes membranes avant et après la réaction acrosomique.
(A) acrosome intact, (B) au cours de la réaction acrosomique et (C) après la réaction acrosomique. La vésicule de l'acrosome est constituée de la membrane externe de l'acrosome (OAM) située sous la membrane plasmique (PM) et la membrane interne de l'acrosome (IAM) collée au noyau (d'après Michaut et al., 2000).

L'acrosome est ancré au noyau par l'intermédiaire d'une structure appelée acroplaxome et au cytosquelette par la thèque périnucléaire. Cette dernière structure contient la phospholipase C zeta (PLCZ) qui est responsable de la cascade d'activation de l'ovocyte (Escoffier et al., 2016). La biogenèse de l'acrosome passe par 3 phases principales : la phase de Golgi, la phase de l'acrosome et la phase de maturation **(Figure 15)** (Sedó et al., 2012).

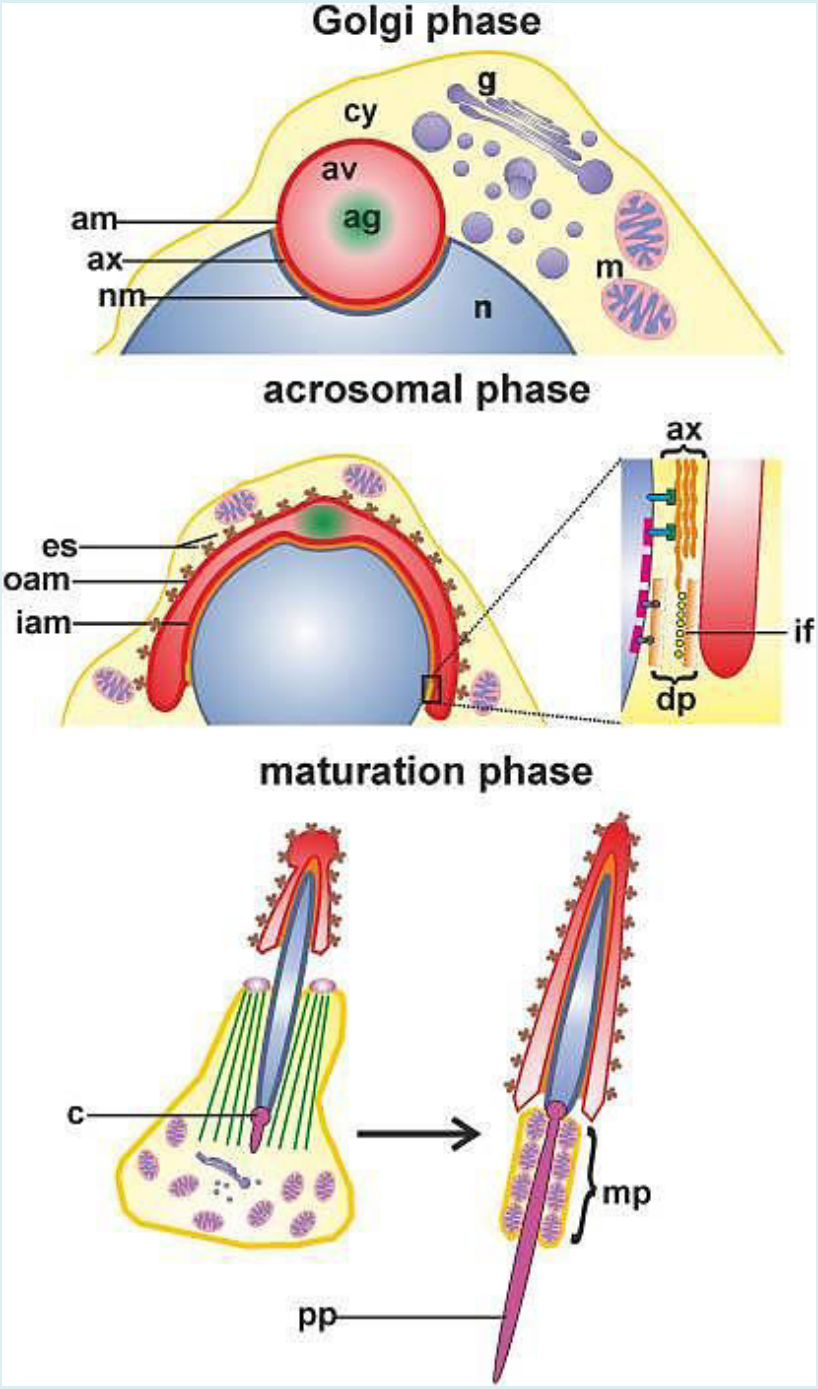


Figure 15. Présentation schématique des principales phases de la biogenèse de l'acrosome.
Abréviations : ag acrosomal granule, am acrosome membrane, av acrosome vesicle, ax acroplaxome, c centriole, cy cytoplasm, dp dense plaque, es apical ectoplasmic specialization, g Golgi apparatus, iam inner acrosomal membrane, if intermediate filaments, m mitochondria, mp midpiece, n spermatid nucleus, nm nuclear envelope, oam outer acrosomal membrane, pp principal piece (d'après Zakrzewski et al., 2017).

Durant la phase de Golgi, des vésicules proacrosomales bourgeonnent à partir des saccules de la face *trans* de l'appareil de Golgi qui est orientée vers le noyau.

Durant la phase de l'acrosome, les vésicules proacrosomales fusionnent entre elles sur l'acroplaxome pour former la vésicule acrosomale. L'acroplaxome est une structure transitoire qui recouvre la surface apicale du noyau. Elle est formée principalement par des microfilaments d'actine, de myosine et de kératine (Kierszenbaum et al., 2003).

Durant la phase de maturation, la manchette, une structure transitoire également, se forme autour de la région acrosomale à partir d'un réseau de microtubules organisés en faisceaux dirigés vers le pôle postérieur de la cellule (Pierre et al., 2012). La vésicule acrosomale commence alors à s'étaler sur le noyau formant une coiffe céphalique.

L'autre rôle de la manchette est d'assurer une fois la vésicule acrosomale formée, le transport actif de l'appareil de Golgi vers le flagelle en formation **(Figure 16)** (Zhou et al., 2009).

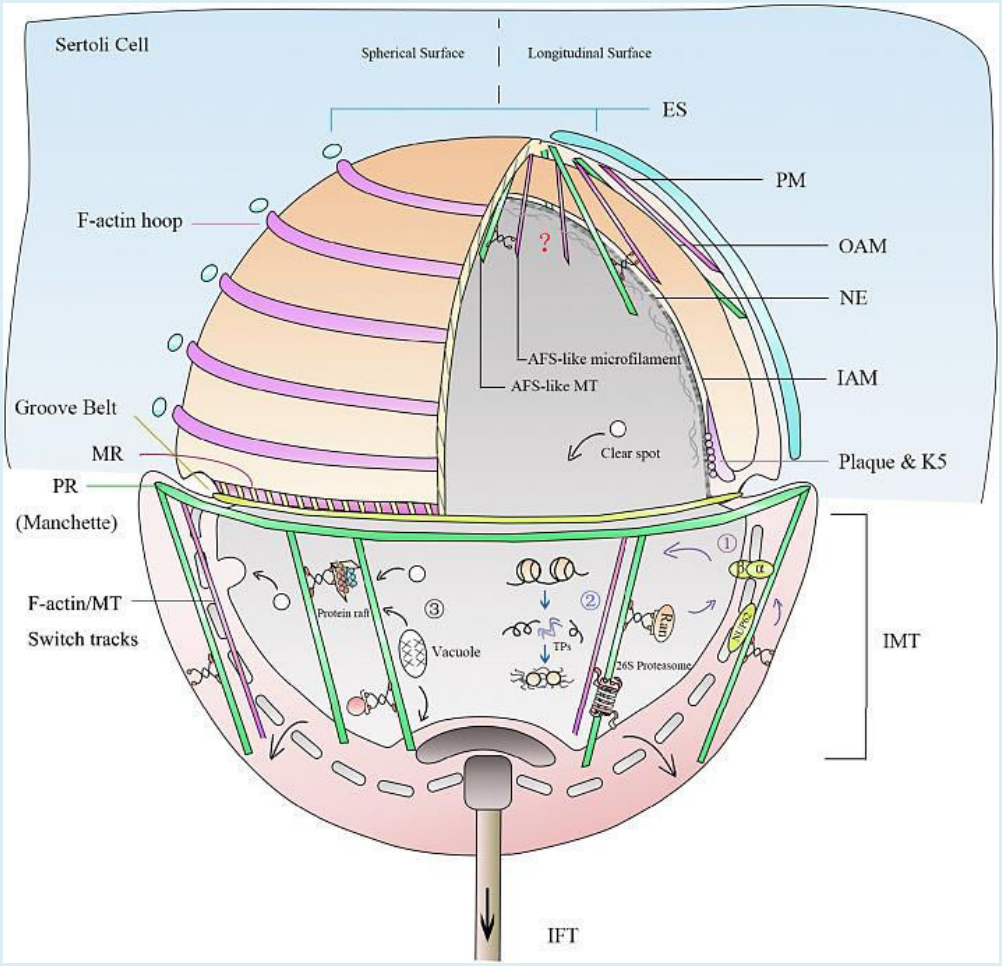


Figure 16. Organisation spatiale de la manchette dans la spermatide en élongation.
La manchette s'étend de l'anneau périnucléaire et recouvre le pôle postérieur du noyau. Cette structure va s'étendre progressivement au cours de l'élongation de la spermatide et étirer l'acrosome et le noyau pour donner la forme finale de la tête du spermatozoïde. Elle va aussi servir de rail pour le transport de vésicules et de protéines vers le pôle postérieur de la cellules pour l'assemblage et l'élongation du flagelle (d'après Wei and Yang, 2018).

7.3.2- Hypercompaction de la chromatine

Au cours de sa maturation, le noyau spermatique va subir une restructuration interne globale. Cette restructuration inclut une condensation progressive de la chromatine avec remplacement des histones riches en lysine, par des protéines nucléaires plus basiques : les protéines de transition, TNP1 et TNP2. Ces dernières sont par la suite remplacées par les protamines **(Figure 17)** (Bao and Bedford, 2016). Chez l'homme, deux types de protamines existent : PRM1 et PRM2 (Balhorn, 2007). Ces deux variants sont exprimés dans le spermatozoïde en quantité équimolaire (Zhang et al., 2006). Les protamines présentent plusieurs domaines de liaison à l'ADN appelés « enchroaching domains » riches en arginine (acide aminé chargé positivement), qui se lie fortement au phosphosquelette de l'ADN (chargé négativement) (Oliva, 2006). Les protamines sont aussi riches en cystéine. La formation de ponts disulfures intra et inter-moléculaires entre ces résidus favorise la compaction et la stabilisation de la chromatine (Ahmadi and Ng, 1999).

Ce réarrangement de la chromatine est nécessaire pour protéger le matériel génétique du gamète male contre les agressions physiques et chimiques durant son transit dans les voies génitales mâles et femelles. La réduction de la taille du noyau qui résulte de la condensation de sa chromatine rend le spermatozoïde plus hydrodynamique et favorise sa mobilité.

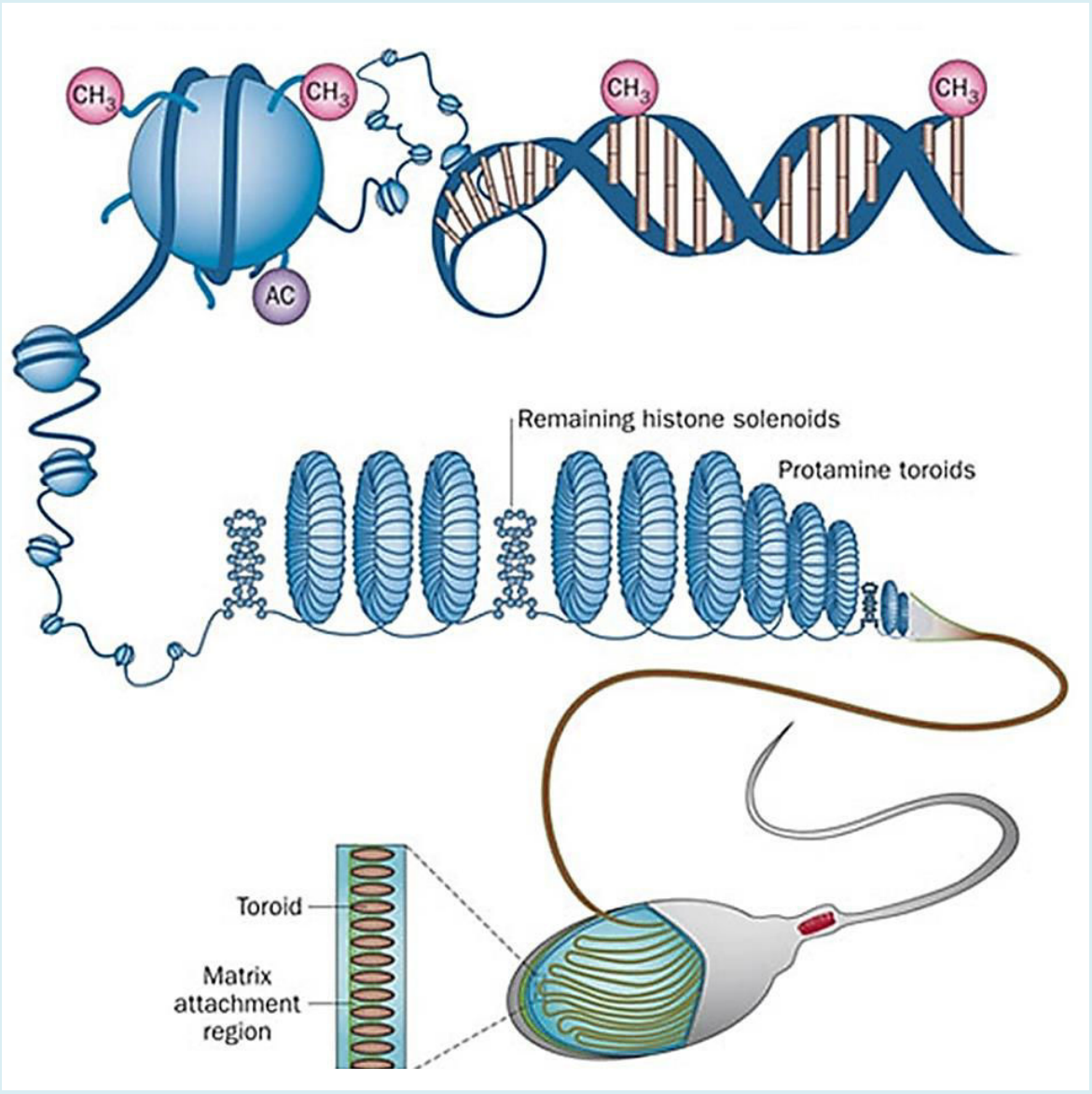


Figure 17. Remplacement des histones par les protamines et hypercompaction de la chromatine spermatique.
Avant le stade de la spermiogénèse, l'ADN génomique est enroulé autour des histones sous forme de nucléosomes puis ces derniers s'associent pour donner des structures « solénoïdes ». A cours de la spermiogénèse, les protamines remplacent les histones et forment des structures « toroïdes ». Approximativement, 85 % des histones sont remplacées par les protamines pendant la spermiogénèse chez l'Homme (d'après Schagdarsurengin et al., 2012).

7.3.3- Formation du flagelle

Le flagelle est responsable de la mobilité du spermatozoïde. Il commence à se développer à un stade précoce de la spermiogénèse. Son élongation se fait à partir du centriole distal du centrosome qui a migré vers le pôle postérieur de la cellule **(Figure 18)**. Le flagelle est subdivisé en trois segments : la pièce intermédiaire, la pièce principale et la pièce terminale.

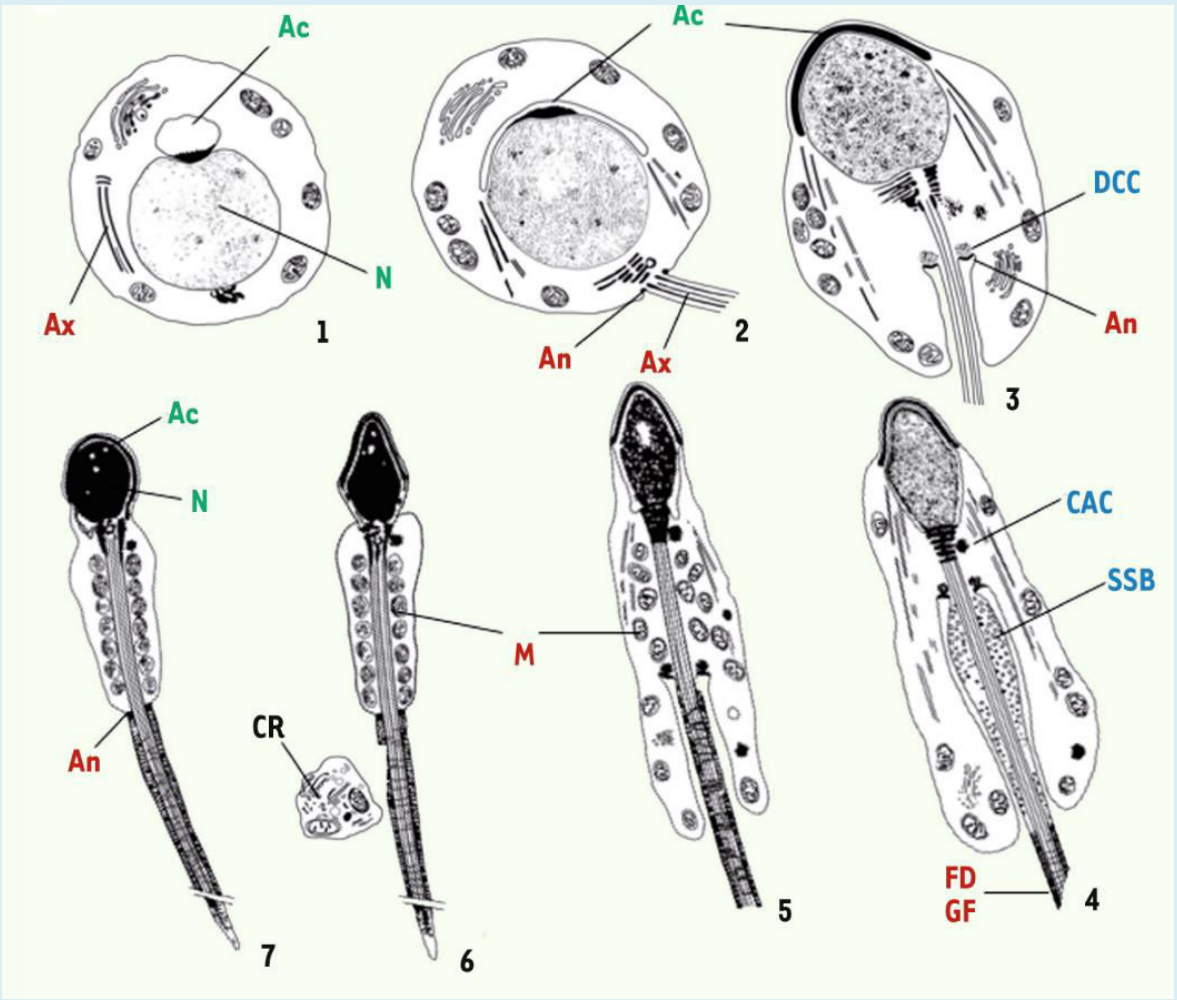


Figure 18. Formation du flagelle au cours de la spermiogenèse.
Le flagelle est formé par l'assemblage de l'axonème (structure microtubulaire) et par la mise en place des structures périaxonémales (fibres denses externes, gaine fibreuse, annulus et gaine de mitochondries). Ce processus est accompagné de l'apparition de structures transitoires (en bleu) (DCC : dérivé du corps chromatoïde, CAC : corps associé au centriole, SSB : spindle shaped body). Ac : acrosome ; An : annulus ; Ax : axonème ; CR : corps résiduel ; FD : fibres denses ; GF : gaine fibreuse ; M : mitochondries ; N : noyau (d'après Escalier and Touré, 2012).

7.3.3.1- L'axonème

L'axonème, en position centrale, parcourt toute la longueur du flagelle (50µm chez l'homme). L'axonème est une structure pivot du flagelle, il est composé d'une paire centrale de microtubules entourée de 9 doublets périphériques (A+B) connectés entre eux par des ponts de nexine. Les microtubules A portent les bras de dynéines internes et externes et sont reliés au complexe central par des ponts radiaires **(Figure 19)** (Nicastro et al., 2006).

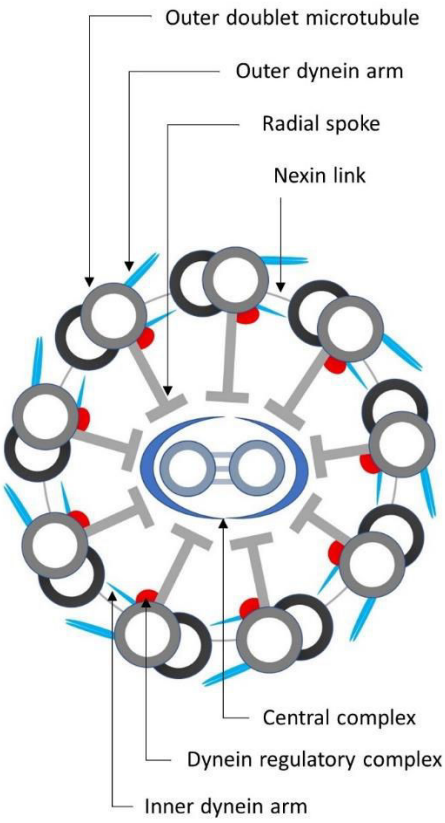


Figure 19. L'axonème du flagelle spermatique.
L'axonème est composé de 9 doublets de microtubules périphériques reliés entre eux par des liens de nexine. Les doublets périphériques communiquent via des ponts radiaires avec un complexe central constitué d'une paire de microtubules entourée par une gaine périphérique. Les complexes de régulation de dynéines sont localisés entre la base des ponts radiaires et les bras internes de dynéines (d'après *Z-E Kherraf, GETI*).

A. Les dynéines axonémales

Les dynéines ont été initialement identifiées et purifiées à partir des axonèmes ciliaires et flagellaires d'eucaryotes. Depuis, une autre famille de dynéines a été identifiée

dans le cytoplasme. Ces dynéines cytoplasmiques se fixent sur les microtubules et grâce à leur activité ATPasique transportent les organelles, les vésicules et les particules comme les protéines et les ARNs vers le centre de la cellule (transport rétrograde) (Sweeney and Holzbaur, 2018). Les dynéines axonémales forment les bras internes de l'axonème. Ces bras sont ancrés dans les microtubules A et dirigés vers les microtubules B du doublet adjacent. Chaque bras est composé de l'association de plusieurs unités de dynéines classées en fonction de leurs masses moléculaires : chaînes légères (≃ 30-80 kDa), chaînes intermédiaires (≃100-160 kDa) et chaînes lourdes (≃500 kDa). Les dynéines ont été largement étudiées chez *Clamydomonas* qui possède 16 gènes de chaînes lourdes (*DHC, dynein heavy chain*). Les mutants de cet organisme caractérisés par une mobilité défectueuse sont typiquement nommés *pf*-mutants (parlyzed flagella). Chez l'Homme, il existe également 16 gènes qui codent pour des DHCs dont 14 sont spécifiques de l'axonème (Roberts et al., 2013). Les chaînes lourdes sont constituées d'une tête globulaire pourvue de 6 domaines moteurs ATPases (*AAA1-6, ATPase associated with cellular activities 1-6*), de deux hélices formées par des domaines *coiled-coil* et d'une unité de liaison aux microtubules B **(Figure 20)**.

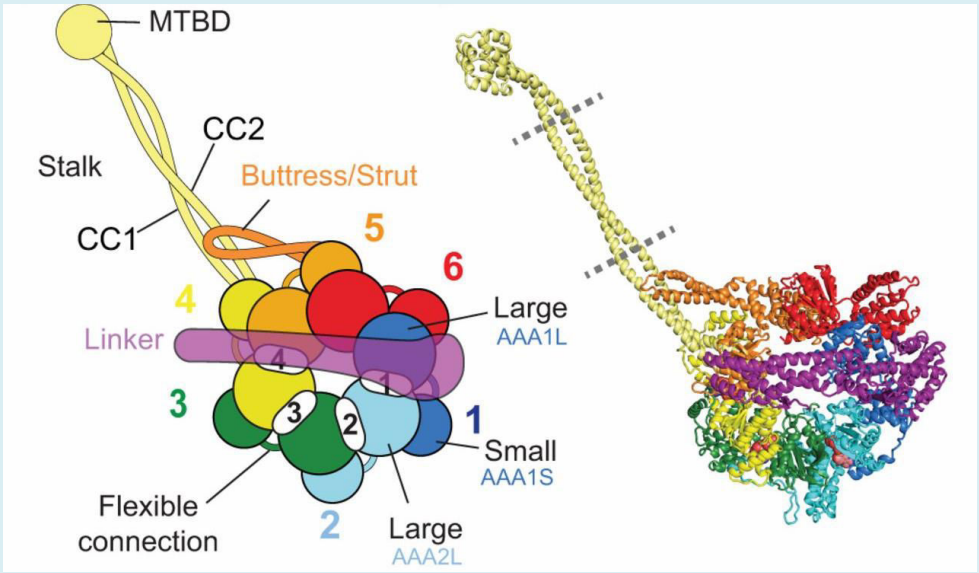


Figure 20. Architecture d'une chaîne lourde de dynéine.
Illustration schématique des domaines structuraux et fonctionnels d'une chaîne lourde de dynéine. CC : coiled-coil domain ; MTBD : microtubule binding domain (d'après Carter, 2013).

Les chaînes légères et intermédiaires jouent un rôle dans l'ancrage et la régulation de l'activité motrice des bras internes et externes. Ces bras sont distribués le long de

l'axonème sous forme d'unités structurales répétées tous les 96 nm. Dans ces unités on dénombre 4 bras externes identiques espacés de 24 nm et 7 bras internes différents. Les bras externes sont composés de deux chaînes lourdes chez les vertébrés et de trois chaînes chez les protistes (α, β et γ) **(Figure 21)**. Ces bras jouent un rôle dans la modulation de la fréquence de battement des cils et des flagelles. Les bras internes se présentent comme des hétérodimères (fα et fβ) ou des monomères (a, b, c, d, e, g) de chaînes lourdes de dynéines. Chaque bras interne exerce une activité distincte dans la modulation de l'amplitude et de la forme des ondes générées par l'activité des bras externes.

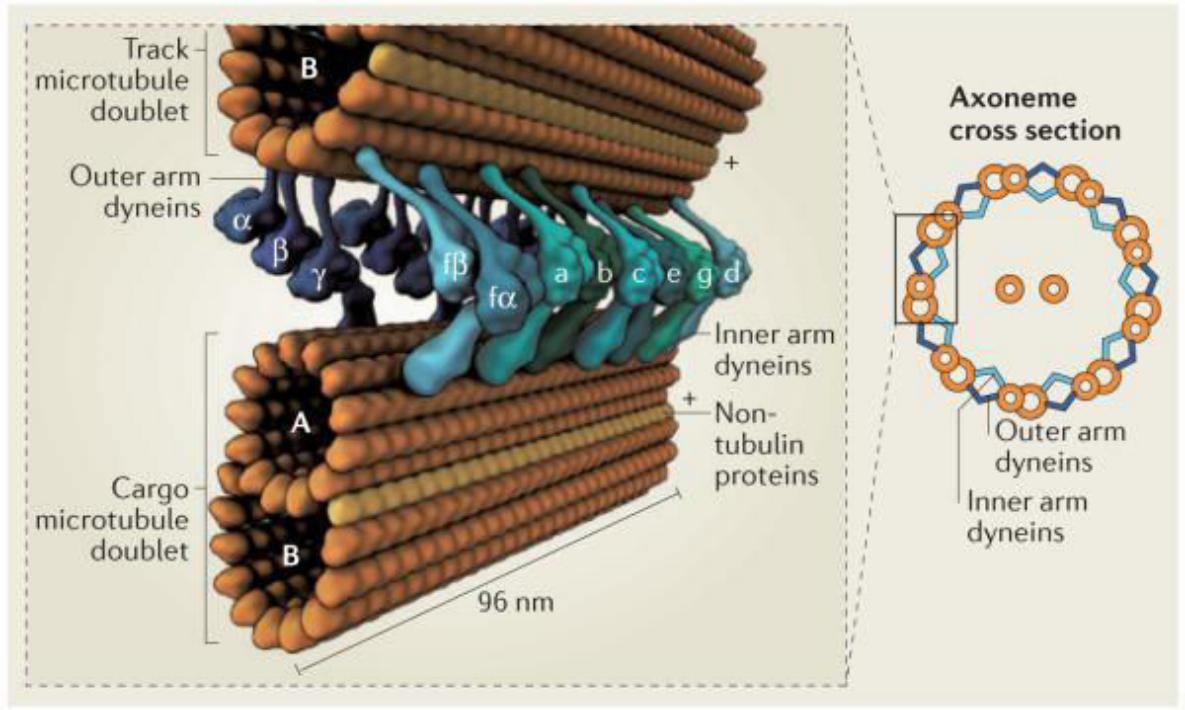


Figure 21. Les bras internes et externes de dynéines.
L'axonème est composé d'un squelette de microtubules de type 9+2 correspondant à 9 doublets périphériques et une paire centrale de microtubules singuliers. Dans les axonèmes des cils mobiles et flagelles, il existe des complexes de dynéines organisés en bras internes et externes. Chez Chlamydomonas, on observe 4 bras externes formés chacun par trois chaînes lourdes de dynéine (α, β et γ) et 7 isoformes différents de bras internes formés chacun par une ou deux chaînes lourdes de dynéine (d'après Roberts et al., 2013).

B. Les ponts radiaires

Les ponts radiaires (*RSs, radial spokes*) sont des complexes d'au moins 23 protéines qui participent ensemble à la transduction des signaux chimio-mécaniques entre le complexe central et les microtubules périphériques. Les RSs prennent la forme de la lettre T

et sont constitués d'une tige ancrée dans les microtubules A des doublets périphériques et d'une tête orthogonale orientée vers le complexe central **(Figure 22)**. Alors que les flagelles de l'algue verte *Chlamydomonas* comptent seulement deux RSs de taille complète (RS1 et 2) chaque 96 nm, les cils et les flagelles d'autres organismes comme le protiste *Tétrahyména*, les oursins et les mammifères en contiennent 3 (RS1, 2 et 3) (Pigino et al., 2011).

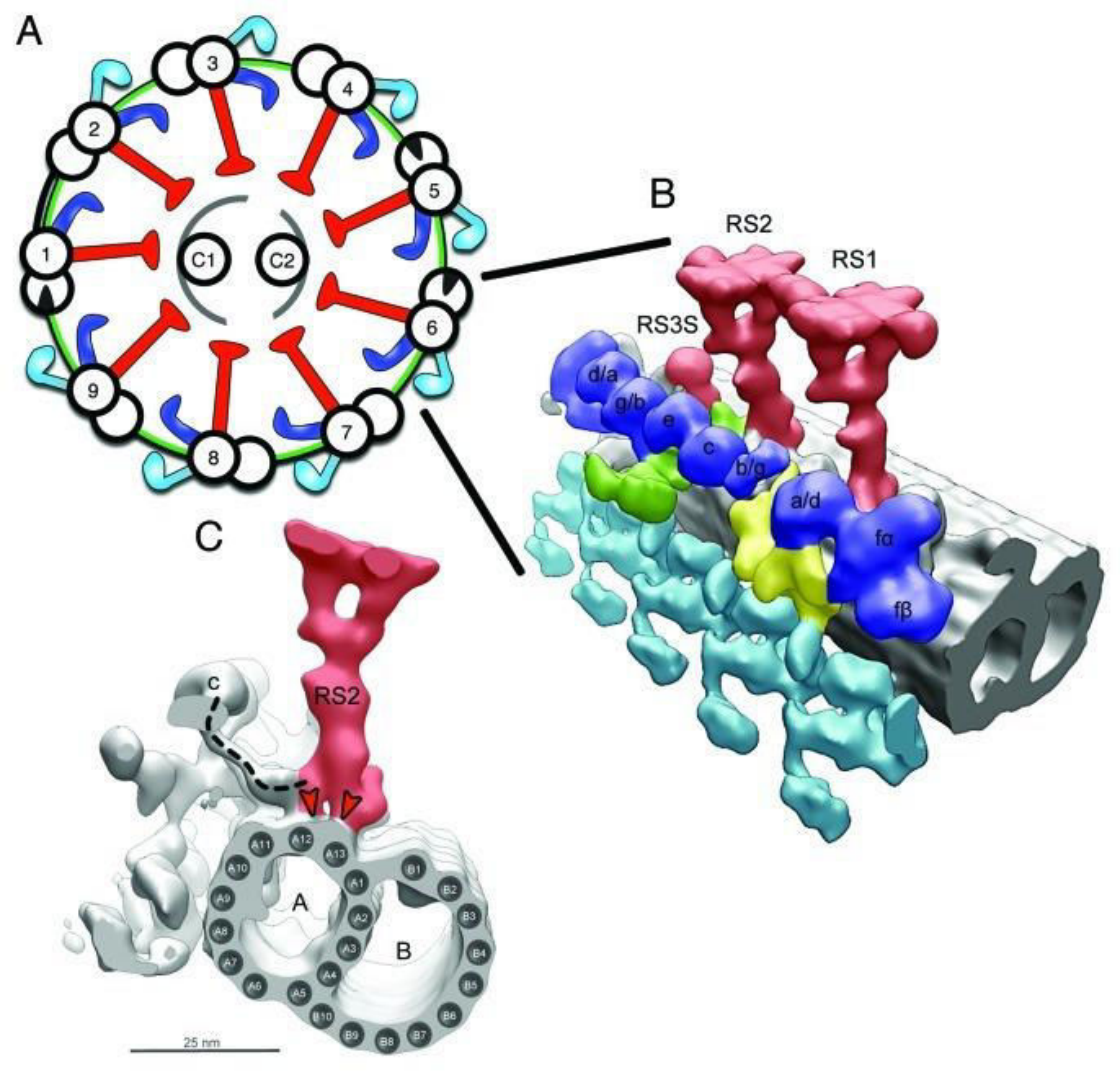


Figure 22. Localisation des ponts radiaires dans l'axonème de *Chlamydomonas*.
(A) Présentation schématique des ponts radiaires (rouge) sur une coupe transversale de l'axonème. (B) Reconstruction tomographiques d'une unité fondamentale de l'axonème avec les trois ponts radiaires (*RS, radial spoke*) : RS1, RS2 et RS3S (rouge), les bras externes de dynéines (bleu), les bras internes de dynéines (turquoise), le complexe N-DRC (vert). (C) Vue latérale du RS2 montrant sa connexion avec la queue terminale du bras interne de dynéine c (pointillés noirs). A, microtubule A ; B, microtubule B (d'après Pigino and Ishikawa, 2012).

C. Le CSC et le N-DRC

Le CSC (calmoduline ans spoke-associated complex) a récemment été identifié comme un complexe associant trois composants majeurs : la base du RS2, le N-DRC (Nexin-dynein regulatory complex) et la base du RS3 (Heuser et al., 2012). Des études de co-immunoprécipitation chez *Chlamydomonas* utilisant un anticorps spécifique de la calmoduline (CaM) à partir d'un lysat de flagelles a permis de co-précipiter trois autres protéines (CaM-IP2,/FAP91, CaM-IP3/FAP61 et CaM-IP4/FAP251) **(Tableau 1)** (Dymek and Smith, 2007).

Name	Predicted molecular mass	Apparent molecular mass	Similarities	Flagellar proteome
CaM-IP2	ND	183kD	ATT-1, AKAP-binding protein	FAP91
CaM-IP3	118kD	140kD	Pyridine-disulfide oxidoreductase domain	FAP61
CaM-IP4	97kD	100kD	WD repeats	FAP251

Tableau 1. Caractéristiques des protéines co-précipitées avec la calmoduline.
(D'après Dymek and Smith, 2007).

Le CSC intervient dans l'ancrage des RS dans les microtubules et leur stabilisation. Il intervient également dans la régulation de la mobilité du flagelle via son interaction, probablement calcium-dépendante, avec les dynéines. En effet, la régulation de la mobilité ciliaire et flagellaire est régulée par des changements de concentration intracellulaire de calcium. Cette régulation pourrait inclure le changement de fréquence, de forme et de direction des oscillations axonémales. L'étude des knock-out de *Tétrahyména* protéines pour *FAP61* et *FAP251* a permis de montrer l'importance de ces gènes dans la mobilité ciliaire et de préciser la localisation des protéines associées au niveau de la base du RS3 **(Figure 23)** (Urbanska et al., 2015).

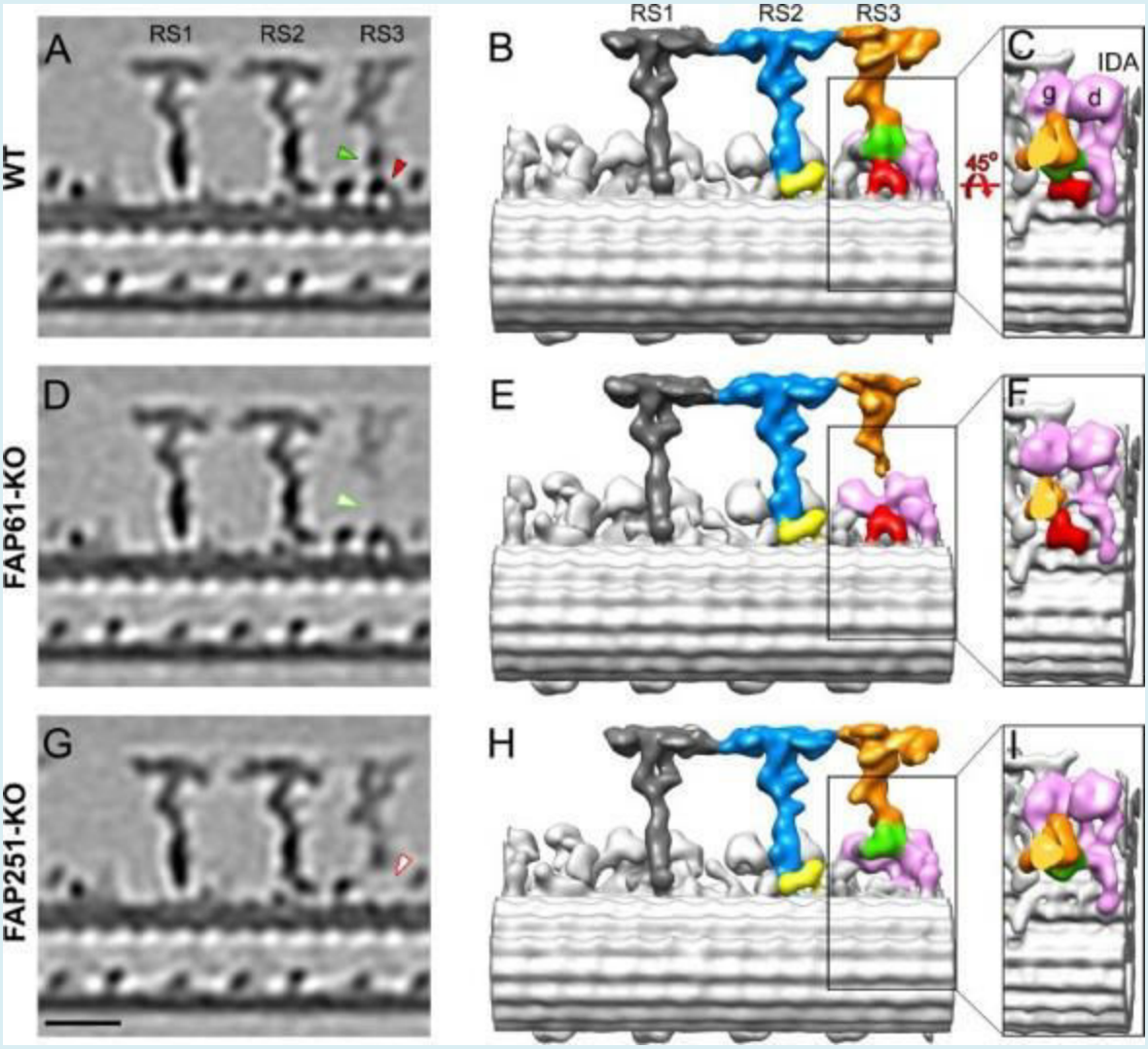


Figure 23. Impact du déficit en FAP61 et 251 sur la structure axonémale de Tétrahyména.
(A, C, D) représentent des coupes tomographiques au niveau des unités axonémales fondamentales chez le wild-type (WT) et les knock-out (KO) pour FAP61 et 251. (B-I) représentent la structure 3D de ces unités. On observe chez le FAP61-KO la disparition de la structure en vert localisée au niveau de la base du RS3 (orange). Chez le FAP251-KO, c'est la structure sous-jacente en arc rouge qui disparaît (d'après Urbanska et al., 2015).

Le N-DRC (Nexin-Dynein Regulatory Complex) chez *Chlamydomonas* forme une structure de 1.5 MDa qui s'étend de la base du RS2 jusqu'à la base du RS3 **(Figure 24)** (Heuser et al., 2009). La nexine forme des ponts entre les doublets de microtubules périphériques adjacents tous les 96 nm. Ces liens permettent de limiter le glissement excessif des microtubules provoqué par l'activité motrice des bras internes et externes de dynéines. Chez la plupart des espèces, le N-DRC contient au minimum 11 sous-unités. Cependant, la

fonction de chaque sous-unité est actuellement inconnue. Des mutations dans trois unités (DRC1/CCDC164, DRC2/CCDC65 et DRC4/GAS8) ont été récemment associées aux dyskinésies ciliaires primitives (Bower et al., 2018).

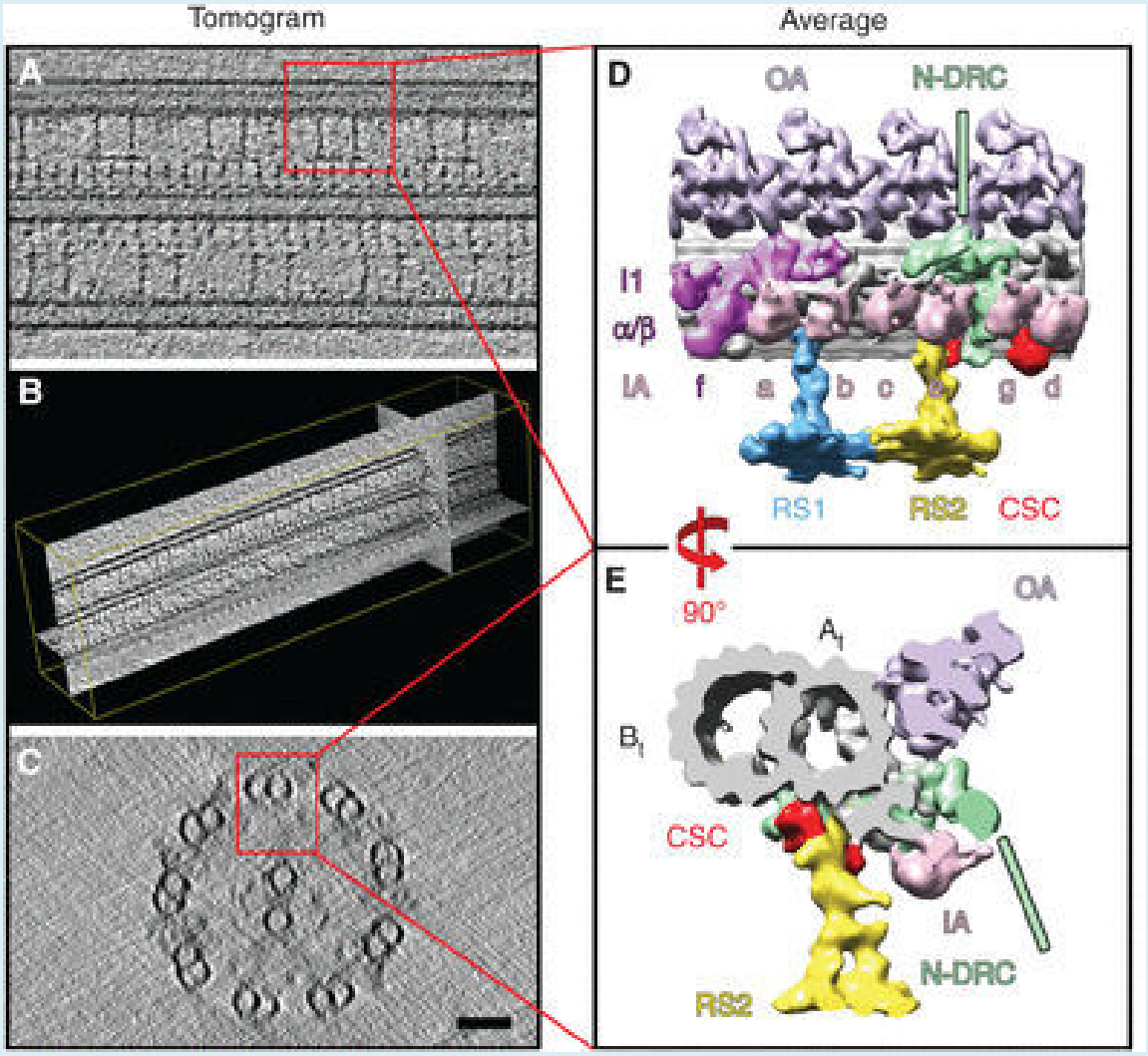


Figure 24. Localisation du N-DRC.
Le N-DRC (vert) se prolonge entre les bases des ponts radiaires 2 et 3 (rouge) et entre les bras internes et externes de dynéines (d'après Viswanadha et al., 2017).

D. Le complexe central

Le complexe central est, comme son nom l'indique, est localisé dans le centre de l'axonème de type « 9+2 », caractéristiques des cils mobiles et des flagelles. Il est composé de deux microtubules singuliers (C1 et C2) chacun associé à des projections asymétriques dirigées vers les ponts radiaires et les doublets de microtubules périphériques **(Figure 25)**. Alors que la perte d'une de ces projections pourrait causer une diminution significative de la

mobilité ciliaire ou flagellaire, la perte de l'un des microtubules centraux cause un phénotype plus sévère de paralysie ciliaire ou flagellaire. Ceci suggère le rôle important que joue le complexe central dans la régulation de la mobilité de ces structures. Cependant, le mécanisme précis de son activité reste inconnu (Carbajal-González et al., 2013).

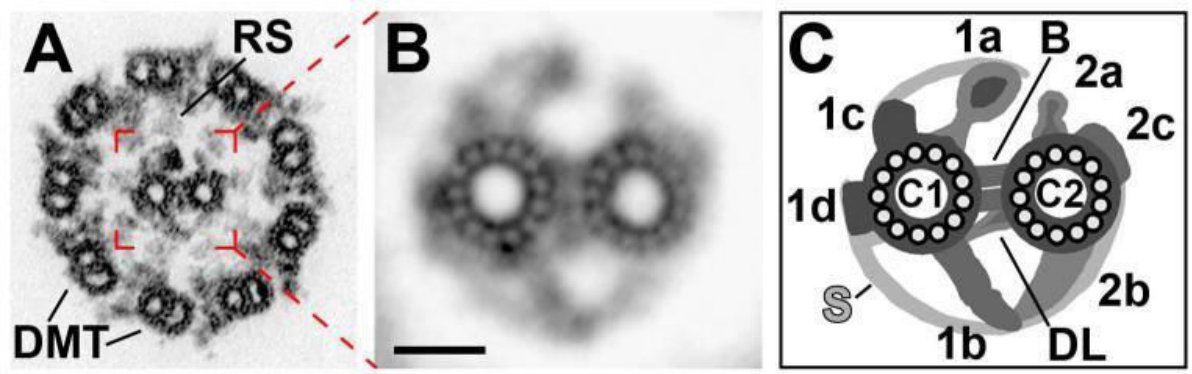


Figure 25. Organisation du complexe central.
(A-B) Images en microscopie électronique. (C) Représentation schématique de la paire centrale composée des deux microtubules C1 et C2 et ses projections dirigées vers les ponts radiaires. DMT= doublets de microtubules. Barre d'échelle en B = 20 nm (d'après Carbajal-González et al., 2013).

7.3.3.2- Les structures para-axonémales

Des structures accessoires existent entre l'axonème et la membrane plasmique et offrent un support mécanique supplémentaire pour le flagelle. La pièce intermédiaire comporte une gaine mitochondriale entourant 9 fibres denses externes appariées aux 9 doublets de microtubules périphériques. On pense que ces fibres maintiennent la structure élastique passive du flagelle et la protègent contre les forces de cisaillement pendant le transit épididymaire des spermatozoïdes et leur éjaculation (Baltz et al., 1990). Les fibres denses n'ont pas la même longueur Les fibres associées aux doublets 1, 5 et 6 sont les plus longs et parcourent jusqu'à trois quarts de la longueur du flagelle et les fibres associées aux doublets 3 et 8 sont les plus courts et se terminent à la jonction des pièces intermédiaire et principale **(Figure 26)**. A partir de la pièce principale, une gaine fibreuse se met en place autour des fibres denses et remplace la gaine mitochondriale. La gaine fibreuse va s'amincir progressivement jusqu'à l'extrémité distale du flagelle qui représente la pièce terminale. La gaine fibreuse et les fibres denses offrent aussi des sites de fixation pour une multitude d'enzymes impliquées dans le métabolisme et les voies de signalisation spermatique (Lindemann and Lesich, 2016).

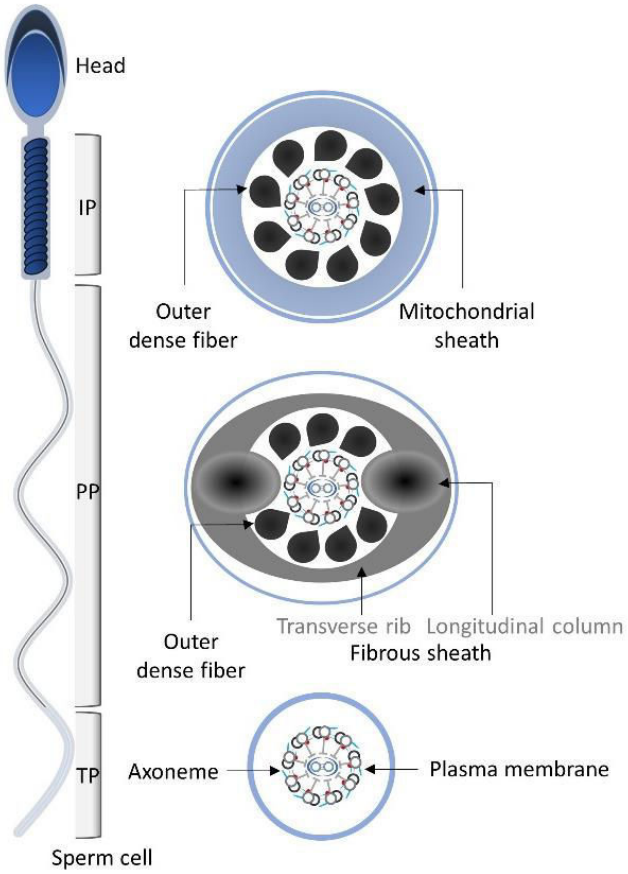


Figure 26. Présentation schématique de l'ultrastructure du flagelle spermatique.
*(*D'après *Z-E Kherraf, GETI)*.

8- La spermiation

La spermiation est un processus moléculaire complexe pendant lequel les spermatides allongées achèvent leur différenciation et se détachent de l'épithélium séminifère sous forme de spermatozoïdes. La spermiation est initiée au stade spermatogénique II chez l'homme (VII-VIII chez la souris) par la transition des spermatides allongées depuis les cryptes profondes de l'épithélium séminifères vers la lumière tubulaire (O'Donnell, 2014). Cette transition est facilitée par le réseau microtubulaire du cytosquelette de la cellule de Sertoli et les plaques jonctionnelles SEs (spécialisations ectodermiques) que cette cellule établit avec les spermatides rondes en développement.

Une fois les spermatides allongées arrivées au contact de la lumière des tubules séminifères, une structure transitoire appelée le complexe tubulobulbaire (TBCs, tubulobulbar complex) apparait dans le cytoplasme de la cellule de Sertoli au contact de la spermatide (Vogl et al., 2014). Ce complexe est formé par un réseau de filaments d'actine juxtaposés entre la membrane cytoplasmique et le réticulum endoplasmique de la cellule de Sertoli. Ce complexe participe au remodelage de la membrane plasmique de la tête spermatique et permet l'internalisation des jonctions apicales entre la cellule de Sertoli et la spermatide (Kusumi et al., 2007).

A la fin de la spermiation, l'excès cytoplasmique reste ancré dans l'épithélium séminifère avant d'être phagocyté par la cellule de Sertoli et la spermatide se détache de l'épithélium séminifère pour se retrouver libre dans la lumière du tubule sous forme de spermatozoïde **(Figure 27)** (O'Donnell et al., 2011). Ce processus est sensible aux facteurs hormonaux qui régulent la spermatogenèse (O'Shaughnessy, 2014). En effet, il a été observé que l'action de la FSH et des androgènes au niveau des cellules de Sertoli module l'expression de miARNs qui agissent de manière coordonnée pour réguler l'adhésion cellulaire et la spermiation (Nicholls et al., 2011).

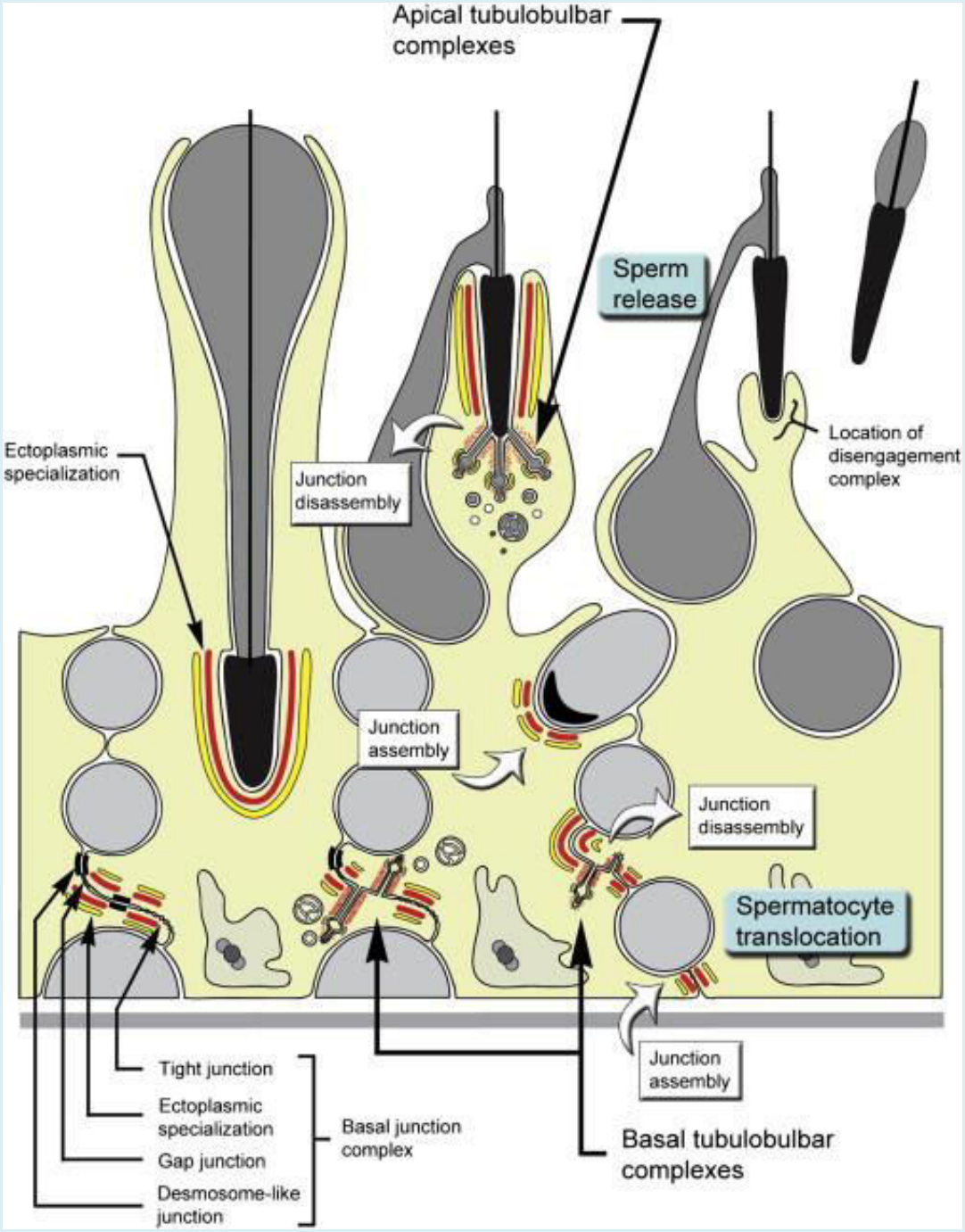


Figure 27. Présentation schématique du processus de spermiation.
(D'après Vogl et al., 2013).

9- La maturation épididymaire des spermatozoïdes

Bien que les spermatozoïdes qui quittent les tubules séminifères semblent être complétement différenciés, leur séjour dans l'épididyme est nécessaire pour leur maturation fonctionnelle (Mawhinney Michael and Mariotti Angelo, 2012). L'épididyme est divisé anatomiquement en trois segments successifs : un segment proximal (tête) connecté au testicule, un segment intermédiaire (corps) et un segment distal (queue) connecté au canal déférent **(Figure 28)**. L'épididyme permet le transport, la concentration et l'immunoprotection des spermatozoïdes. Il joue également un rôle de réservoir spermatique (Belleannée et al., 2012).

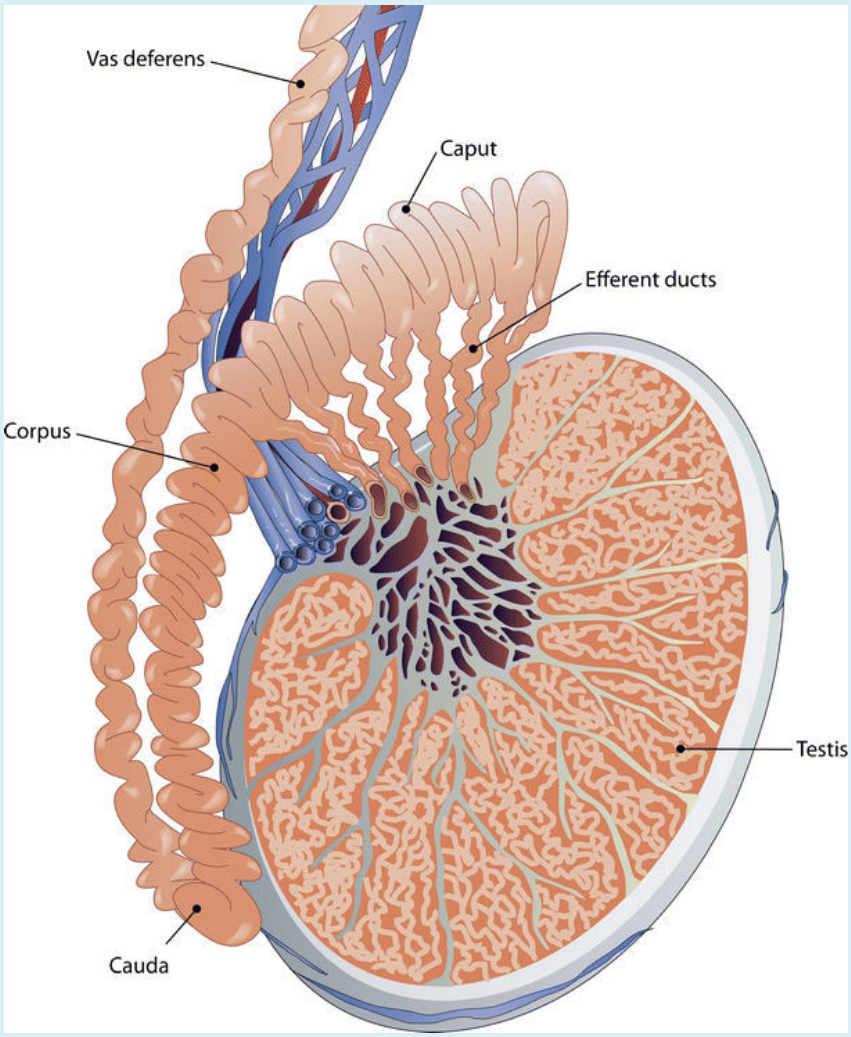


Figure 28. Présentation schématique de l'organisation de l'épididyme chez l'homme.
(D'après Belleannée et al., 2012).

A la différence des spermatozoïdes de la queue, les spermatozoïdes testiculaires et ceux du segment proximal de l'épididyme ne manifestent pas une mobilité progressive. Dans ce contexte, il a été observé que la lyse des membranes spermatiques par un détergent comme le Triton X-100, suivie d'un traitement avec l'ATP et l'AMPc rend ces spermatozoïdes mobiles. Cependant, la mobilité acquise par ce traitement est associée à des oscillations flagellaires différentes de celle observées chez les spermatozoïdes matures (Vadnais et al., 2013). Ces observations montrent que la machinerie moléculaire du flagelle est complétement fonctionnelle à la sortie du spermatozoïde du testicule et suggèrent l'implication de l'épididyme dans l'induction de modifications biochimiques permettant de contrôler et de moduler la mobilité spermatique. Il a été observé un changement progressif de la composition biochimique des spermatozoïdes au cours de leur transit dans l'épididyme (Skerget et al., 2015). Cependant, le mécanisme de ce changement n'est pas encore complétement élucidé. Plusieurs hypothèses ont été proposées pour tenter de l'expliquer : 1) l'absorption passive des molécules sécrétées par l'épididyme, 2) le transfert moléculaire via des exosomes d'origine épididymaire et 3) le transfert intercellulaire direct de ces molécules entre les cellules épithéliales de l'épididyme et les spermatozoïdes (Gervasi M. G. and Visconti P. E., 2017).

Outre, l'acquisition d'une mobilité progressive, la maturation épididymaire des spermatozoïdes leur permet d'acquérir un potentiel fécondant (Dacheux and Dacheux, 2014). La modification du glycocalyx et de la composition lipidique de la membrane spermatique, les modifications post-traductionnelles des protéines acquises durant la spermatogenèse et l'intégration de nouvelles protéines épididymaires sont responsables de l'acquisition de ce potentiel (Sullivan and Mieusset, 2016).

10- La régulation hormonale de la spermatogenèse

Le développement et le maintien de la spermatogenèse est dépendant du contrôle hormonal et en particulier, de l'action des gonadotrophines hypophysaires LH (luteinising hormone) et FSH (follicle stimulating hormone). La synthèse est la sécrétion de ces hormones est sous le contrôle direct de la gonadolibérine hypothalamique (GnRH, gonadotropin-realising hormone). La LH (luteinising hormone) régule la production de la testostérone par les cellules de Leydig et la FSH (follicle-stimulating hormone) agit sur les cellules de Sertoli qui soutiennent les cellules germinales en développement. A leur tour, ces cellules sécrètent des hormones qui participent à la régulation de l'activité sécrétoire des centres hypothalamo-hypophysaires via des boucles de rétrocontrôle (Figure 29).

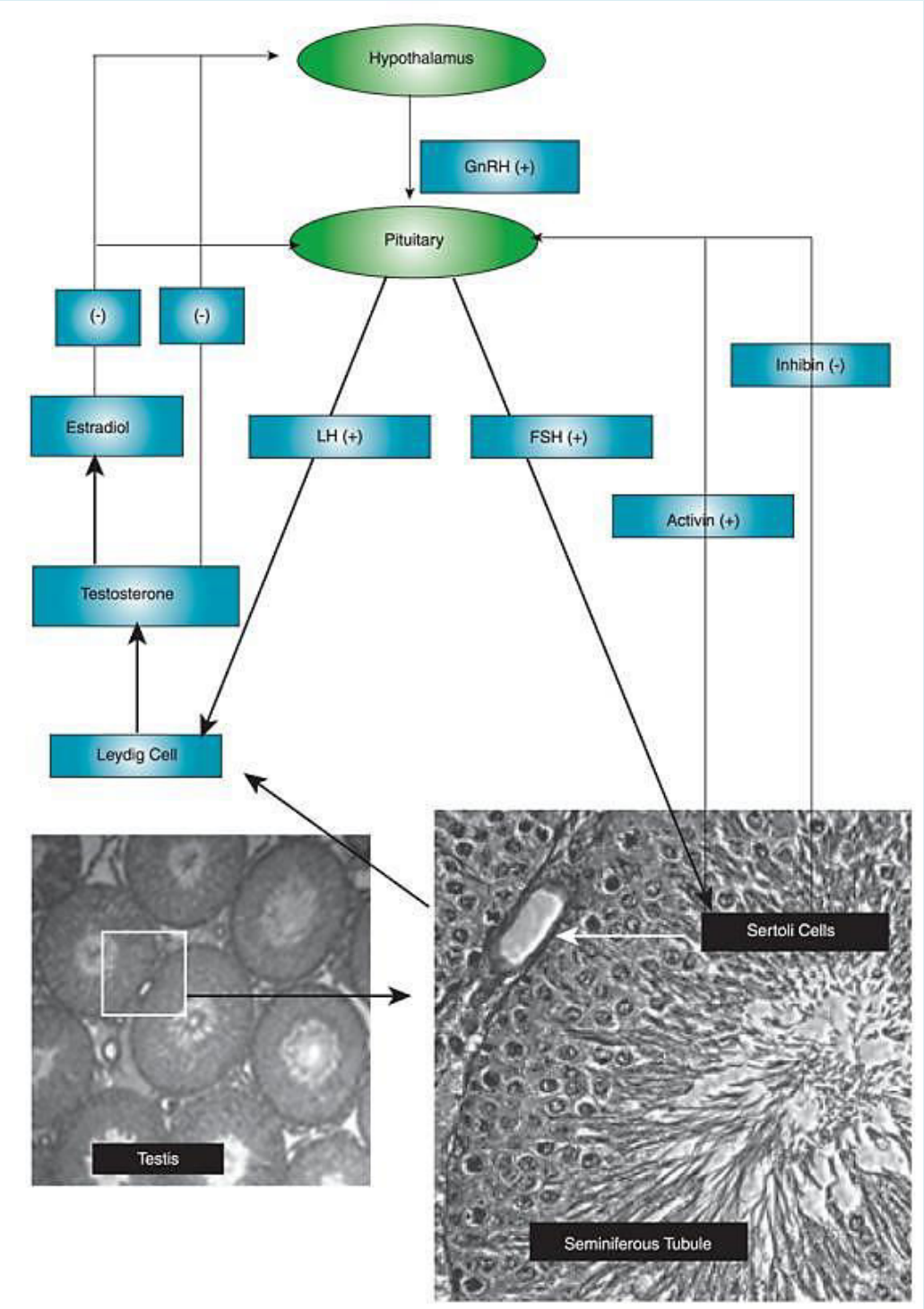


Figure 29. Régulation hormonale de la spermatogenèse.
La production de testostérone est contrôlée par les hormones sécrétées par l'axe hypothalamo-hypophysaire. (+) représente un rétrocontrôle positif alors que (-) représente un rétrocontrôle négatif (d'après Chen et al., 2005).

Durant le premier cycle spermatogénique après la puberté, la FSH agit en synergie avec les androgènes pour limiter la vague apoptotique massive qui survient à ce moment (O'Shaughnessy, 2014). La FSH joue par la suite un rôle dans tous les cycles par l'augmentation du nombre de spermatogonies s'engageant dans ce processus et par conséquence le nombre de spermatocytes rentrant en méiose. Les souris et les hommes qui présentent un déficit de la FSH ou de son récepteur sont cependant fertiles (O'Shaughnessy et al., 2012; Tapanainen et al., 1997). Néanmoins, leurs tubules séminifères sont marqués par une diminution du nombre des cellules germinales suggérant que la FSH n'est pas essentielle à la spermatogenèse mais qu'elle joue un rôle important en synergie avec les androgènes dans l'optimisation de la production des cellules germinales (Huhtaniemi, 2015). Il a été également montré que chez les hommes infertiles, l'administration de la FSH permet d'améliorer dans certains cas les paramètres spermatiques comme la numération et la morphologie (Valenti et al., 2013). D'autres études chez l'homme ont également montré que la FSH stimule la méiose II et le début de la différenciation de la spermatide ronde alors que l'action de de la testostérone permet de potentialiser l'action de la FSH au niveau des dernières étapes de la spermiogénèse (Dimitriadis et al., 2015).

La testostérone est synthétisée dans les cellules interstitielles de Leydig sous l'influence de la LH et elle est fortement concentrée dans les tubules séminifères. En effet, les taux intra-testiculaires de testostérone sont environ 100-1000 fois supérieurs aux taux sériques (Jarow et al., 2001). La testostérone agit sur les récepteurs des androgènes qui sont présents au niveau des cellules de Sertoli et des cellules myoïdes. Les androgènes permettent également de réguler la prolifération des spermatogonies mais jouent surtout un rôle crucial pour l'accomplissement de la méiose depuis les spermatocytes primaires jusqu'aux spermatides rondes (Dimitriadis et al., 2015). Les souris qui présentent un déficit des récepteurs d'androgènes ne parviennent pas à produire des cellules post-méiotiques et sont par conséquence infertiles. De même, les hommes qui présentent un déficit du gène *LHB* (*Luteinizing Hormone Beta Polypeptide*) présentent un hypogonadisme associé à une azoospermie (Lofrano-Porto et al., 2007).

Cependant et de manière surprenante, une étude très récente a montré que chez la souris transgénique knock-out homozygotes pour le gène *Lhr*, l'induction d'une hyperactivation du récepteur de la FSH permet d'inverser le phénotype d'azoospermie et restaurer partiellement leur fertilité (Oduwole et al., 2018). Ceci montre que cette

hyperactivité du récepteur de la FSH permet de compenser l'effet de la LH et suggère qu'un traitement à fortes doses de FSH pourrait améliorer la fertilité des patients hypogonadiques.

Il existe d'autres facteurs qui agissent de manière endocrine ou paracrine pour réguler la spermatogenèse. Cependant la liste est très longue et les mécanismes associés sont complexes et nécessitent d'être mieux élucidés **(Figure 30)** (Schlatt and Ehmcke, 2014).

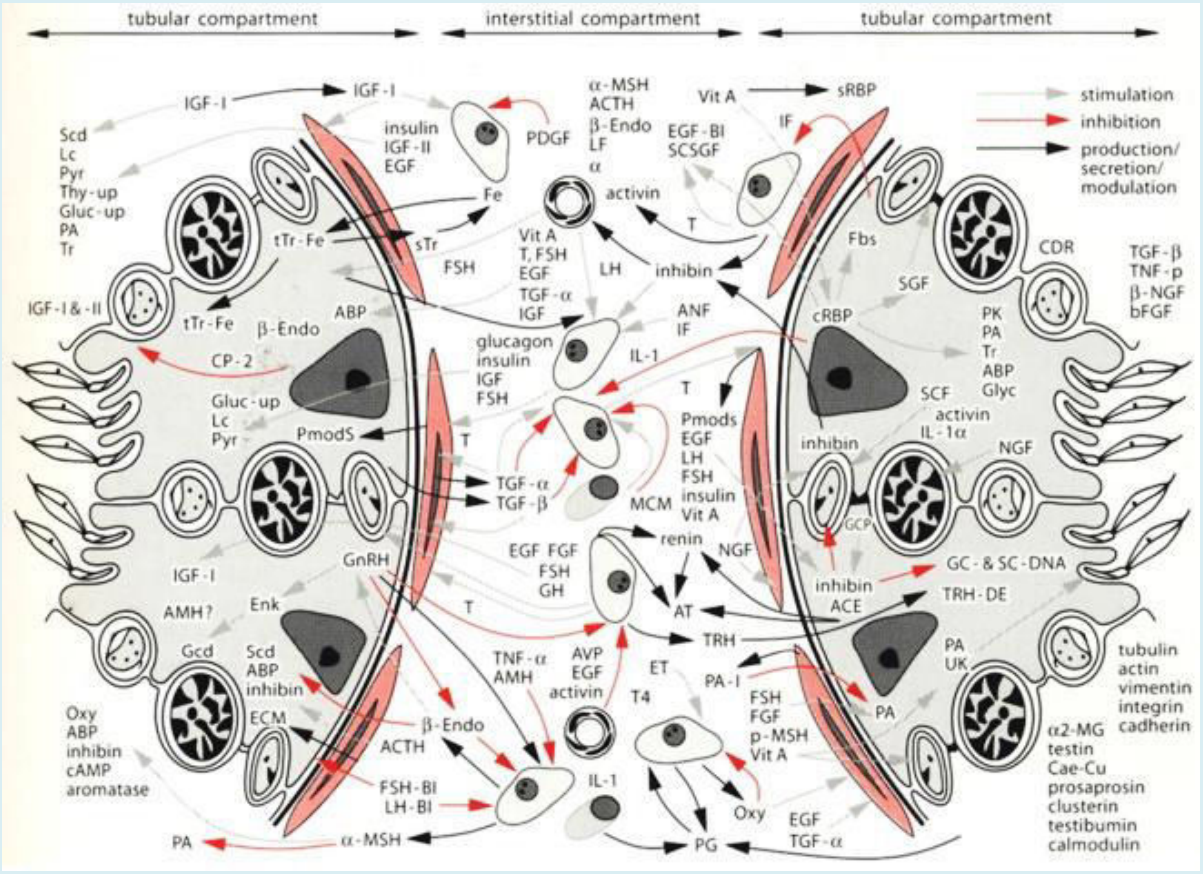


Figure 30. Complexité de la régulation endocrine et paracrine de la spermatogenèse.
(D'après Schlatt and Ehmcke, 2014).

Chapitre II. L'infertilité masculine

1- Définition et données épidémiologiques

Sur le plan nosologique, la fertilité correspond à la capacité d'obtenir un enfant. Selon l'organisation mondiale de la santé (OMS), l'infertilité est définie par l'échec d'obtention d'une grossesse clinique chez des couples en âge de procréer (femme âgée de 18 à 45 ans) au bout de 12 mois de rapports sexuels réguliers non protégés (Zegers-Hochschild et al., 2017). L'infertilité est qualifiée de « primaire » si le couple ne parvient pas à concevoir un premier enfant. Elle est dite « secondaire » en cas d'échec d'obtention de nouvelles grossesses cliniques.

L'infertilité est actuellement considérée par l'OMS comme une préoccupation majeure de santé. Elle affecte approximativement 9% des couples dans le monde et peut atteindre jusqu'à 30% dans certaines régions du monde (Boivin et al., 2007; Ombelet et al., 2008). Cette prévalence semble être constante. Cependant, en raison de la croissance de la population mondiale, on observe une augmentation du nombre de couples infertiles passant de 42 millions (39.6 millions, 44.8 millions) en 1990 à 48.5 millions (45.0 millions, 52.6 millions) en 2010 (Mascarenhas et al., 2012). Il est important de souligner la difficulté liée à l'estimation précise de la prévalence de l'infertilité en raison des disparités régionales dans la définition de l'infertilité et l'absence de données suffisantes concernant le nombre d'hommes infertiles dans certaines régions du monde (Inhorn and Patrizio, 2015).

Une étude française a évalué les étiologies d'infertilité parmi 1686 couples. Le facteur féminin était à l'origine de l'infertilité de couple dans 33% des cas, le facteur masculin dans 20% des cas et une origine mixte dans 39% des cas. Dans 8% des cas l'infertilité était inexpliquée (Thonneau et al., 1991). A l'échelle mondiale, au moins 30 millions d'hommes sont infertiles. Les régions qui présentent les taux les plus élevés de cas d'infertilité masculine sont l'Afrique et l'Europe de l'Est (Agarwal et al., 2015). Chez l'homme les troubles de la spermatogenèse sont les plus fréquemment impliqués dans l'infertilité. Un possible déclin de la qualité spermatique a été rapporté dans plusieurs études (Levine et al., 2017; Vander Borght and Wyns, 2018). Ce déclin concerne l'ensemble des paramètres spermatiques (volume spermatique, numération, mobilité, vitalité et morphologie des spermatozoïdes). Cependant, des biais liés à ces études pourraient être à l'origine d'une fausse interprétation

des données comme l'hétérogénéité des populations recrutées, la spécificité géographique, le changement des habitudes des laboratoires au cours de l'étude et la nature des tests statistiques employés. De plus, la durée d'abstinence sexuelle et la grande variabilité physiologiques inter- et intra-individuelle des paramètres spermatiques pourraient également influencer ces données (Burton, 2013). Les regards se dirigent actuellement vers la possible implication des facteurs environnementaux et soulèvent le débat autour des perturbateurs endocriniens (Gore et al., 2015).

2- Les facteurs pré-testiculaires associés à l'infertilité masculine

2.1- L'hypogonadisme hypogonadotrope

L'hypogonadisme hypogonadotrope traduit une insuffisance de stimulation testiculaire par les hormones hypophysaires : LH (luteinizing hormone) et FSH (follicle stimulating hormone). Ce déficit peut concerner le dysfonctionnement primaire de l'hypophyse (adénome hypophysaire, craniopharyngiome, méningiome etc.) ou être secondaire à l'insuffisance ou à l'absence de sécrétion de la GnRH (gonadotropin-realising hormone). La cause la plus fréquente du déficit en GnRH est le syndrome de Kallmann de Morsier (Hefner et al., 2009). Ce syndrome associe l'absence de synthèse de la GnRH, une perte ou une baisse de l'odorat (anosmie ou hyposmie) et une infertilité. Ces symptômes sont secondaires à un déficit de la morphogenèse des bulbes olfactifs chez le fœtus et l'arrêt de migration des neurones hypothalamiques qui sécrètent la GnRH (Manara et al., 2014). Le syndrome de Kallmann est lié à une forte composante génétique très hétérogène (Layman, 2013).

2.2- L'hyperprolactinémie

La prolactine est une hormone hypophysaire secrétée de façon pulsatile selon un rythme circadien. Les concentrations plasmatiques maximales sont atteintes la nuit pendant le sommeil et les concentrations minimales sont atteintes environ 2-3h après le réveil (Capozzi et al., 2015). L'hyperprolactinémie (Taux de prolactine >15µg/l) chez l'homme est caractérisée cliniquement par une gynécomastie, une galactorrhée, une baisse de la libido et une infertilité. L'hyperprolactinémie inhibe la sécrétion pulsatile de la GnRH conduisant à une diminution de la sécrétion des hormones hypophysaires gonadotropes. La prévalence de ce trouble endocrinien est autour de 11% chez les hommes infertiles atteints d'oligozoospermie (Singh et al., 2011). Le prolactinome (adénome lactotrophe) est la cause la

plus fréquente d'hyperprolactinémie et compte pour environ 40% des adénomes hypophysaires (Dabbous and Atkin, 2018).

2.3- Rôle des perturbateurs endocriniens

La société humaine moderne est largement exposée à des polluants potentiellement toxiques. Parmi ces polluants, les perturbateurs endocriniens émergent comme un potentiel danger pour la santé reproductive. L'exposition à ces polluants synthétiques largement diffusés dans l'environnement se fait par ingestion, inhalation et contact dermique. Une fois dans l'organisme, ces produits interfèrent avec la régulation du système endocrinien aux échelles génomique et physiologique (Annamalai and Namasivayam, 2015). Il existe également une forte évidence que l'exposition prolongée à ces produits durant la vie fœtale soit responsable de dysgénésie testiculaire ou de cryptorchidisme conduisant à une infertilité masculine à l'âge adulte (Juul et al., 2014). Les biphényles polychlorés (BPC), le bisphénol A (BPA), les phthalates et leurs esters, des pesticides comme l'atrazine et les organophosphorés (OP), des insecticides comme le dichlorodiphényl-dichloroéthylène (DDT) et certains métaux lourds tel que le cadmium sont reconnus comme des perturbateurs endocriniens et sont suspectés d'avoir des effets délétères sur la spermatogenèse et les fonctions spermatiques. Bien que leur utilisation dans l'industrie de fabrication des produits courants soit strictement réglementée voire prohibée progressivement pour certains, leur contamination environnementale persistent et il est très difficile de s'en débarrasser rapidement (Mima et al., 2018).

3- Les facteurs testiculaires associés à l'infertilité masculine

3.1- La varicocèle

La varicocèle est une pathologie masculine fréquente dont l'incidence peut atteindre jusqu'à 22% des hommes dans la population générale et jusqu'à 40% des hommes infertiles (Muratorio et al., 2013). La varicocèle représente la pathologie la plus fréquemment observée chez les hommes infertiles. Elle traduit une dilatation anormale des veines spermatiques. Cependant, le mécanisme précis de l'altération de la spermatogenèse reste à ce jour inconnu. Plusieurs hypothèses ont été postulées dans ce contexte comme l'élévation de la température intra-scrotale, l'augmentation de la pression hydrostatique, l'hypoxie et le stress oxydatif (Jensen et al., 2017). Les traitements par ligature chirurgicale ou par embolisation

radiologique couramment proposés à ces patients ne semblent pas améliorer leur fertilité (Baazeem et al., 2011; Kohn et al., 2018).

3.2- Le cryptorchidisme

Le cryptorchidisme est une cause congénitale fréquente d'infertilité masculine qui pourrait être associé à des cancers du testicule (Lee and Houk, 2013). Les enfants nés prématurés sont les plus susceptibles de présenter un cryptorchidisme à la naissance en raison de la descente physiologique tardive des testicules dans le scrotum pendant la vie fœtale (vers le neuvième mois). Son incidence est d'ailleurs proche de 30% chez ces enfants. A l'âge adulte, la fertilité est impactée chez l'ensemble des patients atteints de cryptorchidisme qu'il soit unilatéral ou bilatéral (Goel et al., 2015). L'incidence de l'azoospermie chez les patients présentant une forme unilatérale est de 13% et elle avoisine les 90% dans les formes bilatérales. Une prise en charge médico-chirurgicale précoce dans les premières années de vie permet de réduire cette incidence à l'âge adulte jusqu'à 32% (Chung and Brock, 2011). La pathogénie liée au cryptorchidisme est complexe et très hétérogène. Elles associent des facteurs hormonaux, génétiques, anatomiques et environnementaux (Docampo and Hadziselimovic, 2015).

3.3- Les facteurs génétiques

La gamétogenèse masculine ou spermatogenèse est un processus physiologique très complexe faisant interagir plusieurs centaines de gènes dans un réseau hautement spécialisé. L'observation d'un nombre important de souris KO infertiles ainsi que l'existence d'un nombre important de gènes spécifiquement exprimés dans le testicule (n>500) montrent l'importance de la composante génétique dans les troubles de la spermatogenèse (Tüttelmann et al., 2011). Compte tenu de ces arguments, on peut penser que la grande majorité des cas d'infertilité masculine étiquetés comme idiopathiques soit d'origine monofactorielle et monogénique. Ceci est d'autant plus vrai lorsqu'il s'agit de phénotypes sévères et/ou homogènes comme l'azoospermie non-obstructive et les tératozoospermies monomorphes.

3.3.1- L'azoospermie non-obstructive

Les troubles quantitatifs sévères de la spermatogenèse représentés par l'azoospermie non obstructive (ANO) et l'oligozoospermie sévère (OS) sont les plus susceptibles d'avoir une origine génétique. Ces phénotypes sont d'ailleurs très fréquents

puisqu'on estime qu'ils concernent entre 20-30% des couples infertiles (Thonneau et al., 1991). Les aneuploïdies gonosomiques et les microdélétions du chromosome Y représentent les causes génétiques les plus fréquentes et sont retrouvés dans 2-17% des cas d'ANO et 14% des cas d'ANO d'OS (Jungwirth et al., 2012).

3.3.1.1- Le syndrome de Klinefelter

Les insuffisances de la spermatogenèse peuvent avoir une origine chromosomique. Le plus souvent, il s'agit d'anomalies des chromosomes sexuels parmi lesquelles le syndrome de Klinefelter est le plus fréquent (Lanfranco et al., 2004). Le syndrome de Klinefelter est causé par une aneuploïdie de caryotype 47,XXY **(Figure 31)**. Ce syndrome a été décrit pour la première fois en 1942. Il est caractérisé cliniquement par une atrophie testiculaire associée à un trouble sévère de la spermatogenèse conduisant à une azoospermie dans la majorité des cas. Dans 10% des cas, il existe une mosaïque germinale avec des cellules 46,XY et d'autres 47,XXY qui conduit à un phénotype moins sévère avec une oligozoospermie (Lejeune et al., 2014).

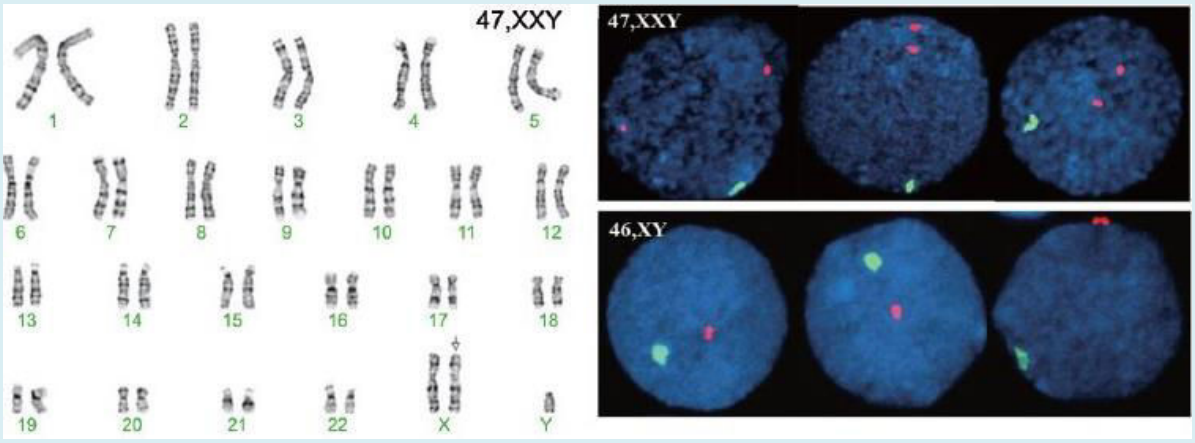


Figure 31. Anomalies chromosomiques dans le syndrome de Klinefelter .
Résultats du caryotype conventionnel 47,XXY et démonstration du mosaïcisme par analyse FISH (fluorescent in situ hybridization). On observe deux types de populations cellulaires avec des cellules normales et d'autres avec deux chromosomes X et un chromosome Y (d'après Mehta et al., 2014).

Dans la mesure où les spermatogonies commencent à disparaitre vers la puberté chez ces individus, un prélèvement précoce avec cryoconservation du tissus testiculaire a été proposé pour tenter de restaurer ultérieurement la spermatogenèse (Van Saen et al., 2012).

3.3.1.2- Les microdélétions sur le bras long du chromosome Y

D'une taille d'environ 53 Mb, le chromosome Y contient plusieurs gènes impliqués dans la différenciation des gonades masculines et la spermatogenèse (Skaletsky et al., 2003). Chez près de 10% des patients souffrant d'ANO et 5% des patients avec OS, des microdélétions du chromosome Y sont observées dans la région AZF (Azoospermia factor) (Hotaling and Carrell, 2014). La région AZF est subdivisée en trois locus AZFa, AZFb et AZFc **(Figure 32)**.

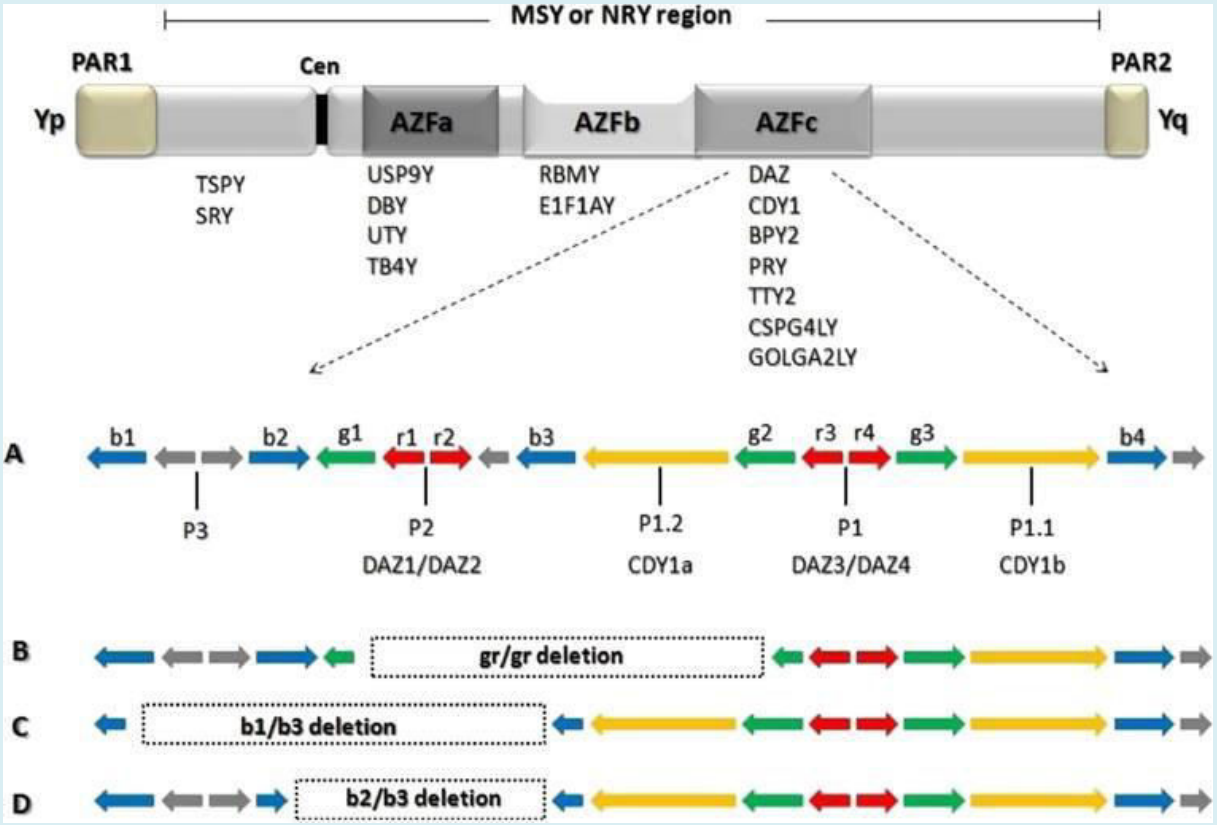


Figure 32. Microdélétions du chromosome Y.
Représentation schématique du chromosome Y et de la région AZF ainsi que des trois sous-régions AZF a, b, c et des principaux gènes compris dans chacune des sous-régions (d'après Tahmasbpour et al., 2014).

Les délétions du locus AZFa causent le syndrome des cellules de Sertoli seules (Sertoli-cell-only) avec absence des cellules germinales dans l'épithélium séminifère (Nuti and Krausz, 2008). Les délétions du locus AZFb sont associées à un blocage variable de la spermatogenèse et donnent une azoospermie.

Les chances de succès d'une récupération de spermatozoïdes après TESE sont inexistantes lorsque la délétion concerne les locus AZFa ou AZFb. Cependant, les délétions du locus AZFc sont responsables d'azoospermie ou d'oligozoospermie du fait de la persistance d'une spermatogenèse résiduelle. Chez les patients porteurs d'une délétion du locus l'AZFc, les chances de succès d'une récupération de spermatozoïdes après TESE sont de l'ordre de 50%. Ces exemples de corrélation génotype/phénotype testiculaire montrent l'importance de développer les tests génétiques dans l'infertilité masculine en particulier pour faire éviter aux patients des explorations invasives inutiles.

3.3.2- Les tératozoospermies monomorphes

Bien que la majorité des cas d''infertilité masculine soit liée à un trouble quantitatif de la spermatogenèse, beaucoup de cas sont associés à une altération qualitative de ce processus qui aboutit à la production de spermatozoïdes anormaux et non fonctionnels. On parle de tératozoospermie lorsque les anomalies constatées sont morphologiques. Les tératozoospermies monomorphes sont définies par la présence dans l'éjaculat de spermatozoïdes atypiques concernés par les mêmes anomalies morphologiques. Trois formes rares de tératozoospermie monomorphe, représentant environ 1 % des infertilités masculines, sont présentées ici : la macrozoospermie, la globozoospermie et le phénotype MMAF (multiple morphological anomalies of the sperm flagellum).

3.3.2.1- La macrozoospermie : les spermatozoïdes macrocéphales multi-flagellés

La macrozoospermie est un phénotype rare caractérisé par la présence dans l'éjaculat de 100% de spermatozoïdes présentant une tête surdimensionnée et de multiples flagelles. Ce phénotype est souvent associé à une oligo-asthéno-zoospermie (Kahraman et al., 1999). Depuis sa description il y a 40 ans (Ray et al., 2017), plusieurs études ont rapporté un taux élevé d'aneuploïdie et de polyploïdie dans ces spermatozoïdes. Ces observations suggèrent que cette anomalie soit liée à la non-disjonction des chromosomes au cours de la méiose. En 2007, une étude réalisée par l'équipe GETI de Grenoble a permis d'identifier la même mutation homozygote (c.144delC) dans le gène *AURKC (aurora kinase C*) chez l'ensemble des patients inclus dans l'étude soit 14 patients nord-africains (Dieterich et al., 2007). Ce gène code pour une protéine appartenant au CPC (chromosomal passenger complex) qui régule certains événements méiotiques comme la ségrégation des chromosomes et la cytokinèse. La mutation identifiée génère un codon stop prématuré dans le transcrit qui est rapidement dégradé par le système NMD (nonsense-mediated decay) (Ben Khelifa et al.,

2011). L'étude par cytométrie en flux a montré l'existence d'une tétraploïdie homogène dans ces spermatozoïdes expliquant que les cellules germinales ne peuvent pas se diviser après la réplication de leur ADN (Dieterich et al., 2009). Malgré ce défaut méiotique, la spermiogénèse n'est pas abolie chez ces patients et aboutit à la production de spermatozoïdes macrocéphales multi-flagellés **(Figure 33)** (Ben Khelifa et al., 2011).

La prise en charge de ces patients devrait commencer par la recherche de cette mutation dans le gène *AURKC*. La FIV-ICSI est alors formellement contre-indiquée chez les patients présentant des mutations homozygotes en raison de la présence constante de polyploïdie incompatible avec le développement embryonnaire. Concernant les patients qui ne présentent pas de mutations du gène *AURKC*, l'ICSI pourrait être envisagée après une analyse chromosomique par FISH (Fluorescence in situ hybridization) des spermatozoïdes afin d'évaluer le taux d'aneuploïdie. Un suivi rapproché de la grossesse devrait alors être instauré. Un diagnostic préimplantatoire pourrait également être envisagé dans certains cas présentant un taux intermédiaire d'aneuploïdie.

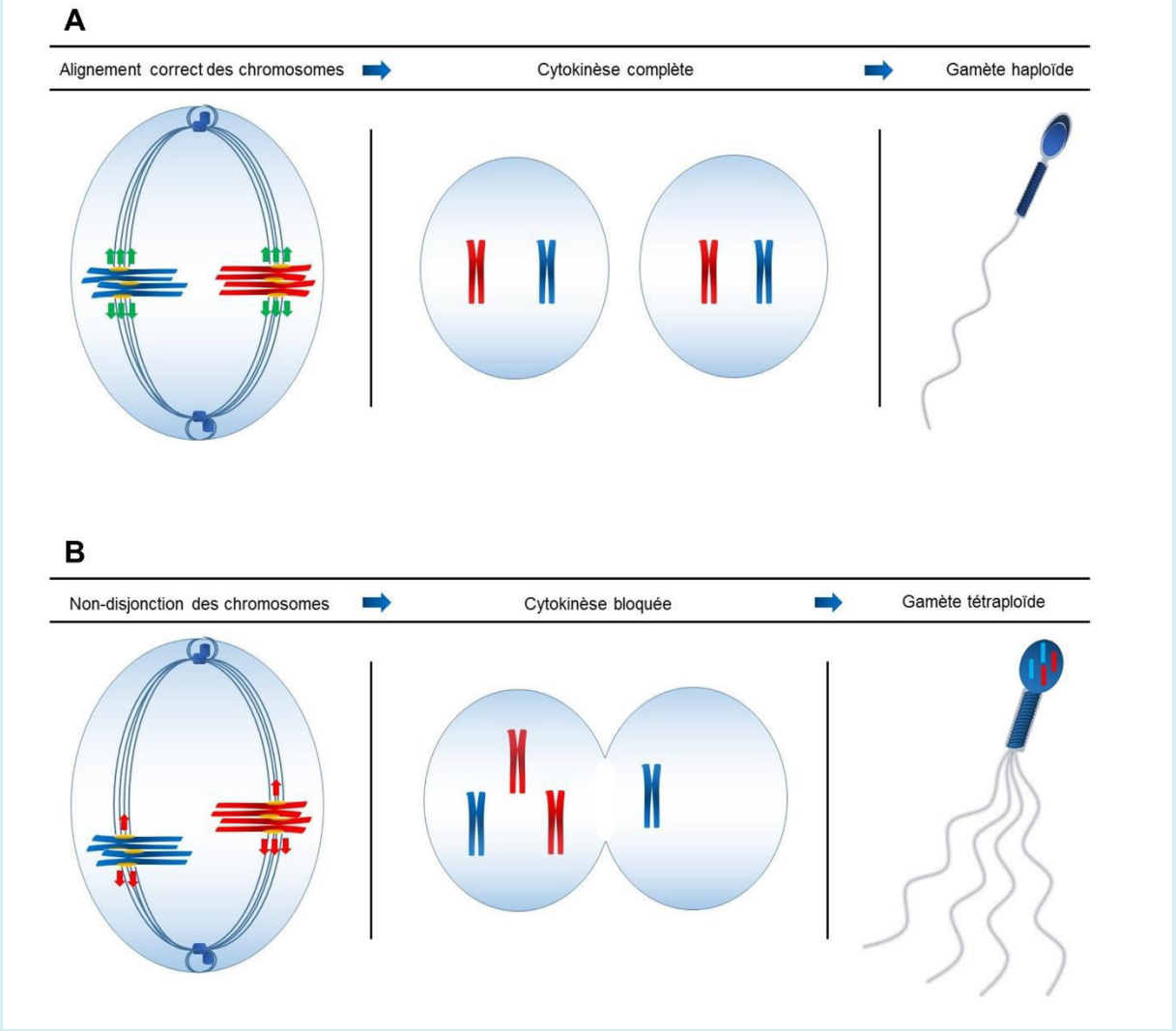


Figure 33. La macrozoospermie : Rôle de l'AURKC durant la méiose et la cytokinèse.
(A) Au cours de la méiose, les chromosomes homologues s'apparient et s'alignent au niveau de la plaque équatoriale grâce à l'équilibre des tensions subies par chacun des kinétochores (en jaune) d'un chromosome. A la fin de chaque division méiotique (ici la méiose II n'est pas présentée), les cellules filles se séparent (cytokinèse). Le gamète mâle produit est haploïde.
(B) Chez les patients porteurs de mutations homozygotes du gène AURKC, on observe un déséquilibre d'alignement et une non-disjonction des chromosomes au cours de la méiose I. On observe aussi un échec de la cytokinèse et les spermatozoïdes produits sont tétraploïdes et pourvus de plusieurs flagelles (d'après *Z-E Kherraf, GETI*).

3.3.2.2- La globozoospermie : les spermatozoïdes microcéphaliques

La globozoospermie est aussi un phénotype rare observé chez moins de 0.1% des hommes infertiles. Elle est caractérisée par la présence dans l'éjaculat d'une majorité de spermatozoïdes présentant une tête globulaire dépourvue d'acrosome **(Figure 34)**. La présence de mariages consanguins dans les familles affectées par la globozoospermie suggère la contribution d'une composante génétique transmise sous un mode autosomique récessif. Deux gènes impliqués dans la biogenèse de l'acrosome et l'organisation de la tête du spermatozoïdes ont été identifiés dans la globozoospermie (Coutton et al., 2015).

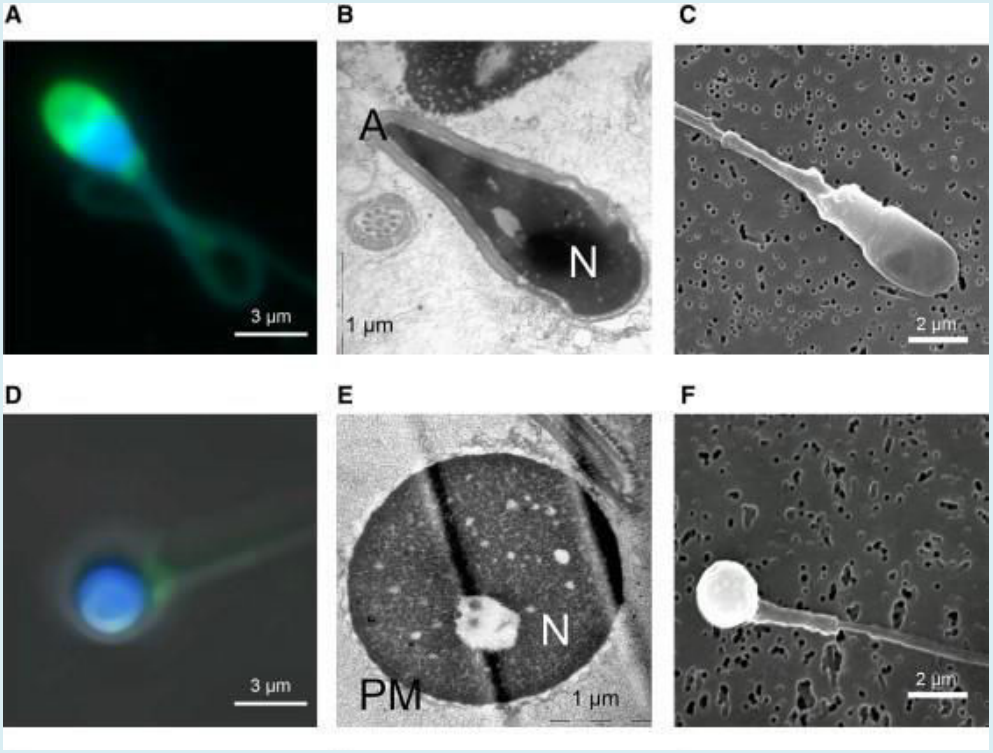


Figure 34. La globozoospermie.
(A-C) Spermatozoïde normal ; (D-F) spermatozoïde globocéphalique.
(A, D) Images de microscopie confocale prise après co-marquage des noyaux (en bleu) et de l'acrosome (en vert). On observe la disparition du marquage vert précédemment localisé sur le pôle nucléaire apical du spermatozoïde globocéphalique. (B, E) Images de microscopie électronique par transmission montrant l'absence de l'acrosome (E). (C, F) images de microscopie par balayage (d'après Harbuz et al., 2011)

En 2007, Dam et al. ont identifié une mutation homozygote (c.848G>A, p.R283Q) dans le gène *SPATA16* (spermatogenesis-associated protein 16) chez trois frères d'une famille consanguine ashkénaze (Dam et al., 2007). La protéine produite est exprimée pendant la

spermiogénèse. Elle est localisée dans l'appareil de Golgi et les vésicules proacrosomales qui fusionnent sur le pôle antérieur du noyau pour former l'acrosome. Deux autres études récentes ont permis d'identifier de nouvelles mutations du gène *SPATA16* chez les patients globocéphales (ElInati et al., 2016; Karaca et al., 2014).

En 2011, Harbuz et al. ont étudié une cohorte de 20 sujets tunisiens globozoospermiques et ont identifié une délétion homozygote de 200 kb emportant le gène *DPY19L2* (dpy-19 like 2) dans 75% des cas (n=15) (Harbuz et al., 2011). D'autres études ont par la suite confirmé la prévalence élevée des altérations du gène *DPY19L2* dans la globozoospermie dans d'autres régions du monde (Coutton et al., 2012; ElInati et al., 2012; Zhu et al., 2013). La protéine DPY19L2 est localisée dans la membrane interne du noyau à partir du stade de spermatide ronde. Elle est impliquée dans le positionnement de l'acroplaxome, une structure dérivée du cytosquelette et localisée au niveau du pôle apical du noyau spermatique. L'acroplaxome permet l'ancrage de l'acrosome au niveau de l'enveloppe nucléaire. L'absence de DPY19L2 fragilise l'acroplaxome et cause le détachement et la perte de l'acrosome au cours de la spermiogénèse **(Figure 35)**.

L'absence de l'acrosome rend les spermatozoïdes globocéphales incompatibles avec la fécondation in vitro (FIV) standard. Quant à la FIV-ICSI, l'échec de fécondation observé est attribué à un déficit en facteur d'activation ovocytaire. Il s'agit d'une phospholipase spécifique des cellules germinales testiculaires, la PLCζ qui est perdu avec l'acrosome dans la globozoospermie. L'absence de la PLCζ peut être remplacée par l'injection concomitante d'un ionophore calcique avec le spermatozoïde lors de l'ICSI. Cependant, les taux de grossesses sont modérés en raison de la présence de défauts de condensation de la chromatine et une fragmentation de l'ADN spermatique. Par ailleurs l'utilisation d'un ionophore calcique n'est actuellement pas permise en France.

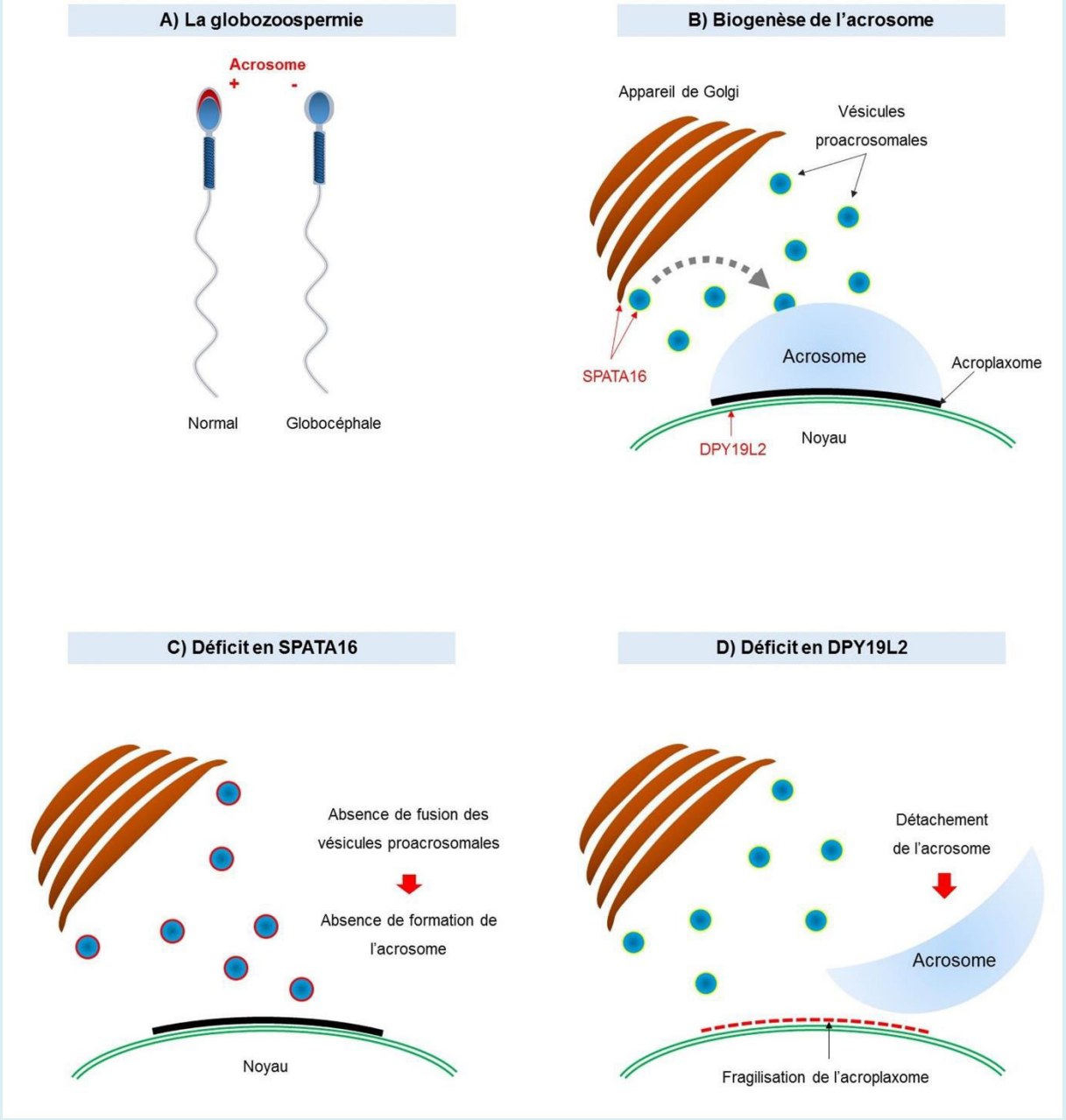


Figure 35. Physiopathologie de la globozoospermie.
(A) Représentation schématique de la morphologie d'un spermatozoïde globocéphale dépourvu d'acrosome (à droite) comparé à un spermatozoïde normal (à gauche).
(B) Biogenèse normale de l'acrosome au cours de la spermiogénèse : l'appareil de Golgi produit des vésicules proacrosomales qui fusionnent sur l'acroplaxome positionné sur le pôle antérieur du noyau de la spermatide pour former l'acrosome.
(C) Le déficit en SPATA16 chez les patients globozoospermiques empêche la fusion des vésicules proacrosomales et l'acrosome ne se forme pas.
(D) En cas de déficit en DPY19L2, l'acroplaxome est fragilisé et l'acrosome en formation se détache du pole antérieur du noyau des spermatides. Finalement, l'acrosome est perdu avec l'excès cytoplasmique du spermatozoïde au cours de sa maturation (d'après *Z-E Kherraf, GETI*).

3.3.2.3- L'asthénozoospermie et le phénotype MMAF

L'asthénozoospermie est définie par la réduction ou l'absence de mobilité spermatique. Selon l'OMS, le diagnostic est établi lorsqu'on observe moins de 32% de spermatozoïdes mobiles-progressifs (Cooper et al., 2010). Ce phénotype pourrait être associé à la nécrozoospermie ou résulter d'un défaut structurel et/ou fonctionnel du spermatozoïde. L'asthénozoospermie est souvent inscrite dans un tableau syndromique autosomique récessif de dyskinésie ciliaire primitive dont le principal symptôme caractéristique est l'infection récurrente des voies respiratoires. Ces infections sont secondaires à un dysfonctionnement ciliaire et une stase de la clairance des sécrétions muqueuses.

Parmi les gènes candidats identifiés dans l'asthénozoospermie, on peut citer *SEPT12*, *CATSPER1* et *CATSPER2*. Le gène *SEPT12* code pour une GTPase impliquée dans la formation des filaments du cytosquelette. Les mutations rapportées dans ce gène conduisent à l'altération de l'intégrité structurelle du flagelle. Cette anomalie est secondaire à une perturbation de la polymérisation des filaments de l'annulus, une structure en forme d'anneau qui fait jonction entre la pièce intermédiaire et la pièce principale du flagelle (Kuo et al., 2012). Les gènes *CATSPER1* et *CATSPER2* codent pour des protéines appartenant à un canal cationique spécifiquement localisé dans la membrane cytoplasmique de la pièce principale du flagelle. Ce canal module les flux calciques à travers la membrane flagellaire de manière pH-dépendante et il est responsable de la capacitation et de l'hyperactivation de la mobilité du spermatozoïde (Avenarius et al., 2009; Avidan et al., 2003).

Le phénotype MMAF (multiple morphological abnormalities of the sperm flagella) est caractérisé par une asthénozoospermie associée à des anomalies morphologiques multiples du flagelle incluant des flagelles courts, épais, enroulés et/ou de calibre irrégulier (Coutton et al., 2015). En 2014, Ben Khelifa et al. ont montré que le gène *DNAH1* (*dynein axonemal heavy chain 1*) est muté chez approximativement un tiers des patients analysés porteurs de ce syndrome. Lorsque la protéine DNAH1 est absente, l'architecture axonémale est désorganisée avec souvent, une perte de la paire centrale de microtubules **(Figure 36)** (Ben Khelifa et al., 2014). Plusieurs études ont montré une élévation du taux d'aneuploïdie et une altération de la qualité nucléaire spermatique chez les patients atteints de MMAF. Cependant, chez les patients mutés pour *DNAH1*, ces anomalies sont rares et la FIV-ICSI a un bon pronostic (Wambergue et al., 2016).

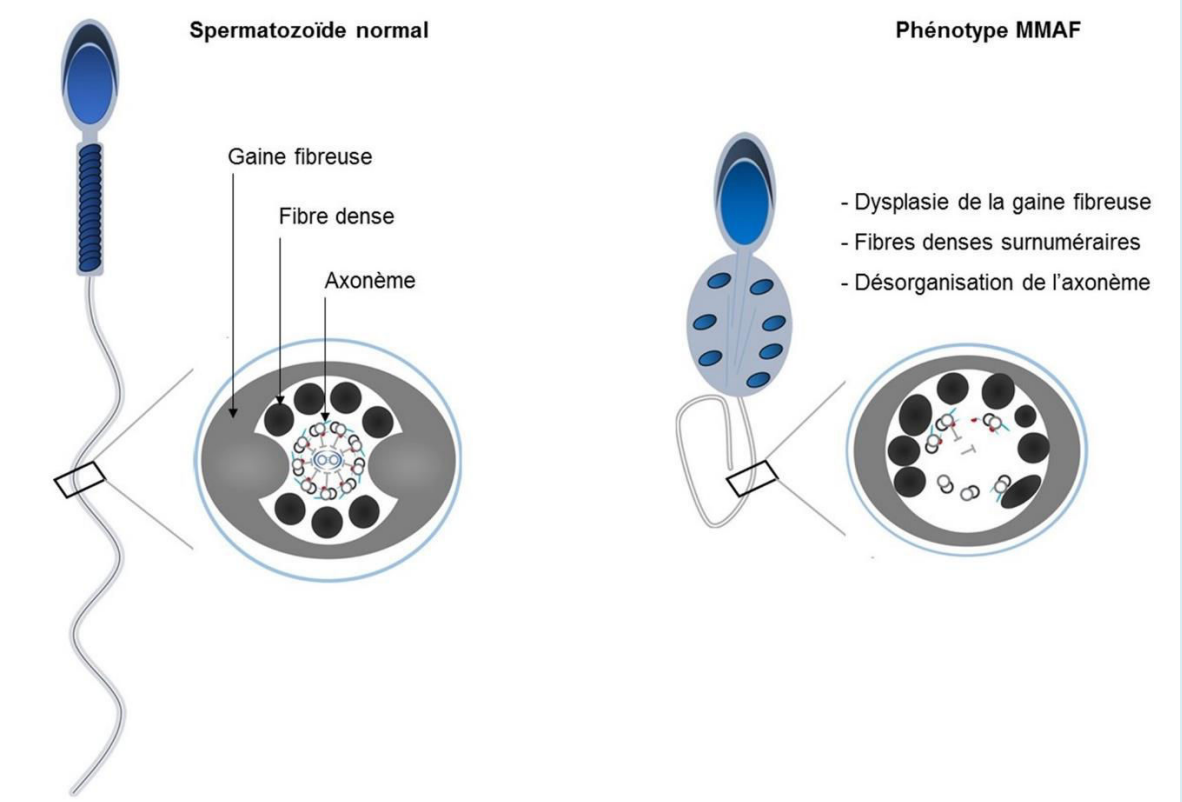


Figure 36. Le phénotype MMAF.
A gauche, un spermatozoïde normal et une coupe transversale au niveau de la pièce principale du flagelle. On note du centre vers la périphérie : un axonème avec un complexe central et 9 doublets périphériques de microtubules, sept fibres denses, une gaine fibreuse et une membrane cytoplasmique.
A droite, un spermatozoïde typique d'un patient MMAF avec un flagelle de morphologie anormale et une structure interne profondément désorganisée (d'après *Z-E Kherraf, GETI)*.

4- Les infertilités obstructives

Elles traduisent une anomalie obstructive des voies excrétrices des spermatozoïdes située entre le testicule et le carrefour urogénital. Le pronostic reproductif est relativement bon puisque la spermatogenèse est souvent normale à la différence de l'azoospermie non obstructive où la spermatogenèse est altérée **(Figure 37)**. Les anastomoses chirurgicales des voies spermatiques lorsqu'elles sont possibles peuvent présenter une solution thérapeutique efficace. Le prélèvement des spermatozoïdes en amont de l'obstruction en vue d'une utilisation en FIV-ICSI peut être considéré comme une alternative thérapeutique palliative.

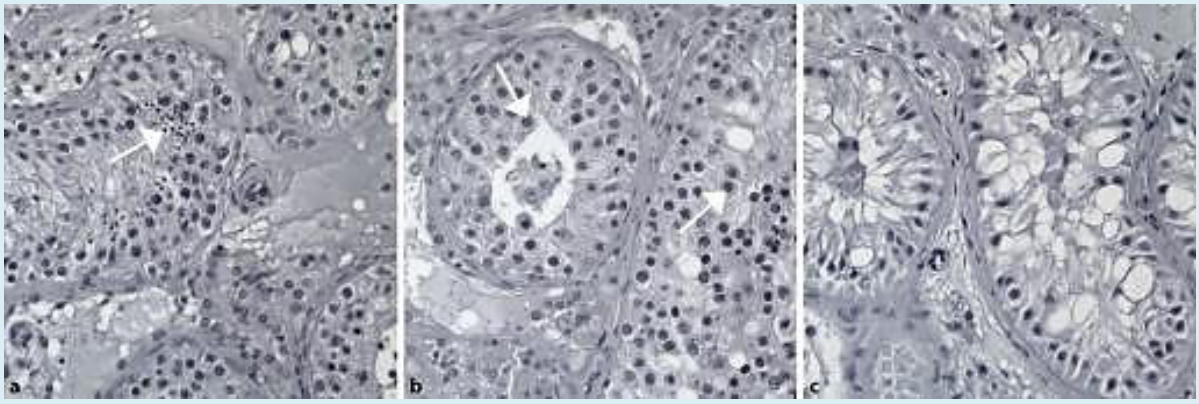


Figure 37. La spermatogenèse dans l'azospermie.
Images de coupes histologiques de tubules séminifères provenant d'un patient présentant une azoospermie obstructive (a) caractérisée par une spermatogenèse normale, un patient présentant un arrêt méiotique de la spermatogenèse (b) et un patient atteint du syndrome des cellules de Sertoli seules (c) (d'après Tüttelmann et al., 2018).

4.1- L'absence bilatérale des canaux déférents (ABCD)

L'absence bilatérale des canaux déférents (ABCD) est le phénotype d'infertilité masculine le plus fréquemment associé à des anomalies génétiques. Ce phénotype concerne 6-25% des cas d'azoospermie obstructive (Yu et al., 2012). L'ABCD est causée par la stase et l'épaississement des sécrétions muqueuses qui favorisent les infections puis l'obstruction cicatricielle. C'est pour cette raison que le terme « absence » est aujourd'hui préférable au terme « agénésie » dans l'acronyme ABCD.

4.1.1- Le gène CFTR (cystic fibrosis transmembrane regulator)

Le gène *CFTR* est muté dans approximativement 80% des cas (O'Flynn O'Brien et al., 2010). L'ABCD est présente de façon constante chez les patients atteints de

mucoviscidose (Radpour et al., 2008). Le gène *CFTR* est localisé dans le chromosome 7 q31.2. Il a une taille de 190 kb et il est composé de 27 exons. Plus de 2000 mutations sont listées dans la base de données hébergée sur le site Web : *http://www.genetsickkids.on.ca/cftr/*. Les mutations les plus fréquemment observées dans les pays européens sont : delta F308 (66.8%), G542X (2.6%), N1303K (1.6%), G551D (1.5%) et W1282X (1.0%) (Estivill et al., 1997).

Six classes de mutations ont été décrites **(Figure 38)** : classe I, donnant une protéine tronquée non fonctionnelle (G542X) ; classe II, causant un défaut d'adressage de la protéine vers la membrane cellulaire apicale (ΔF508) ; classe III, entrainant une altération de la régulation par l'AMPc et de l'ouverture du canal Chlore (G551D) ; classe IV, causant un déclin partiel de la conductance du Chlore (R117H) ; classe V, caractérisée par une baisse du taux des protéines CFTR fonctionnelles (3849+10kbC-T) et classe VI causant une diminution de la stabilité de la protéine (Q1412X). Les mutations des classes IV, V et VI sont responsables d'un phénotype clinique moins sévère du fait de la persistance d'une activité résiduelle des protéines CFTR (Hamada et al., 2013).

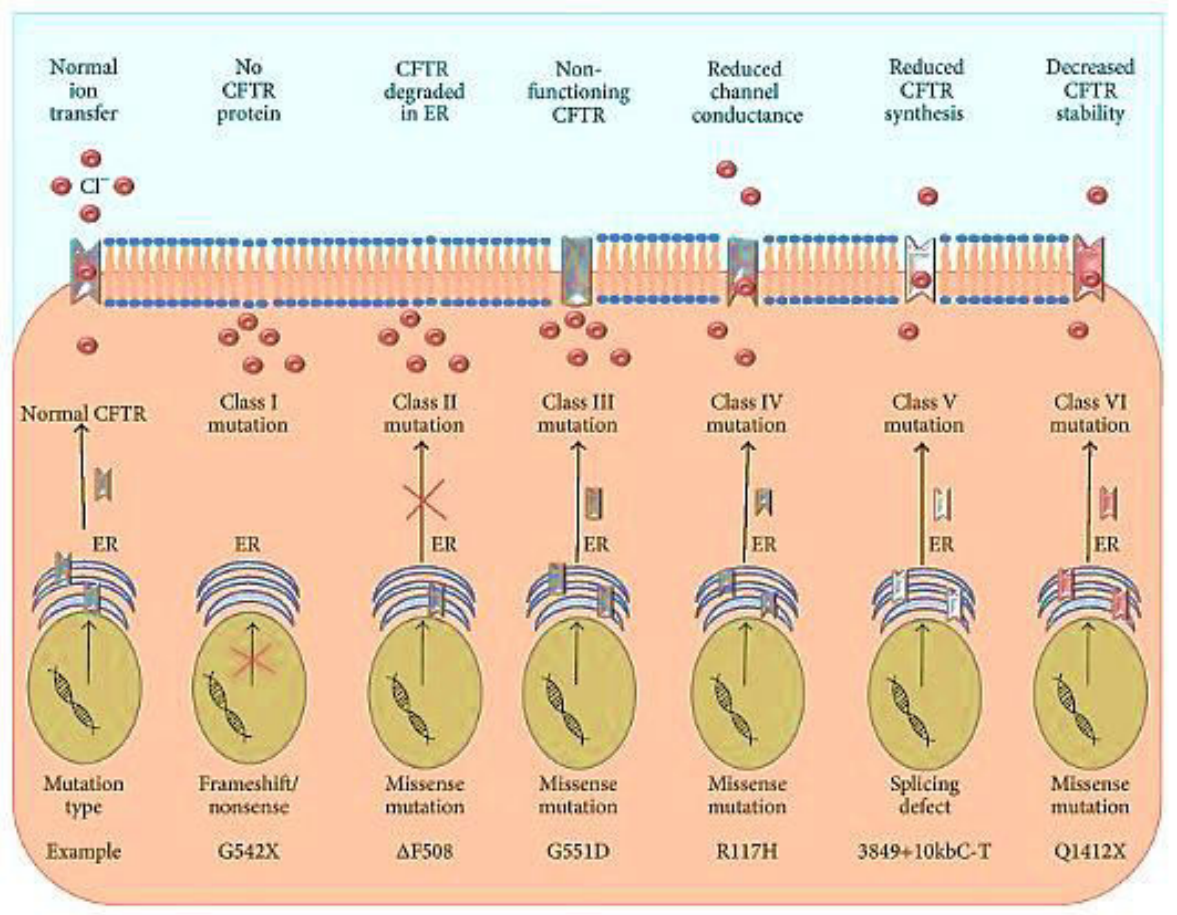


Figure 38. Classification des mutations du gène *CFTR*.
(D'après Lavelle et al., 2016).

Les patients atteints d'ABCD présentent soit deux mutations majeures (classes I, II ou III) ou la combinaison d'une mutation majeure avec une mutation mineure (classe IV ou V) ou un polymorphisme d'épissage IVS8-5T de l'intron 8 (Dequeker et al., 2009).

La FIV-ICSI est une solution efficace pour le traitement de l'azoospermie des patients présentant des mutations du *CFTR*. Cependant un conseil génétique voire un diagnostic pré-implantatoire (DPI) doivent être conseillés aux patients dont la partenaire présente une mutation sur ce gène pour limiter la transmission des anomalies génétiques aux descendants.

4.1.2- Le gène ADGRG2 (adhesion G protein-coupled receptor G2)

Dans environ 20% des cas d'ABCD, on ne retrouve pas de mutation sur le gène *CFTR*. En 2016, Patat et al. ont analysé une large cohorte de patients présentant une ABCD

confirmée non-associée à des mutations du gène *CFTR* (Patat et al., 2016). Avec l'utilisation du séquençage haut débit chez 12 patients et du séquençage ciblé chez 14 autres patients, ils ont réussi à identifier trois mutations frameshift dans le gène *ADGRG2* chez 4 sujets testés. L'*ADGRG2* est un gène situé sur le chromosome X et est exprimé de manière spécifique dans les canaux déférents. Des études antérieures avaient montré que les souris *Adgrg2* knock-out (KO) développaient elles aussi une infertilité obstructive (Davies et al., 2004). L'analyse moléculaire du gène *ADGRG2* devrait alors être envisagée chez les patients atteints d'ABCD chez qui on n'a pas retrouvé de mutation sur le gène *CFTR*. Cette nouvelle analyse pourrait réduire le nombre de cas étiquetés actuellement comme idiopathiques et améliorer le conseil génétique chez ces patients en tenant en compte du mode de transmission lié à l'X de ces mutations.

4.2- Le syndrome de Young

C'est une pathologie très rare dont la prévalence n'est pas connue. Elle associe une bronchiectasie, une sinusite chronique et une azoospermie excrétoire dues à une obstruction bilatérale des épididymes par des sécrétions épaisses. L'étiologie de ce syndrome demeure inconnue. Il ne semble pas y avoir d'association avec des mutations du gène *CFTR* (Friedman et al., 1995).

4.3- Les obstructions post-infectieuses et iatrogènes

Une mauvaise cicatrisation tissulaire à la suite d'infections génitales comme une prostatite, une urétrite à gonocoque ou à chlamydiae ou une tuberculose urogénitale peut aboutir à une obstruction des conduits spermatiques. Une infertilité obstructive peut également être secondaire à des séquelles chirurgicales : c'est le cas des lésions accidentelles du canal déférent à l'occasion d'une chirurgie herniaire, vésicale, prostatique ou pelvienne extensive.

Chapitre III. Les investigations cliniques et paracliniques de l'infertilité masculine

1- L'examen clinique

L'interrogatoire présente le temps principal d'une consultation d'andrologie. Il s'attache à rechercher les facteurs pouvant perturber la fonction testiculaire et les antécédents personnels et familiaux d'infertilité et de maladies génétiques. Le deuxième temps de consultation est consacré à l'examen physique avec la recherche d'éventuels signes d'hypoandrisme témoignant de l'altération de la fonction endocrine du testicule puis l'examen de l'appareil génital dans sa globalité (hypo ou hypertrophie testiculaire, dilatation épididymaire, absence des canaux déférents et/ou des vésicules séminales, varicocèle, hypospadias etc.).

2- L'analyse du sperme

2.1- Le spermogramme

En 1951, Mac Léod étudia la relation qui pourrait exister entre la fertilité masculine et les caractéristiques du sperme de deux groupes d'hommes, féconds et inféconds. Depuis, l'analyse du sperme constitue la première investigation à envisager pour explorer l'infertilité masculine. Cet examen devrait être réalisé selon les dernières recommandations de l'OMS *(World Health Organization. WHO Laboratory Manual for the Examination and Processing of Human Semen, in 5th edn. 2010)* (Esteves, 2016). Il doit être interprété en tenant compte des données cliniques du patient **(Tableau 2)**.

Paramètre	Limite basse de référence
Volume du sperme (mL)	1,5 (1,4-1,7)*
Nombre total de spermatozoïdes (10^6/éjaculat)	39 (33-46)*
Concentration de spermatozoïdes (10^6/mL)	15 (12-16)*
Mobilité totale (Pr + NP, %)	40 (38-42)*
Mobilité progressive (Pr, %)	32 (31-34)*
Vitalité (spermatozoïdes vivants, %)	58 (55-63)*
Morphologie (formes normales, %)	4 (3,0-4,0)*

Tableau 2. Limites inférieures de référence des paramètres du sperme humain (OMS, 2010). *5e percentile (et intervalle de confiance à 95%). Pr = Spermatozoïdes progressifs, NP = Spermatozoïdes non progressifs.

Des facteurs pré-analytiques peuvent biaiser l'interprétation du spermogramme. La collection doit concerner la totalité de l'éjaculat après une période d'abstinence de 2-7 jours. L'acheminement de l'échantillon au laboratoire est réalisé à température ambiante. Au laboratoire, le sperme est liquéfié avant d'être analysé, Tout aspect anormal est noté telle qu'une hémospermie ou une forte viscosité.

Le volume traduit essentiellement les capacités sécrétoires des glandes annexes, le volume des spermatozoïdes dans l'éjaculat ne représente que 10-15% du volume total. Un volume spermatique faible inférieur à 1.5 mL (hypospermie) traduit une atteinte des glandes annexes (prostatite, vésiculite, agénésie des glandes séminales) ou un trouble éjaculatoire. Le diagnostic d'éjaculation rétrograde est confirmé par l'absence d'éjaculat visible avec l'observation de spermatozoïdes dans un échantillon d'urine post-éjaculatoire.

Le pH normal du sperme est compris entre 7.4 et 8. Des valeurs plus basses orientent vers un défaut de sécrétion des vésicules séminales (normalement alcaline) alors que des valeurs alcalines orientent plutôt vers un défaut de sécrétion de la prostate (normalement acide).

La numération des spermatozoïdes est exprimée en concentration (millions/mL). Une numération normale compte 15-200 millions de spermatozoïdes par millilitre. En dessous de 15 millions par millilitre, on parlera d'oligozoospermie. Devant

l'absence de spermatozoïdes, il est important de rechercher leur présence dans le culot de centrifugation (3000 g pendant 15 min) avant de conclure à une azoospermie.

La mobilité des spermatozoïdes est d'abord appréciée au microscope optique puis peut être analysée par vidéomicrographie assistée par ordinateur (system CASA, Computer-Assisted Sperm Analysis). Ce système permet une mesure automatisée de la trajectoire de la tête du spermatozoïde qui est un bon reflet de l'activité du flagelle. Devant une mobilité abaissée, un test de vitalité par coloration à l'éosine-nigrosine (test de Williams) s'avère nécessaire pour différentier la mort cellulaire de l'asthénozoospermie. Ce test consiste à évaluer le pourcentage des spermatozoïdes morts qui prennent la coloration rose-orangé sur fond noir et le pourcentage des spermatozoïdes vivants incolores.

Les cellules rondes doivent être comptées et leur nature appréciée. Il peut s'agir de cellules épithéliales provenant des voies urogénitales, de cellules germinales immatures ou de leucocytes.

2.1.1- Le spermocytogramme

Cet examen vient compléter les données du spermogramme en quantifiant les anomalies morphologiques des spermatozoïdes en microscopie optique après coloration de Harris Shorr ou de Papanicolaou modifiée (les plus utilisées). L'observation est faite sur cent spermatozoïdes. Les anomalies morphologiques sont décrites selon la classification de David modifiée par Auger et Eustache (la plus utilisée). Cette classification répertorie sept anomalies de la tête (allongée, amincie, microcéphale, macrocéphale, multiple, présentant un acrosome anormal ou absent, présentant une base anormale), trois anomalies de la pièce intermédiaire (angulée, grêle, présence de restes cytoplasmiques) et cinq anomalies de la pièce principale (absente, écourtée, de calibre irrégulier, enroulée, multiple). En dessous de 4% de formes typiques on parle de tératozoospermie. L'existence d'anomalies morphologiques multiples doit faire calculer l'index d'anomalies multiples, IAM (nombre d'anomalies observées / nombre total de spermatozoïdes atypiques) qui doit être inférieur à 1,6.

Si le spermogramme-spermocytogramme est normal, il n'est pas nécessaire de le contrôler. Cependant si une anomalie a été observée, il faudrait le répéter sur un autre recueil à au moins un mois d'intervalle après le précédent en raison de la grande fluctuation physiologique des paramètres spermatiques (Dua and Vaidya, 1996).

2.2- Le dosage des biomarqueurs séminaux

Le liquide séminal est produit principalement par l'épididyme et les glandes annexes. Le dosage des marqueurs biochimiques séminaux est réalisé en cas d'anomalies du spermogramme ou d'une suspicion d'infection génitale. La L-carnitine et l'α-glucosidase sont deux marqueurs caractéristiques de l'épididyme, le fructose est caractéristique des vésicules séminales et l'acide citrique, les phosphatases acides et le zinc sont caractéristiques de la prostate. Dans le cadre de l'azoospermie, les résultats de ces dosages apportent des renseignements précieux et permettent de distinguer une azoospermie sécrétoire (testiculaire) d'une azoospermie excrétoire (obstructive) et de préciser le niveau d'obstruction dans ce dernier cas.

3- L'exploration hormonale de l'axe gonadotrope

Lorsqu'un bilan hormonal est indiqué, il consiste à doser en première intention la FSH et éventuellement l'inhibine B pour évaluer la spermatogenèse en plus de la testostérone totale pour évaluer l'androgenèse. Si la testostéronémie est basse, il convient de contrôler le résultat sur un nouveau prélèvement et de réaliser le dosage de la LH et de la prolactine.

4- Les investigations génétiques de routine

Il est recommandé chez les patients présentant une oligozoospermie (numération < 5 millions/mL) ou une azoospermie non obstructive de réaliser un caryotype et une recherche de microdélétions du chromosome Y. Les aneuploïdies gonosomiques et les microdélétions du chromosome Y sont retrouvées respectivement dans 2-17% et 14% des cas d'azoospermie non obstructive (ANO) et d'oligozoospermie sévère (OS).

La présence d'une absence bilatérale des canaux déférents (ABCD) doit indiquer une étude moléculaire du gène *CFTR* (*cystic fibrosis transmembrane conductance regulator*) chez le patient. Le *CFTR* est trouvé muté dans 60-90% des cas d'ABCD. Suite à un diagnostic positif, la conjointe doit être génotypée avant toute biopsie testiculaire et ICSI compte tenu du risque de transmission de la mucoviscidose à la descendance.

4.1- Identification de nouveaux gènes candidats par le séquençage exomique

4.1.1- Introduction

Beaucoup de maladies humaines résultent d'anomalies génétiques. L'identification des facteurs génétiques sous-jacents dans une pathologie spécifique pourrait aider dans l'établissement d'un diagnostic précis et guider le patient dans le parcours de soin (Smith, 2017). L'achèvement du projet du génome humain (the human genome project) et l'introduction des techniques de génotypage haut débit vers le début du 21ème siècle ont permis le développement d'études d'association génomique (genome-wide association studies, GWAS) qui ont facilité la détection de certains loci dans le génome humain en association avec des maladies. Cependant, cette approche n'a pas été suffisamment puissante pour couvrir de façon complète les causes génétiques associées aux maladies rares et/ou complexes. Les instruments de séquençage de nouvelle génération (*NGS, next-generation sequencing*) sont devenus disponible à partir de 2005 et le nombre de publications scientifiques utilisant ces techniques a augmenté de manière exponentielle depuis 2010. Comme il existe non seulement une mais plusieurs méthodes de séquençage de nouvelle génération, le terme de NGS ne semble pas être adapté et le terme de séquençage haut débit (*HTS, hight throughput sequencing generation*) est actuellement préféré.

Dans le domaine de la génétique humaine, le développement des méthodes d'enrichissement ciblé comme le séquençage exomique (*WES, whole exome sequencing*) qui permet le séquençage sélectif des régions génomiques codantes (exons). Ces régions représentent approximativement 1% du génome mais sont concernées par la grande majorité des mutations pathogéniques. L'emploi du séquençage exomique a permis de réduire la quantité des séquences à générer et de ce fait, le rapport coût-efficacité de la technique (Bamshad et al., 2011). L'utilisation du séquençage exomique dans la recherche translationnelle a significativement amélioré les rendement du diagnostic génétique dans plusieurs domaines de la médecine (Petersen et al., 2017). Le WES est alors devenu une méthode de choix pour identifier les nouvelles causes génétiques responsables de maladies héréditaires et il est actuellement largement utilisé dans la pratique clinique de routine dans beaucoup de laboratoires. Avec le progrès continu apporté aux techniques de séquençage haut débit, l'analyse bio-informatique et le stockage des données, le WES va être prochainement

remplacé par le séquençage du génome complet (*WGS, whole genome sequencing*) qui grâce à sa plus grande couverture va permettre d'améliorer encore plus le rendement diagnostique.

4.1.2- Principes du séquençage et étapes de l'analyse bio-informatique des données

Afin d'identifier et de sélectionner un excellent gène candidat dans le phénotype étudié, plusieurs étapes sont nécessaires. Globalement, ce processus se divise en deux étapes principales qui sont schématisées dans la **Figure 39** et qui sont détaillées ci-dessous.

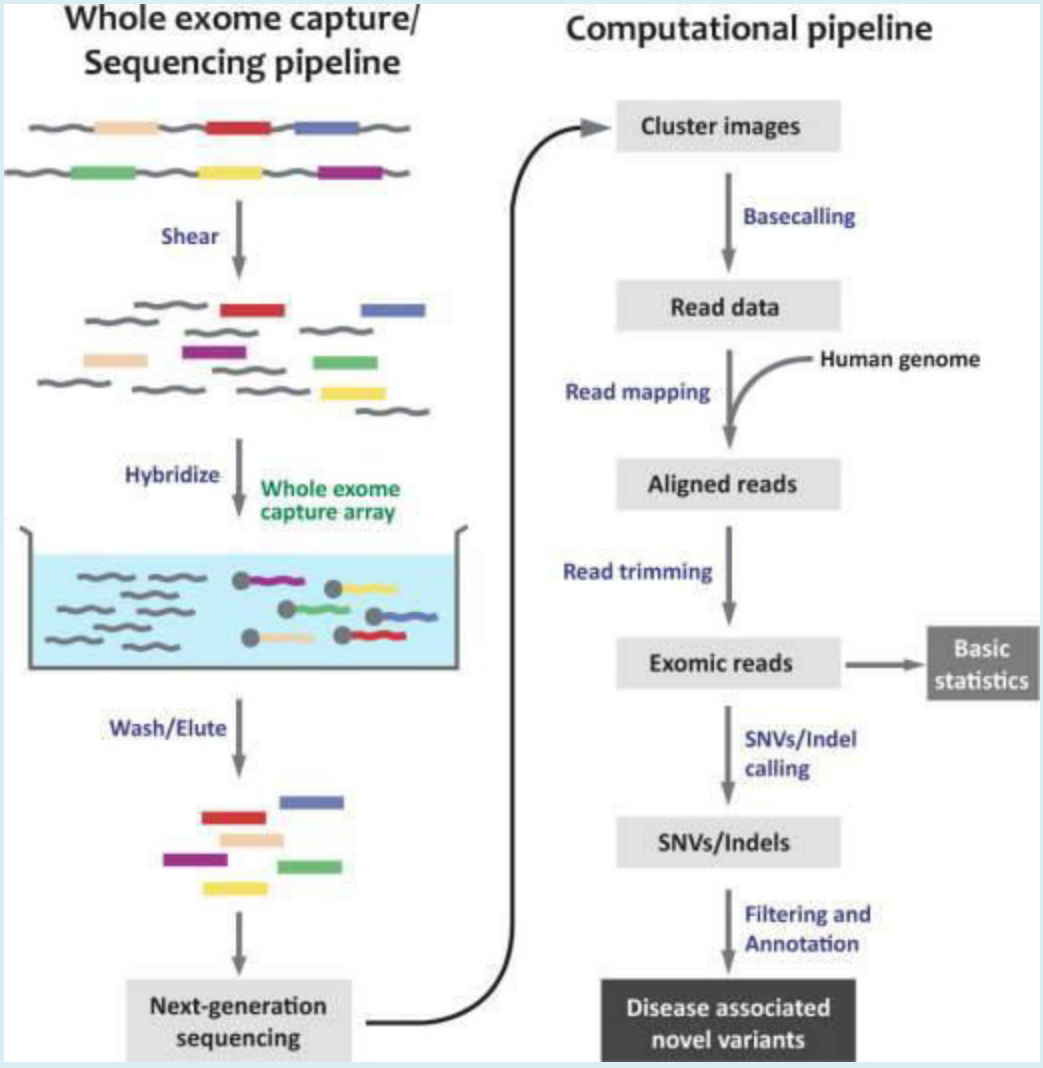


Figure 39. Vue d'ensemble de la stratégie d'identification d'un variant candidat par séquençage exomique.
(D'après Goh and Choi, 2012).

4.1.2.1- Rappel : le séquençage conventionnel de Sanger

Le séquençage Sanger est une méthode introduite en 1977 par F. Sanger, S. Nicklen, et A. R. Coulson pour déterminer l'ordre séquentiel des quatre nucléotides Adénosine (A), Guanine (G), Cytosine (C) et Thymine des acides désoxyribonucléiques (ADN) (Sanger et al., 1977). Le principe de la méthode est celui d'une synthèse enzymatique sélective. Une étape d'amplification de l'ADN par PCR est réalisée en amont de la réaction de séquence pendant laquelle, les amplicons sont produits. Cette réaction utilise une polymérase spécifique dépourvue de son activité exonucléase. Le mixte est composé de 4 désoxyribonucléotides (dATP, dGTP, dCTP et dTTP) et un seul type de didésoxyribonucléotides (ddATP, ddGTP, ddCTP ou ddTTP). Les ddNTPs rentrent en compétition avec les dNTPs et comme ils sont dépourvus de leur groupe hydroxyle 3'-OH, ils agissent comme des bloqueurs d'élongation et donc de terminateurs de chaine. La réaction est réalisée 4 fois en présence d'un ddNTP différent dans chaque réaction. Il en résulte alors un mélange de fragments de tailles croissantes chacun terminé par un ddNTP donnée correspondant au nucléotide de la séquence cible. Pour identifier la position de chaque ddNTP et donc les positions des nucléotides correspondants dans la séquence à déterminer, les fragments subissent une migration différentielle selon leurs tailles en électrophorèse. Grâce aux progrès techniques continus apportés à cette méthode pendant plusieurs décennies, le procédé est devenu plus simple et rapide. Grâce à l'utilisation de ddNTPs marqués par des fluorophores différents, une seule réaction de séquence est réalisée. Quant à la séparation des fragments, c'est l'électrophorèse capillaire qui est employée actuellement en couplage avec une lecture informatique instantanée des séquences. L'automatisation de cette technique a permis, en 2001, de réaliser le premier séquençage du génome humain entier (Lander et al., 2001; Venter et al., 2001). Cependant, ceci a pris 13 ans pour sa réalisation et pour un coût total estimé à 2.7 billions de dollars américains. Les nouvelles technologies de séquençage du génome doivent alors répondre à trois questions cruciales :1) Comment augmenter la taille des fragments ? 2) Comment accélérer le débit des réactions ? Et 3) comment réduire le coût de la procédure ?

4.1.2.2- Le séquençage exomique

E. La préparation de l'ADN

Pour être séquencé dans son intégralité par les méthodes haut débit, le génome doit être digéré par des procédés enzymatiques, chimiques ou physiques, en fragments d'une

taille comprise entre 50 et 500 pb et ce sont ces petites séquences qui feront l'objet de lecture par le séquenceur de manière parallèle. Et c'est d'ailleurs pour cette raison que l'on a décrit souvent le séquençage haut débit comme un séquençage parallèle massif. Cette technologie a comme principaux avantages par rapport au séquençage Sanger, la capacité de lecture rapide d'un nombre conséquent de fragments en même temps et à cout réduit. Ces avantages pratiques et économiques ont permis une diffusion rapide de cette nouvelle technologie dans la recherche clinique.

F. L'enrichissement

Dans un souci de réduction de la masse de données à stoker et du coût de la procédure, des méthodes d'enrichissement se sont développées par la suite permettant de réaliser un séquençage massif mais ciblé sur certaines régions génomiques précises comme des panels de gènes d'intérêt dans des pathologies particulières ou encore l'ensemble des régions codantes du génome (l'exome). L'étude de l'exome est cliniquement intéressante. Bien que l'exome ne représente que 1% du génome humain, les mutations associées représentent près de 90% des causes responsables de maladies génétiques. Pour ces raisons, le séquençage exomique est considéré actuellement comme le standard pratique et accessible dans le cadre de recherche sur des pathologies génétiques et se révèle être un outil robuste pour l'identification de variants associés aux pathologies.

G. L'amplification et la réaction de séquence

Afin d'améliorer la qualité du signal recueilli, le séquençage est précédé par une étape d'amplification qui permet l'obtention de plusieurs milliers de clones pour chaque fragment d'ADN capturé. Cette étape est réalisée selon les technologies : soit sur une phase solide (Illumina) soit en phase liquide en émulsion (Ion Torrent). L'ensemble des clones produits seront séquencés en parallèle générant des signaux protoniques ou photoniques qui seront ensuite traduits en séquences nucléotidiques (Heather and Chain, 2016).

4.1.2.3- L'analyse bio-informatique

H. Objectifs et principes

L'objectif de l'utilisation du séquençage exomique dans cette thèse était d'identifier de nouveaux variants et gènes candidats dans l'infertilité masculine. Un des challenges de l'analyse des données obtenues est de sélectionner les variants responsables du phénotype de chaque individu parmi les autres variants comme ceux non pathogéniques ou

résultant d'une erreur technique du séquençage. En effet, le séquençage exomique permet d'identifier environ 25000 variants SNV (single nucleotide variant) et indels (petites insertions et délétions) par individu. Parmi ces variants, près de 95% sont déjà connus comme de simples polymorphismes dans la population régionale ou générale. Les autres variants peuvent être pathogènes et altérer la fonction de la protéine. De manière synthétique, la détection de ces variants à partir des données du séquençage exomique et la sélection du variant causal suivent trois étapes principales :

1. L'alignement des séquences des fragments (reads) sur un génome de référence ;
2. Localisation et appel des variants dans les génomes séquencés ;
3. Filtration des variants pour éliminer ceux de faible qualité et annotation des variants retenus avec des informations supplémentaires.

L'étape de filtrage est considérée comme une étape critique dans l'analyse des données du séquençage exomique. Nous avons défini des paramètres de filtrage par défaut qui peuvent être modifiés ultérieurement en fonction des besoins. Ces paramètres sont les suivants :

L'union des variants : lorsque des individus porteurs d'un même phénotype présentent des liens de parenté, seul les variants communs chez l'ensemble des individus seront retenus.

Génotypes des variants : Du fait de la forte consanguinité qui caractérise nos patients, nous postulons que les anomalies génétiques associées à leur infertilité sont transmises selon un mode autosomique récessif. Dans ce cas, la présence de deux allèles mutés est nécessaire. Nous éliminons donc les variants hétérozygotes et nous conservons les variants homozygotes et hétérozygotes composites. Dans ce dernier cas, il est impossible de déterminer le phasage des variants (en *cis* : sur le même allèle ou en *trans* : sur deux allèles différents) et une analyse familiale s'impose. En effet, seuls les variants hétérozygotes en *trans* seront retenus.

Impact du variant : afin de ne conserver que les variants les plus susceptibles d'être délétères et d'impacter la fonction de la protéine, seuls sont retenus, les variants présents dans les exons codants y compris les mutations faux-sens prédites comme délétères par des outils bio-informatiques tels que PolyPhen et SIFT, et les régions d'épissage flanquant les exons et en particulier, les sites consensus accepteurs et donneurs d'épissage.

Fréquence des variants : Dans les maladies autosomiques récessives, en raison de la pression de sélection négative, les variants délétères attendus sont généralement rares. Les variants fréquemment retrouvés dans la population générale (fréquence allélique ≥1%) sont donc éliminés.

Présence des variants chez des sujets contrôles : Ce filtre permet de détecter les variants faux-positifs résultant d'artéfacts liés aux différentes étapes en amont de l'analyse bio-informatique. Il permet aussi, de mettre en évidence les variants propres à une population donnée. En effet, les variants homozygotes communs chez deux groupes d'individus présentant des phénotypes différents sont moins susceptibles de causer le phénotype des individus étudiés et sont donc rejetés.

I. Cas particulier : détection des variants de nombre de copies (CNVs, copy number variants)

Lorsque la couverture d'une région est inférieure ou supérieure à celle des régions voisines ou de la même région provenant d'un génome de référence, cette information pourrait être utilisée pour appeler une délétion ou une duplication. Pour les données du séquençage exomique, cette analyse s'avère difficile en raison de la grande hétérogénéité des couvertures relatives aux différents reads (profondeur de lecture). Parmi les logiciels d'analyse des CNV à partir des données d'exomes, on retrouve le logiciel ExomeDepth. Ce logiciel permet de réaliser des moyennes des profondeurs de lecture par exon cible. Après une normalisation, il compare les moyennes obtenues pour chaque exon par rapport à la couverture du même exon séquencé chez environ une dizaine de contrôles. Nous avons utilisé ce logiciel et avons identifié avec succès une large délétion intragénique d'environ 8 kb qui emporte deux exons du gène WDR66. Les résultats détaillés ainsi que leur discussion sont présentés dans l'article 3 de ce manuscrit.

4.2- Validation des gènes candidats chez le modèle murin CRISPR/Cas9

4.2.1- Introduction

En utilisant le WES, l'identification de potentiels gènes candidats dans un phénotype donné est devenue actuellement relativement facile, rapide et abordable. La difficulté vient surtout du nombre important de variants identifiés chez chaque individu et la sélection du variant causal. La validation des variants candidats et des gènes associés aux phénotypes explorés représente actuellement un défi majeur. Pour cela, l'utilisation des modèles cellulaires et animaux est critique pour confirmer le lien entre le gène muté et le phénotype observé et caractériser la physiopathologie associée aux anomalies génétiques. En raison de la forte homologie génétique et physiologique entre la souris et l'Homme, la souris représente un des meilleurs modèles pour valider les gènes candidats, en particulier dans l'infertilité masculine.

4.2.2- Historique de la transgénèse murine

Les souris transgéniques étaient traditionnellement générées par mutagenèse insertionnelle ou par des stratégies de transgénèse ciblée (Wang et al., 2013). En utilisant la méthode conventionnelle de transgénèse ciblée, les mutations sont introduites par intégration d'une matrice de réparation exogène par recombinaison homologue dans les cellules souches embryonnaires. L'injection de ces cellules modifiées dans les blastocystes permet alors la production de souris chimériques présentant un mosaïcisme germinal. (Capecchi, 2005). Ce processus est cependant laborieux et couteux puisque la fréquence d'évènements désirables est extrêmement rare (<0.01%) (Capecchi, 1989) et c'est d'ailleurs la raison pour laquelle d'autres technologies alternatives se sont développées pour accélérer ce processus. Récemment, l'utilisation des nucléases artificielles cible-dépendantes comme les ZFNs (zinc finger nuclease) et les TALENs (transcriptional activator-like effector nucleases) a permis la manipulation et l'édition du génome sans faire recours aux cellules souches embryonnaires (Barman et al., 2017). Ces nucléases sont composées de domaines spécifiques de liaison sur l'ADN et couplées avec des nucléases de restriction non spécifiques. Il a été observé que l'introduction de ces nucléases par une injection dans le zygote murin pouvait causer une cassure double-brin dans un site génomique spécifique. Cependant, les difficultés pratiques associées au design et à la synthèse de ces nucléases ont limité la diffusion de leur utilisation.

4.2.3- Le système CRISPR/Cas9

Bien que la technique CRISPR (clustered regularly interspersed short palindromic repeats)/Cas9 (CRISPR-associated protein 9) soit développée en parallèle des systèmes ZFNs et TALENs, son utilisation s'est rapidement et largement diffusée puisqu'elle ne nécessite pas la synthèse d'une protéine spécifique pour chaque cible génomique (Sapranauskas et al., 2011).

Le technologie CRISPR/Cas9 est dérivée d'un remarquable système immunitaire microbien adaptatif (Barrangou et al., 2007) retrouvé dans la plupart des bactéries et des archae. Chez ces organismes, les loci CRISPR sont composés de séquences palindromiques répétées et séparées par de courtes séquences d'ADN dérivées d'éléments infectieux antérieurs comme les phages et les plasmides. Ces éléments sauvegardés comme des *protospacers* dans les loci CRISPR représentent la mémoire immunitaire de ces organismes (Barrangou et al., 2007). La transcription de ces loci et le processing de l'ARN transcrit permettent de produire deux types d'ARNs, les crRNAs (CRISPR RNAs) et les tracrRNAs (trans-activating crRNAs). Ces deux ARNs sont hybridés pour pouvoir se lier à la Cas9 et former un complexe effecteur Cas9-crRNA-tracrRNA (Deltcheva et al., 2011). Ce complexe est guidé par une séquence guide de 20 nucléotides dans le crRNA et cause une cassure double-brin dans la région génomique cible (Garneau et al., 2010). Il a été démontré que l'hybride crRNA-tracrRNA pouvait être synthétisé sous forme d'une seule molécule d'ARN chimérique (sgRNA, single guide RNA) capable de guider la Cas9 vers la séquence cible (Jinek et al., 2012). En plus du sgRNA, la Cas9 nécessite, pour cliver la région cible, la reconnaissance préalable d'un motif PAM (protospacer adjacent motif), une séquence de trois nucléotides directement adjacente de l'extrémité 3' du brin complémentaire de la séquence cible. Dans le cas du Streptococcus pyogenes, la Cas9 reconnait la séquence 5'-NGG-3'(N correspondant à n'importe quelle base) comme un motif PAM (Doudna and Charpentier, 2014).

En pratique, le sgRNA et la Cas9 sont introduits dans la cellule soit par micro-injection soit par électroporation. Une fois dans la cellule, le sgRNA s'hybride avec la Cas9 et la guide vers la séquence génomique cible pour induire une cassure double-brin généralement trois nucléotides avant le motif PAM. La cassure générée peut être auto-réparée par les mécanismes intrinsèques de la cellule manipulée soit par le système de réparation aléatoire,

NHEJ (nonhomologous end joining) ou le système de réparation dirigée, HDR (homology-directed repair).

Quand la réparation se fait par le système NHEJ, des mutations de type InDels (petites insertions et délétions) peuvent se produire. Si la séquence cible est contenue dans une région codante et si le nombre de nucléotides concernés n'est pas un multiple de trois, on parle alors de mutation « frameshift » qui induisent un décalage du cadre de lecture et l'instabilité de l'ARNm. L'ARN transcrit à partir de ce gène muté peut être reconnu et dégradé par le système NMD (non-sense mediated mRNA decay) ou être transcrit en protéine tronquée non fonctionnelle. Ce mécanisme est intéressant pour supprimer la fonction d'un gène donné. La réparation par le système HDR nécessite la co-injection d'une matrice de réparation sous forme d'un simple-brin d'ADN (ssDNA single stranded DNA) d'une taille moyenne de 100 pb (Hsu et al., 2014). Cette matrice doit présenter une homologie avec les séquences flanquant la région cible. Ce mécanisme est intéressant pour introduire des modifications désirées dans la région cible comme des mutations ou une séquence codant pour un tag fluorescent ou immunogène.

Nous avons utilisé cette nouvelle technologie pour produire des souris knock-out et knock-in pour valider les gènes candidats dans le phénotype MMAF. La méthodologie, les résultats et leur discussion sont présentés dans l'article suivant (Article 1).

4.2.4- Article 1 : Creation of knock out and knock in mice by CRISPR/Cas9 to validate candidate genes for human male infertility, interest, difficulties and feasibility

Zine-Eddine Kherraf, Beatrice Conne, Amir Amiri-Yekta, Marie Christou-Kent, Charles Coutton, Jessica Escoffier, Serge Nef, Christophe Arnoult & Pierre F. Ray.

Molecular and Cellular Endocrinology. Mars 2018.

Contexte et principaux résultats de l'étude :

Le séquençage haut débit (HTS, high throughput sequencing) et le CRISPR/Cas9 sont deux technologies récentes qui sont en train de révolutionner la recherche clinique et biologique. L'HTS permet d'identifier de nouveaux variants génétiques dans différentes pathologies et le CRISPR/Cas9 permet de créer des modèles cellulaires et animaux pour valider l'impact de ces variants, caractériser la fonction des gènes candidats identifiés et tester de nouvelles approches thérapeutiques ciblées pour corriger les anomalies moléculaires et cellulaires associées.

Nous avons analysé une cohorte de 78 sujets présentant des anomalies morphologiques multiples des flagelles spermatiques (MMAF), une forme sévère d'infertilité masculine. En utilisant, le séquençage exomique (WES, whole exome sequencing), nous avons identifié des mutations homozygotes dans des gènes candidats autosomiques chez 63% des sujets analysés. Afin de valider ces gènes candidats dans le phénotype MMAF et caractériser leurs fonctions dans la biogenèse du flagelle au cours de la spermiogénèse, nous avons produit avec la technique CRISPR/Cas9 quatre lignées de souris knock-out (KO) et une lignée knock-in (KI). Avec cette technique, nous avons pu obtenir 31% de souris nées vivantes et présentant des mutations dans les séquences génomiques ciblées. Tous les évènements identifiés étaient des insertions ou des délétions localisées à proximité du PAM (protospacer adjacent motif), une séquence de trois nucléotides directement adjacente à la région cible. De manière surprenante, nous avons observé des taux élevés de mosaïcisme germinal avec 30% des animaux de la génération F1 portant des mutations différentes de celles identifiées chez leurs parents à partir de tissus somatiques (queue). Ceci montre que le système CRISPR/Cas9 reste actif au cours des divisions cellulaires successives à partir du zygote injecté.

Au total, nous avons produit 5 lignées de souris transgéniques différentes et dans chaque cas, un animal homozygote pouvait être obtenu dans un délai de 6 mois. Ces résultats montrent que la combinaison du WES et du CRISPR/Cas9 est une stratégie robuste et efficace pour identifier et valider les gènes candidats responsables d'infertilité masculine.

Molecular and Cellular Endocrinology 468 (2018) 70–80

Contents lists available at ScienceDirect

Molecular and Cellular Endocrinology

journal homepage: www.elsevier.com/locate/mce

Creation of knock out and knock in mice by CRISPR/Cas9 to validate candidate genes for human male infertility, interest, difficulties and feasibility

Zine-Eddine Kherraf [a, b], Beatrice Conne [c], Amir Amiri-Yekta [a, b, d], Marie Christou Kent [a], Charles Coutton [a, e], Jessica Escoffier [a], Serge Nef [c], Christophe Arnoult [a, 1], Pierre F. Ray [a, b, *, 1]

[a] Genetic Epigenetic and Therapies of Infertility, Institute for Advanced Biosciences, Inserm U1209, CNRS UMR 5309, Université Grenoble Alpes, F-38000, Grenoble, France
[b] UM GI-DPI, CHU Grenoble Alpes, Grenoble, F-38000, France
[c] Department of Genetic Medicine and Development University of Geneva Medical School, CH 1211, Geneva 4, Switzerland
[d] Reproductive Biomedicine Research Center, Royan Institute for Reproductive Biomedicine, ACECR, Tehran, PO Box 16635-148, Iran
[e] UM de Génétique Chromosomique, CHU de Grenoble, Grenoble, F-38000, France

ARTICLE INFO

Article history:
Received 12 October 2017
Received in revised form
2 March 2018
Accepted 3 March 2018
Available online 6 March 2018

Keywords:
High throughput sequencing
CRISPR/Cas9
Spermatogenesis
Sperm flagella
Knock out mice

ABSTRACT

High throughput sequencing (HTS) and CRISPR/Cas9 are two recent technologies that are currently revolutionizing biological and clinical research. Both techniques are complementary as HTS permits to identify new genetic variants and genes involved in various pathologies and CRISPR/Cas9 permits to create animals or cell models to validate the effect of the identified variants, to characterize the pathogeny of the identified variants and the function of the genes of interest and ultimately to provide ways of correcting the molecular defects.

We analyzed a cohort of 78 infertile men presenting with multiple morphological anomalies of the sperm flagella (MMAF), a severe form of male infertility. Using whole exome sequencing (WES), homozygous mutations in autosomal candidate genes were identified in 63% of the tested subjects. We decided to produce by CRISPR/cas9 four knock-out (KO) and one knock-in (KI) mouse lines to confirm these results and to increase our understanding of the physiopathology associated with these genetic variations. Overall 31% of the live pups obtained presented a mutational event in one of the targeted regions. All identified events were insertions or deletions localized near the PAM sequence. Surprisingly we observed a high rate of germline mosaicism as 30% of the F1 displayed a different mutation than the parental event characterized on somatic tissue (tail), indicating that CRISPR/Cas9 mutational events kept happening several cell divisions after the injection.

Overall, we created mouse models for 5 distinct loci and in each case homozygous animals could be obtained in approximately 6 months. These results demonstrate that the combined use of WES and CRISPR/Cas9 is an efficient and timely strategy to identify and validate mutations responsible for infertility phenotypes in human.

Contents

1. Introduction 71
2. Material and methods 73
 2.1. CRISPR/Cas9-mediated mice genome edition 73
 2.2. Mice genotyping strategy 74

* Corresponding author. Genetics of Infertility and PGD, Institut de Biologie et Pathologie, R+2, CHU Grenoble Alpes 38 043, Grenoble Cedex 9, France.
E-mail address: pray@chu-grenoble.fr (P.F. Ray).
[1] Shared leadership.

https://doi.org/10.1016/j.mce.2018.03.002
0303-7207/

3. Results 74
3.1. Genetic results 74
3.2. Creation of F0 KO mice for *Cfap43*, *Cfap44*, *FlagC*, *FlagF* and of a HA tagged *FlagF* gene using the CRISPR/Cas9 technique 75
3.2.1. Creation of KO mice for *Cfap43* 75
3.2.2. Creation of KO mice for *Cfap44* 75
3.2.3. Creation of KO mice for *FlagC, FlagF and tagged mice FlagF-HA* 75
3.3. Size and position (in regards to the PAM) of the obtained mutations 76
3.4. Overall results 76
4. Discussion 77
4.1. Frequent occurrence of F0 mosaic 77
4.2. Localization of the mutational events and off target mutations 78
4.3. Technical difficulties 78
4.4. Timing of CRISPR/Cas9 gene editing projects 78
5. Conclusion 78
Conflicts of interest 79
Acknowledgments 79
Supplementary data 79
References 79

1. Introduction

Fertility disorders affect more than 15% of couples worldwide and are considered as a major health concern by the world health organization (WHO). Approximately half of these cases are related to male factors but despite the large number of genes specifically expressed in testes during spermatogenesis (Soumillon et al., 2013) and the multitude of KO infertile male mice described in the literature, only few genes have so far been associated with human male infertility and its molecular basis remains largely unknown (Coutton et al., 2015). The recent development of new high throughput sequencing (HTS) techniques and efficient targeted enrichment methods such as whole exome sequencing (WES) now allow the sequencing of all the coding sequences of any human being. The increased use of WES in human genetics allowed the identification of a large number of new Mendelian disease genes (Alkuraya, 2016) and the technique is now increasingly used in the field of male infertility (Ray et al., 2017). Using WES, the identification of plausible candidate genes for a particular phenotype is now relatively easy, quick and affordable. The difficulty comes from the large number of likely deleterious variants identified for each patient which renders difficult the identification of the causal mutation(s). For each candidate variant/gene, the main challenge is to validate the direct link between the identified variant and the patient's phenotype. Cellular or animal models are thus critical to confirm the link between a gene defect and a phenotype and to characterize the physiopathology associated with the identified genetic anomalies. Because of the high genetic homology between mice and men, the mouse model represents one of the best models for confirming and characterizing gene defects identified in human and this is particularly true for male reproductive disorders.

Mutant mice were traditionally generated using insertional mutagenesis or gene-targeting strategies (Wang et al., 2013). Using conventional gene targeted strategy, mutations are introduced by integrating exogenous repair templates through homologous recombination in embryonic stem (ES) cells. Injection of modified ES cells into blastocysts then allows the production of chimeric mice (Capecchi, 2005). This process is however laborious and expensive as the frequency of desired events was extremely rare (<0.01%) (Capecchi, 1989) and this is the reason why alternative technologies bypassing these difficulties were needed to accelerate this process. Recently, the use of artificial site-specific nucleases, such as zinc finger nuclease (ZFN) and transcriptional activator-like effector nucleases (TALENs) has enabled genome editing without the use of ES cells (Barman et al., 2017). However, difficulties associated with nucleases design and synthesis limited the widespread adoption of these technologies. The CRISPR (clustered regularly interspersed short palindromic repeats)/Cas9 (CRISPR-associated protein 9) technique which does not require the synthesis of target specific proteins has recently emerged as the method of choice for producing mutations in cells or animal models (Sapranauskas et al., 2011). The system requires the use of a single guide RNA (sgRNA) which will define the nucleotide sequence to be targeted and the Cas9 nuclease which will create a double strand break (DSB) underneath the RNA/DNA hybrid complex (Fig. 1A). The DSBs are usually repaired by the error-prone non-homologous end joining (NHEJ) mechanism that often creates small insertions and deletions (indels) thereby initiating truncating mutations (Fig. 1B). DSB can also be repaired by homology-directed repair (HDR) (Hsu et al., 2014) if a DNA template with homology to the sequence surrounding the DSB is found during the repair mechanism (Fig. 1C). The co-injection of a single stranded DNA (ssDNA) molecule homologous to the sequences flaking the targeted region will therefore favor HDR over NHEJ and will permit the replacement of the original sequence by the newly introduced sequence. Using a ssDNA slightly modified from the target genomic DNA permits to create a specific knock in (KI) or to insert a defined sequence that will be transcribed and translated. For the creation of mutant mice, the sgRNA and the Cas9 protein and potentially a ssDNA can be directly microinjected in a zygote to create a mutation at a predefined locus (Fig. 2). Homozygous mutant animals can be obtained after 1 or 2 back crossing permitting to create specific KO animals in only a few months (Fig. 2).

The recruitment of patients presenting with specific morphological anomalies of the spermatozoa enabled us to identify several genes associated with specific infertility phenotypes. We have demonstrated that approximately 80% of patients with large headed multiflagelled spermatozoa have a defective *AURKC* gene (Dieterich et al., 2009, 2007) and that in excess of 70% of men with round acrosomeless spermatozoa have a homozygous deletions of the *DPY19L2* gene (Coutton et al., 2012; Harbuz et al., 2011). The absence of spermatozoa in the ejaculate can also be caused by post meiotic defects and WES allowed to demonstrate that mutations in *SPINK2*, coding for a protease inhibitor targeting the acrosin could in fact prevent acrosome formation and block the spermiogenesis process (Kherraf et al., 2017). We now focus our research on

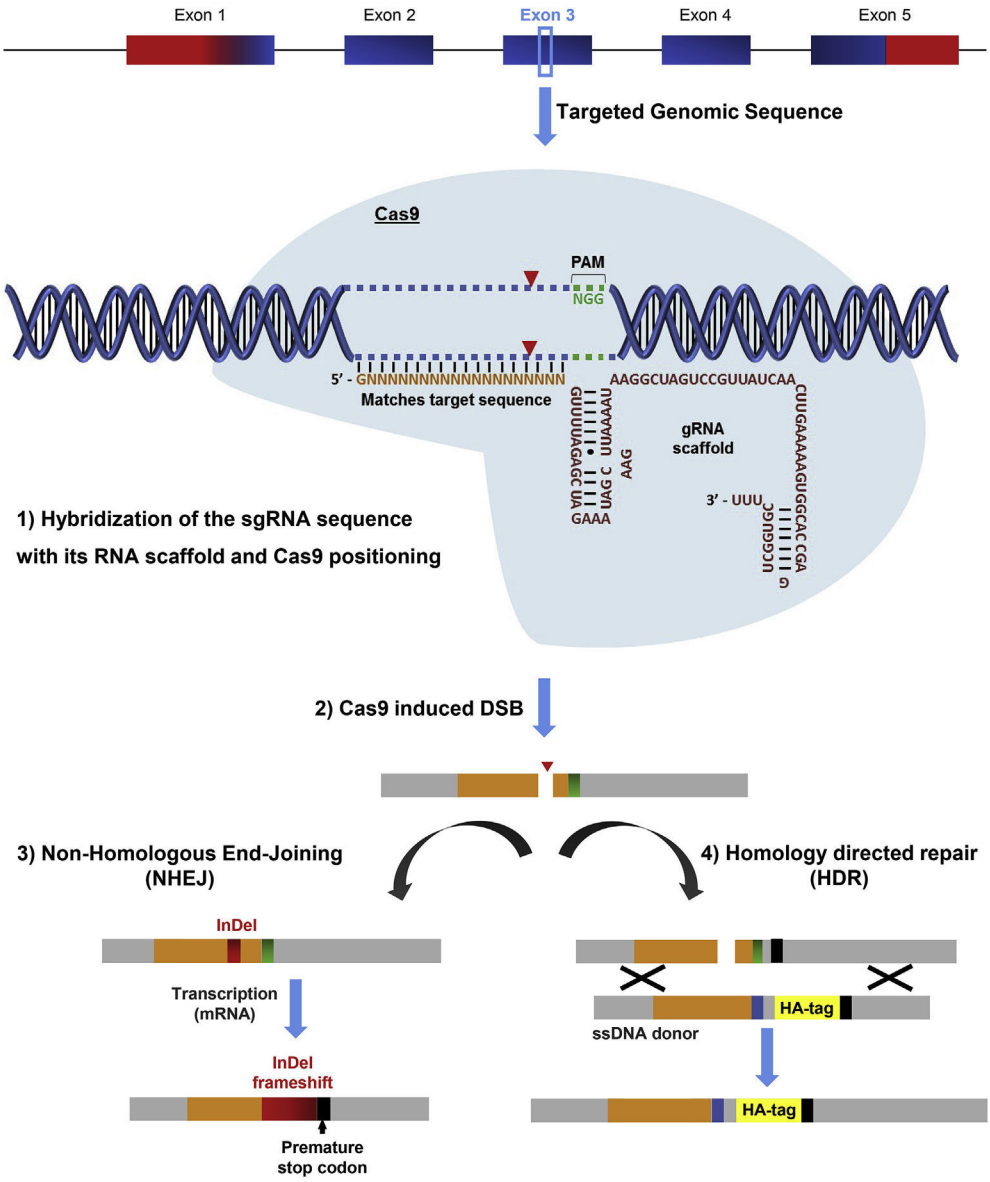


Fig. 1. The CRISPR/Cas9 system for genome editing.
1) The Cas9 endonuclease is directed to the targeted sequence by base pairing between the specific sgRNA (orange) and its target within the genomic DNA sequence. A "NGG" PAM sequence (Green NGG) located immediately downstream from the target sequence in the non-complementary strand is required for Cas9 recognition.
2) The Cas9 induces a DSB (red arrow-head) within the sgRNA targeted genomic DNA (orange box), normally between the third and the fourth bases from the PAM (orange box).
3) The error-prone NHEJ repair mechanism induces unpredictable insertions or deletions (InDels) represented by a red box, close to the initial DSB and usually within the sgRNA targeted sequence (orange box). The majority of InDels (which are not a multiple of 3 nucleotides) induce a frameshift that introduces a premature stop codon (black rectangle) which induces mRNA degradation (by non-sense mediated mRNA decay) and/or give a truncated non-functional protein. In most cases a null allele is obtained.
4) Co-injection of a ssDNA template homologous to the targeted sequence permits to repair the ds break by homology directed repair (HDR). Here the ssDNA template contains a small sequence (27 bp) coding for an HA-tag (yellow box) inserted immediately at the end of the protein of interest just before the stop codon (black box) to add a 9 amino acid tag to the targeted protein to facilitate its detection in-vivo. Compared to the targeted sequence, the ssDNA is also modified at the PAM sequence, the NGG (green box) is replaced by a TTG (blue box) to prevent new targeting by the sgDNA and ds breaks of the modified allele.

subjects presenting with Multiple Morphological Anomalies of the sperm Flagella (MMAF) characterized by the presence of short, bent, and coiled flagella and flagella of irregular width or absent mainly due to numerous ultrastructural defects of the axoneme leading to terato-asthenozoospermia. In 2014, we identified pathogenic mutations in the dynein axonemal heavy chain 1 (*DNAH1*) gene in 28% of the MMAF patients analyzed (Ben Khelifa et al., 2014). This result indicates that MMAF is genetically heterogeneous phenotype and that *DNAH1* is likely to be only one of the genes involved in this phenotype. Analysis of familial MMAF cases subsequently allowed us to confirm the implication of *DNAH1* in this syndrome (Amiri-Yekta et al., 2016). We then carried out WES analysis on 78 patients presenting with severe MMAF phenotype and identified *CFAP43* and *CFAP44* as two other genes involved in this phenotype (Coutton et al., 2018). Others have also reported mutations in *CFAP43* and *CFAP44* as important causes of MMAF in human (Tang et al., 2017; Sha et al., 2017). Further analysis of WES data allowed us to identify 9 other candidate genes which are currently being analyzed (data not published). To investigate the role of all these newly identified genes we performed gene

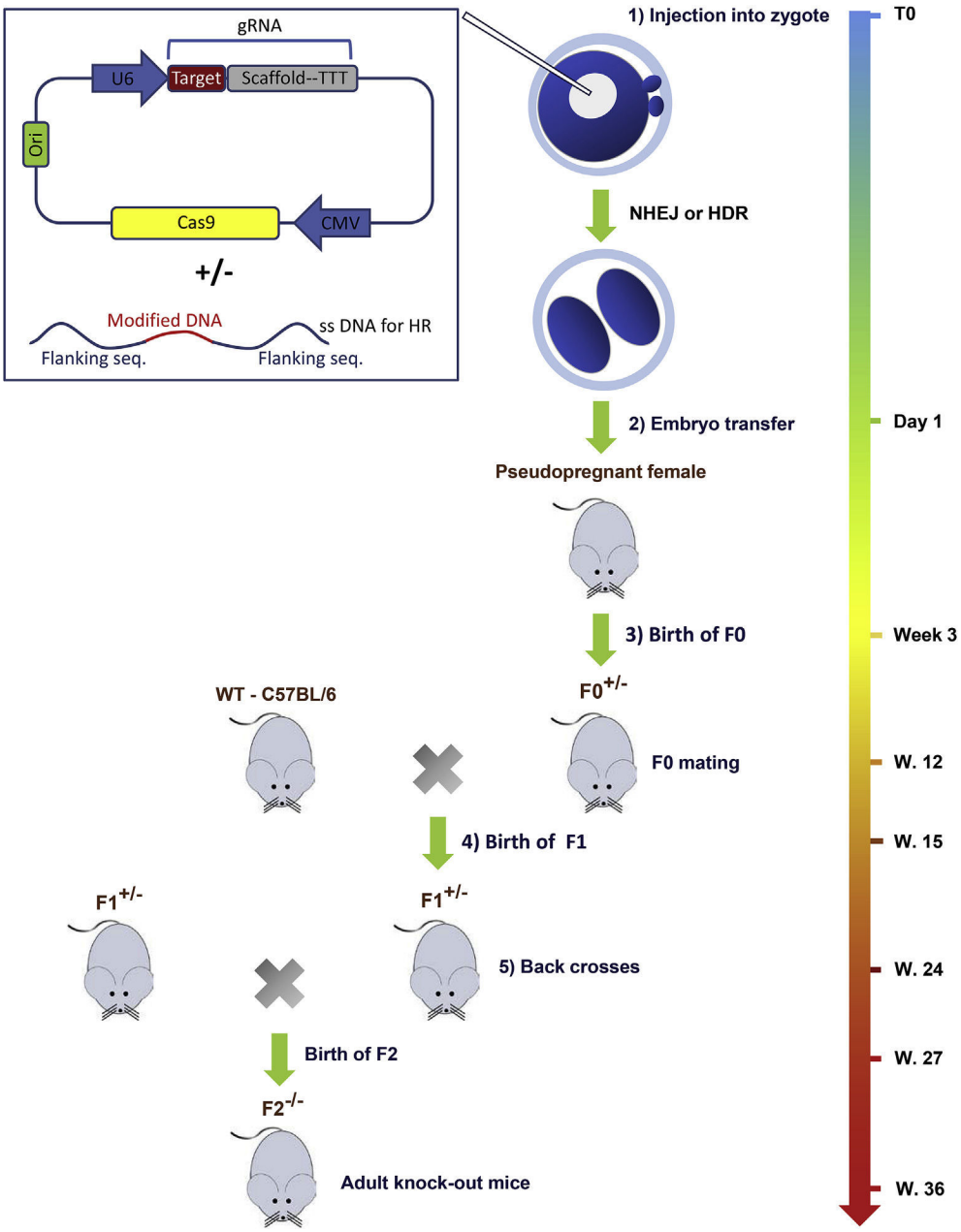


Fig. 2. Creation of KO/KI mice using the CRISPR/Cas9 system.
1) A minimal injection will contain a eukaryotic plasmid allowing the transcription of the sgRNA and the Cas9 (box on the left). The Cas9 protein will then be produced by the oocyte endogenous machinery. Knock in (KI) production can be achieved by the coinjection of a single stranded (ss) DNA for HDR. For insertion of a 5′ Tag, the sgRNA target (red line) is centered on the stop codon (TGA) marking the end of the coding sequence. The plasmid and, if necessary, the ssDNA are microinjected into the pronuclei of mice zygotes.
2) Obtained embryos are transferred into pseudopregnant females.
3) Genotyping of the obtained pups by Sanger sequencing allows the identification of founder mice (F0) carrying frameshift indels (or the correctly inserted sequence of choice). Because CRISPR/Cas9 system can remain active after several mitotic divisions founder mice can be mosaics carrying different mutations.
4) Breeding selected founder mice with wild-type (WT) C57BL/6J mice allows alleles segregation and transmission to the F1 mice generation. Sanger sequencing allows the identification of F1 mice with the same heterozygous frameshift indel.
5)Breeding these mice together leads to the production of F2 mice normally comprising 50% of heterozygotes, 25% of mutated homozygotes and 25% of WT mice. The mouse lineage is established and can be maintained for phenotypic characterization and further functional work.

invalidation in mouse using CRISPR/Cas9 technology for *Cfap43* and *Cfap44* and two of the 9 newly identified genes referred as *FlagC* and *FlagF*. Furthermore, as we could not obtain any specific antibodies for *FlagF* by conventional methods, we have generated a mouse lineage containing a HA-tagged *FlagF* gene. We describe here the techniques used and the results obtained to create these 5 transgenic mouse lines in a context of male infertility.

2. Material and methods

2.1. CRISPR/Cas9-mediated mice genome edition

The aim of this work was to generate 4 mouse lineages invalided for the *Cfap43*, *Cfap44*, *FlagC* and *FlagF* genes, respectively, to validate the candidate genes by assessing the spermatogenesis of the

corresponding knock-out mice. Additionally, the generation of a knock in (KI) was also carried out consisting of inserting 27 nucleotides coding for the HA tag at the end of the *FlagF* gene.

Generation of CRISPR/Cas9 edited mice was done at University of Geneva (UNIGE). All the procedures were done in Geneva until the birth of the modified litters. Mice were housed with unlimited access to food and water and were sacrificed by cervical dislocation after 8 weeks old, which means that they were pubescent and that their reproductive organs were fully established. All animal procedures were ran according to the Swiss and French guidelines on the use of animals in scientific investigations with the approval of the French local Ethical Committee (ComEth Grenoble N° 318, ministry agreement number #7128 UHTA-U1209-CA) and the Direction Générale de la Santé (DGS) for the State of Geneva.

All KO mice were created using the CRISPR/Cas9 technique as described in Figs. 1 and 2. To maximize the chances of generating deleterious mutations, two gRNA located in two distinct coding exons located at the beginning of the targeted genes were always used. For each gene the two gRNAs were inserted in two distinct plasmids, each plasmid also containing the Cas9 sequence. The Cas9 gene was driven by a CMV promoter and the gRNA and its RNA scaffold by an U6 promoter. Full plasmids (pSpCas9 BB-2A-GFP (PX458)) containing the specific sgRNA were ordered from Genescript (https://www.genscript.com/crispr-cas9-protein-crRNA.html). Both plasmids were co-injected in the zygotes' pronuclei as described in Fig. 2 at a concentration of 2.5 ng/μL. We note that we directly injected the plasmids as delivered by the supplier avoiding the *in vitro* production and purification of Cas9 proteins and sgRNA.

For the realization of the KI, one plasmid containing one sgRNA and the Cas9 was injected (5 ng/μL) with the single stranded DNA (50 ng/μL) for homology directed repair (Fig. 2). The single stranded DNA was 187 nt long and contained two 80 nucleotide-long arms homologous to the nucleotides located before and after *FlagF* stop codon, surrounding the 27 nucleotides encoding the HA Tag ending with a TAG stop codon (TAC CCA TAC GAT GTT CCA GAT TAC GCT TAG). The sgRNA sequence overlapped the native *Flagf* stop codon which was followed, 8 nucleotides further, by a TGG suitable as a Protospacer Adjacent Motif (PAM) sequence, which is an essential targeting component for efficient Cas9 binding and cleavage. PAM sequence consists of any nucleotide followed by two G (XGG). It is important to understand that if successful, the sgRNA will induce a double strand break near the site where the DNA modification is wanted (here at the end of the target gene to allow the insertion of a fusion HA Tag) and the break will either be repaired by HDR with the ssDNA and introduce the Tag, or, if unsuccessful by NHEJ. Due to the persistence of Cas9 and sgRNA several dsbreak can occur successively in the zygote or in early cleavage stage embryos. A successfully inserted Tag by HDR can therefore undergo a new DSB and repair by NHEJ thus altering the initial event and likely preventing the traduction of the Tag sequence. To limit this risk, in the ssDNA the TGG PAM sequence was changed to TTG. As the presence of a PAM sequence is necessary for DSB, this will normally prevent the occurrence of secondary events leading to the alteration of correctly inserted Tag. The synthetic ssDNA was purchased from IDT (https://www.idtdna.com/pages/products/dna-rna/ultramer-oligos). After the injection, zygotes were left for 4–6 h before being introduced into pseudopregnant host females and carried to term.

Edited founders were identified by Sanger sequencing from digit biopsies. Mice carrying the desired modification events (frameshift mutation or tag insertion) were crossed with C57BL6/J to ensure germline transmission and eliminate any possible mosaicism. Heterozygous animals with the same modification were then mated to generate homozygous offspring.

2.2. Mice genotyping strategy

DNA for genotyping was isolated from tail/digit biopsies. Tissues biopsies were digested in 200 μl of DirectPCR Lysis Reagent (Tail) (Viagen Biotech Inc, CA, USA) and 0.2 mg of proteinase K for 12–15 h at 55 °C followed by 1 h at 85 °C for proteinase K inactivation. The digested tissues were directly used for PCRs. Thirty-five cycles of PCR amplification were carried out with a hybridization temperature of 60 °C. Sequencing reactions were performed using BigDye Terminator v3.1 (Applied Biosystems). Sequence analyses were carried out on ABI 3130XL (Applied Biosystems). Sequences were analyzed using seqscape software (Applied Biosystems).

3. Results

3.1. Genetic results

WES had been performed on a total of 78 subjects with a MMAF phenotype (Coutton et al., 2018). After applying a series of computing filters permitting in particular to exclude all the non-rare variants (>1% in the ExAC database), we considered a gene to be a candidate gene if a homozygous loss of function variant was found in at least two subjects and if the gene was described to be highly expressed in the testis. A total of 11 genes corresponding to these criteria were identified (Table 1). Ten patients had a homozygous loss of function variant in *CFAP43*, six in *CFAP44* and six in *DNAH1*, three genes previously described as MMAF related genes (Coutton et al., 2018). Eight additional undisclosed autosomal genes were found to harbor homozygous mutations in at least two patients and are referred as *FLAG A-H*. *FLAGA* was mutated in 7 subjects, *FLAGB* in 4 subjects, *FLAG C and F* in 6 patients and *FLAG G* and *H* were found to be mutated in 4 patients. The variants localization and their respective impact on the FLAGC and FLAGF proteins are shown in Supp. Table 1. If all mutations are confirmed to be causal, a total of 49 subjects will have been diagnosed by WES with an overall efficiency of 63% (49/78).

To confirm the involvement of these genes in MMAF and to further characterize their function we decided to create KO animals for *Cfap43, Cfap44* (Coutton et al., 2018) and *FlagC and FlagF* (characterization still in progress). Seven subjects carried a homozygous *FlagA* deleterious variant but as this gene had already been described to be coding for an axonemal protein and the creation of KO animals was not undertaken for this gene. Finally, KO for the other genes (*FlagB, D, E, G, H*) may be done at a later stage.

Table 1
List of autosomal genes identified as candidate after exome sequencing of 78 MMAF patients and number of patients harboring a homozygous deleterious homozygous mutation.

Gene	Number of mutated subjects	% of tested subjects
CFAP43 [a]	10	12.8
FLAGA	7	9.0
CFAP44[a]	6	7.7
DNAH1	6	7.7
FLAGB	4	5.1
FLAGC[a]	3	3.8
FLAGD	3	3.8
FLAGE	3	3.8
FLAGF[a,b]	3	3.8
FLAGG	2	2.6
FLAGH	2	2.6
Total	49	62.8

[a] Indicates the genes for which KO mice were produced by CRISPR/Cas9.
[b] The gene for which a HA Tag was inserted. Identity of FLAG (for Flagellar gene) A-F is currently undisclosed.

3.2. Creation of F0 KO mice for Cfap43, Cfap44, FlagC, FlagF and of a HA tagged FlagF gene using the CRISPR/Cas9 technique

3.2.1. Creation of KO mice for Cfap43

Two separate series of injections were performed to generate mutated mice carrying frameshift indels. Data related to injected/transferred zygotes were not recorded for these first two experiments. The first injection allowed the production of 12 animals but none carried a mutation in either of the targeted exons (Supp. Table 2). The second injection resulted in 9 births with 5 mutated mice. Two mice had heterozygous complex indels, probably caused by the presence of two different variants on both alleles (not characterized) in the target region 1 (TR1), two mice had a heterozygous deletion (delAAGG or delTGTGAAGGGC) in the target region 2 (TR2) and a fifth mouse had a homozygous mutation delG in the TR2 (Supp. Tables 1 and 2). By chance, this latter mouse was a male and enabled us to analyze the morphology of the sperm of a KO *Cfap43* animal without further crossing. All the observed spermatozoa were immotile with very short flagella, typical of the MMAF phenotype (Coutton et al., 2018). Although it was impossible to generate a lineage from this infertile male, this first result strongly suggested that *CFAP43* was one of the genes responsible of the MMAF syndrome. Because we could not rule out an off target effect, a new colony was established by mating the F0 mouse carrying the delAAGG variant with a C57BL/6J WT animal for phenotype confirmation and further investigations. Of the 15 F1 pups obtained in two separate litters, 5 (33%) inherited the genotyped parental mutation (Supp. Table 3) and three additional F1 (20%) harbored a four nucleotides insertion (insTCCA) that had not been detected in the mutant parent (Supp. Tables 1 and 2). Homozygous *Cfap43*$^{-/-}$ delAAGG F2 males generated from F1 animals were infertile. All spermatozoa showed similar defects to those observed in the initial F0 homozygous delG mouse and were short (approximately 1/5th of the normal size) or absent and did not show any movement (Fig. 3).

3.2.2. Creation of KO mice for Cfap44

A total of three injections were performed in order to invalidate this gene. Despite the large number of pups obtained (n = 26), only one mouse was mutated and had an insertion (insTCAGATA) in the TR1 (Supp. Tables 1 and 2). Backcrossing of this mouse permitted to obtain three F1 litters (Supp. Table 3). Out of the 27 newborns, 8 mice (30%) had inherited the parental mutation (insTCAGATA). Two additional mutations (delTAC and delTATA in same target region, TR1) were found in a total of six mice (23%). Breeding were done for InsTCAGATA and delTATA animals and homozygous males were generated for both these *Cfap44* mutations. All *Cfap44*$^{-/-}$ male mice were infertile. Their sperm flagella were of normal size but often presented irregular caliber of the midpiece (Fig. 3). All spermatozoa were however immotile, often with a mild vibrating pulse.

3.2.3. Creation of KO mice for FlagC, FlagF and tagged mice FlagF-HA

A total of five injections were carried out for *FlagC, FlagF and FlagF-HA* (Supp. Table 2). An average of 139 zygotes were injected per session. Seventy percent of the zygotes survived the 1-cell stage and could be transferred into pseudopregnant females. On average 30 embryos were transferred per female. The transfer of 479 injected embryos permitted the birth of 65 pups indicating that only 14% of the embryos transferred led to a live birth (Supp. Table 2). Three of the five injections realized permitted to obtain >50% mutated live pups (Supp. Table 2). At least two mutated pups were obtained following each injection session.

Only one *FlagC* targeted injection was performed which permitted to obtain two F0 heterozygous animals. The first carried two uncharacterized events on TR1 and the second had a heterozygous delTAC on TR1 and a 12 nucleotide deletion in TR2 (delAAACCCGTTAGC). Following the backcrossing of the second mouse, 7 out of 19 (37%) newborn inherited the parental mutated allele (delTAC + delAAACCCGTTAGC) and 5 (26%) presented an insT in TR1. *FlagC*$^{-/-}$ animals have just been produced. Homozygous male are unfertile and we observed that all spermatozoa present a

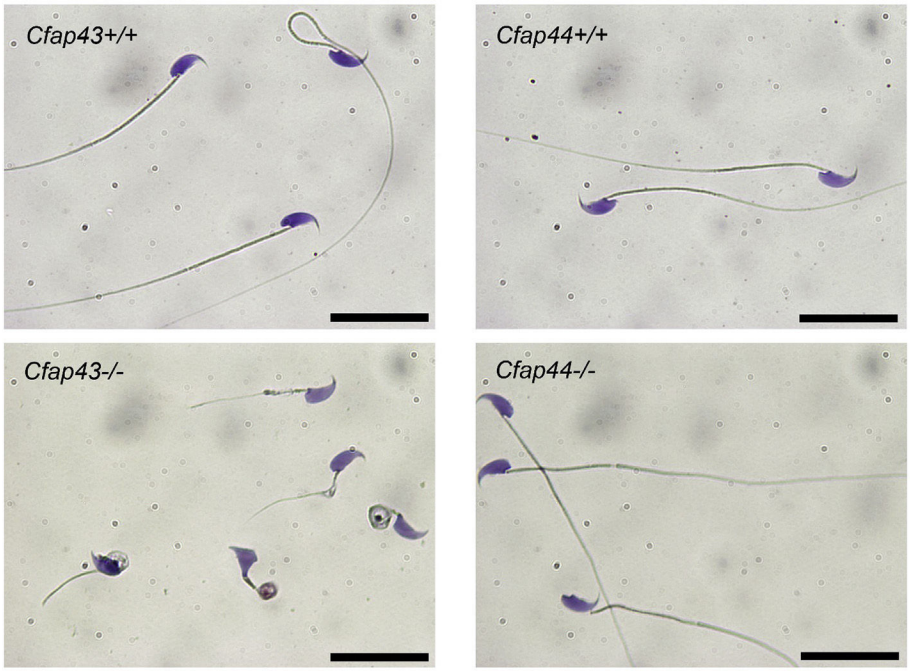


Fig. 3. Morphology of spermatozoa from *Cfap43* and *Cfap44* KO mice.
All spermatozoa from *Cfap43*$^{-/-}$ mice were short (approximately 1/5th of the normal size) or absent. Spermatozoa from *Cfap44*$^{-/-}$ mice were of normal size but often presented a kink at the end of the mid-piece. They did not show any progressive movement but sometimes a mild vibration. Scale bar 20 μm.

short flagella, a feature typical of the MMAF phenotype and very similar to what was observed on Cfap43$^{-/-}$ mice (data not shown). Creation of KO *FlagF* animals was also successful. Seventeen animals were created following two series of injections. A delG mutation was identified in one homozygous female mouse. Two animals carried a heterozygous event in TR1 (delTTCG, delCAGCTGGAA) and one had a complex event (uncharacterized) with the presence of a complex genomic rearrangement. Furthermore one F0 had a 12 nucletides insertion in TR2 (insAATGTACATAG). Generation of F1 animals carrying the delTTCG is still in progress.

Finally, we decided to insert an HA tag at the end of *FlagF* gene to compensate the lack of specific antibodies. The co-injection of the plasmid with the target RNA and the Cas9 gene along with the single stranded DNA for HDR was very successful in creating mutated pups. Six animals had the correct HA tag inserted by HDR, four of which had an indel on the other allele (Supp. Tables 1 and 2). Eight animals had a heterozygous indel created by NHEJ. In total 9 animals were compound heterozygotes (with two indels or one indel and one TAG insertion or perhaps sometimes a third event, possibly in mosaic. Due to the complexity of the events, we estimate that at least 12 events could not be characterized.

The two mice with only the HA tag insertion were bred with WT C57BL/6J animals to generate *FlagF* tagged mice. FlagF-HA homozygous males showed a normal fertility demonstrating that the tag does not interfere with the protein function. Studies of FlagF-HA localization and pulldown experiments are currently ongoing.

3.3. Size and position (in regards to the PAM) of the obtained mutations

Overall, following 10 series of injection targeting a total of 9 distinct sequences, we obtained (in F0 and F1) a total of 24 characterized individual *de novo* mutations (with a total of 20 different mutations). All the characterized mutations were indels and the HA tag was inserted correctly in 6 F0 animals. In addition, there were at least 18 bi-allelic heterozygous variants, often in mosaic that were not characterized. Deletions removed 1 to 21 nucleotides and insertions added 1 to 11 nucleotides. Approximately 80% of the indels were deletions. Half the events were small deletion of −1 to −4 nucleotides. The longest deletion and insertion were 15 and 11 nucleotides long respectively (Supp. Table 4 and Fig. 4).

All observed indels were located less than 20 nucleotides away from the PAM and the median point for indels was located 4 nucleotides before the PAM's first nucleotide. Only 4 events concerned some nucleotides located after the PAM (Supp. Table 4 and Fig. 5).

3.4. Overall results

Overall a total of 10 injection sessions were realized and all but three permitted to obtain at least one mutated pup. We can see that despite the three failures, we obtained an average of 3.4 mutated pups per injection session, highlighting the fact that the injections were very successful in producing mutated pups (Supp. Table 2).

Mutated animals were obtained for all genes investigated. Phenotypic characterization of the mutated animals has so far only been done for *Cfap43* and *Cfap44* genes. We obtained two independent mouse strains or homozygous animals for *Cfap43* and *Cfap44* with different insertions/deletions of a few nucleotides, all inducing a translational frameshift expected to lead to complete absence of the protein or production of a truncated protein. We observed in each case the same reproductive phenotype, so for *Cfap43* we restricted our study to a strain with a 4 bp deletion in the exon 21 (delAAGG) and for *Cfap44* we focused on a strain with a 7

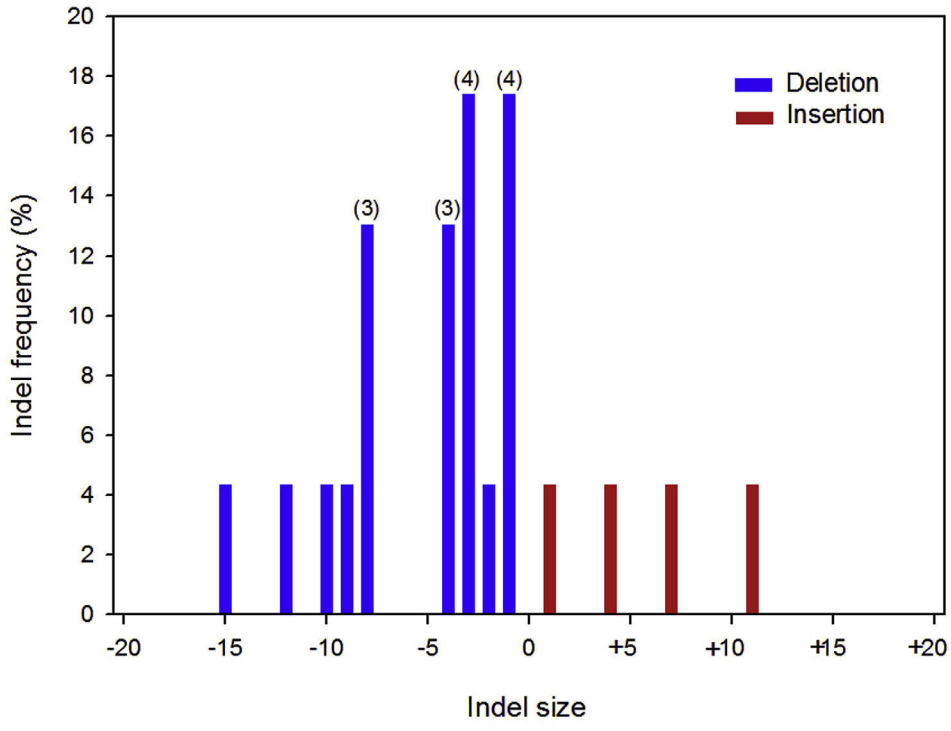


Fig. 4. Distribution of all *de novo* indels identified according to their size.
The graph shows the frequency of the indels according to their size in nucleotide. We counted all events observed in F0 animals as well as the new, non-parental events observed in F1. Identical indels identified in several F0 are independent mutational events and were all taken into consideration whereas identical non-parental indels transmitted by the same F0 parent were considered as a single event (due to parental germline mosaicism) and were counted only once.
As the HDR mechanism does not result in the production of indels, insertions of the HA tag which occurred during *FlagF-HA* production were not taken into consideration here. Are presented here all the observed NHEJ events (including those which took place following injections for *FlagF-HA* production). All indels are described in Supp. Table 2.

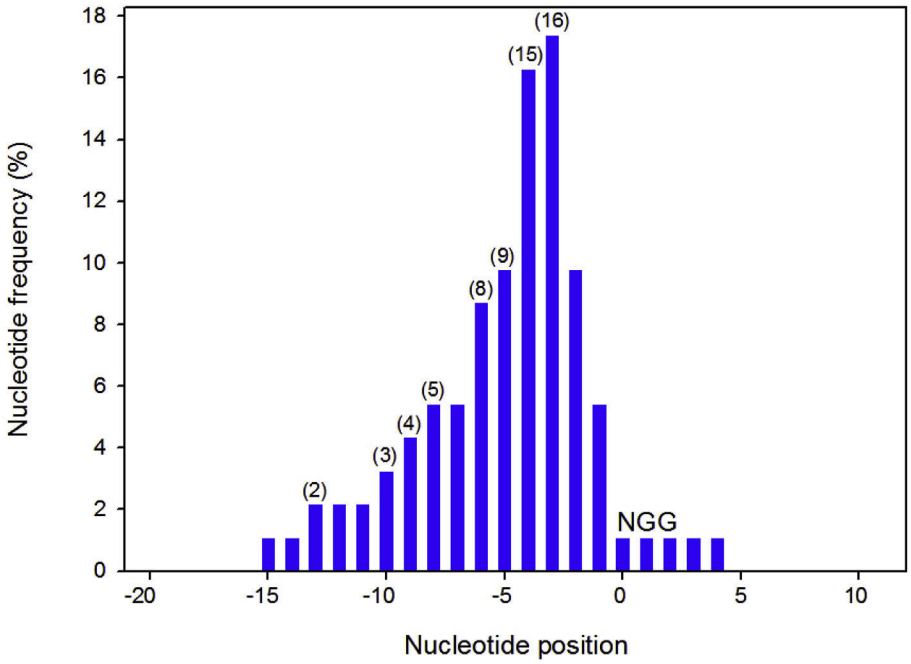


Fig. 5. Localization of NHEJ events on the sequence surrounding the PAM.
As in Fig. 4, all F0 and F1 *de novo* indels are taken into consideration. The plot indicates the number of times (frequency) a nucleotide was concerned by an indel. For the deletions, all deleted nucleotides were taken into account. For the insertions, both nucleotides located before and after the insertion were considered as affected by the NHEJ and counted. When an event occurred within a single nucleotide repeat, the most 5′ nucleotide was taken into consideration. All indels are described in Supp. Table 2. NGG corresponds to the PAM sequence.

bp insertion in exon 3 (InsTCAGATA). RT-PCR was performed on testis ARN from *Cfap43*$^{-/-}$ and *Cfap44*$^{-/-}$ mice which confirmed the production of abnormal transcripts in both mutants leading to a premature stop codon.

Reproductive phenotype was studied for both KO mice models. Homozygous KO males exhibited complete infertility when mated with WT females. Sperm concentrations fell within the normal values for mouse but all spermatozoa were immotile. For *Cfap43*$^{-/-}$, 100% of sperm had a short flagella whereas *Cfap44*$^{-/-}$ had a flagella of normal length with some moderate morphological abnormalities but with a complete absence of motility (Fig. 3). In both cases the axonemal structure was altered, often with an absence of the central microtubule pair. We note that the short flagella observed in *Cfap43*$^{-/-}$ animals is one of the hallmark of the MMAF phenotype in human, the KO model thus producing a faithful reproduction of the human anomalies. Sperm from *Cfap44*$^{-/-}$ animals do not present any obvious morphological anomalies of the flagella but are immotile and present the same ultrastructural defects observed in human. This confirms the human genetic results but highlights some slight differences between human and mouse flagellar growth (Coutton et al., 2018).

4. Discussion

Whole exome sequencing is a powerful technology that enables discovering new candidate genes implicated in different human disorders such as infertility. Validation of these genes through the study of transgenic animal models is important and enables to greatly improve our understanding of spermatogenesis and male reproduction. We observed here that CRISPR/Cas9 delivers its promises as mutated F0 animals were usually obtained in one month after the first injection and stable lineages producing homozygous animal could be obtained in approximately 6 months. Interestingly, we observed that the three KO models we have created (i.e. the *Cfap43*, *Cfap44 and FlagC* genes) reproduced the human phenotype. This confirms that mouse is a good model to study human male infertility and spermatogenesis. Furthermore, we have worked on several other models and have so far always observed a good correlation between the mouse and the human phenotype: we observed for instance that *Dpy19l2*$^{-/-}$ mice perfectly mimicked the globozoospermia phenotype (Pierre et al., 2012) and *Spink2*$^{-/-}$ mice had azoospermia, like our mutated patients (Kherraf et al., 2017).

4.1. Frequent occurrence of F0 mosaic

It has been reported that gRNA and Cas9 can be passed on through several successive mitoses initiating double strand break and NHEJ in different daughter cells thus producing mosaic animals carrying different mutations in different tissues (Inui et al., 2014; Midic et al., 2017; Yen et al., 2014). Here F0 animal were genotyped using tail/digit DNA, containing mostly bone, cartilage, muscle and skin tissues. We observed that all bred F0 animals did not only pass the initially identified mutation but that 23% of their offspring carried one of two different indels not present in the F0's tail. Most of these mutations were found in several siblings thus indicating that the analyzed F0 presented a germinal mosaic, often with three different mutations. As such variants were not detected in the tail/digit biopsies it seems that germ cell progenitors have been more subjected to CRISPR induced double strand breaks than the progenitors of the tested somatic cells. However, during the course of Sanger sequencing we observed a few abnormal electropherograms of very low intensity overlapping the normal sequence that could sign of the presence of low level mosaicism. When breeding these F0 animals we never obtained any mutations in the F1s, thus indicating that these abnormal sequences could have been the signs of low level somatic mosaicism.

The high rate of mosaicism observed here could be due in part to the fact that we chose to inject plasmids permitting the expression of the sgRNA and the Cas9 protein. It is likely that the plasmids will not be as quickly degraded than native RNA and protein thus permitting to induce some mutational even after several cell divisions. This can be an advantage as we could obtain some new mutations and an increased number of mutated animals. It however complicates the analysis and to limit the extent of mosaicism we can suggest to reduce the quantity of plasmid injected or to directly inject the protein and RNA. Interestingly, it was observed in human embryos that the rate of mosaicism was reduced when injecting the MI oocyte directly during the fertilization by ICSI rather than to inject the zygote (Ma et al., 2017) as is usually done and as was done here. Earlier injection of the CRISPR/Cas9 elements could therefore be an interesting alternative to reduce the occurrence of mosaicism.

4.2. Localization of the mutational events and off target mutations

The PAM, a short sequence motif adjacent to the gRNA-targeted sequence on the invading DNA, plays an essential role in Cas9 binding and DNA break (Shah et al., 2013). In our experience a suitable PAM sequence can always be found, especially if the objective is to create KO animals allowing great flexibility for choosing suitable target regions. For the creation of KI, gRNA has to be located near the target region leaving fewer choices and thus restricting the number of available gRNA sequences. We only have one experience with KI production but we had excellent results and could quickly obtain animals with the desired insertion, these results confirms that CRISPR/Cas9 is an efficient tool to generate KI animals (Seruggia and Montoliu, 2014). Double strand breaks normally occur 3 nucleotide prior the PAM sequence. As the breaks are repaired we cannot have a direct localization of the break. What can be observed is the localization of the modified nucleotide during the NHEJ repair process. We observed in Fig. 5 that the nucleotides that are most often modified are in fact the nucleotides located in positions 3 and 4 before the PAM. These results thus confirm that the DSB occurs most often between these two nucleotides and confirm that the obtained indels are located within the target sequence.

Many studies describe a relative high level of off-target mutagenesis induced by the CRISPR-Cas9 system when injected into cultured human cells (Fu et al., 2013; Kuscu et al., 2014) whereas many others showed a low incidence of this undesirable effect when this system is injected into mouse zygotes (Hryhorowicz et al., 2017; Iyer et al., 2015). Concerning our experience, we only sequenced the targeted sequences and have no way of assessing the frequency of such events. We therefore expect that Cas9 usage will slightly increase the overall rate of *de novo* indels. Simple statistic tell us that 90% of these events will not occur in coding sequence and an even greater percentage will have no deleterious effect. To be absolutely sure that the observed phenotype is due to the selected mutation, we used the following two strategies: 1) breeding with WT C57BL6/J animal and use F3-F5 animals. As off target events are unlikely to be allelic with the targeted sequence, the selection and backcrossing of the selected KO/KI will quickly eliminate the unwanted mutations. 2) Analyze several animals with different mutations. As we have seen here the CRISPR/Cas9 technique allows to obtain several animals with different indels. Furthermore, some F0 can have a homozygous mutation. An acceptable solution is to phenotype one or two F0 homozygous animal and to establish a stable line with a different mutation. The observation of the same phenotype in several animals with different indels rules out the possibility of a non-specific effect from an off target event.

4.3. Technical difficulties

The microinjection of Cas9-sgRNA into embryos requires a relatively high skill level. Recently, electroporation was used as an alternative to microinjection and demonstrated a high efficiency for introduction of Cas9-sgRNA into zygotes (Kaneko, 2017; Wefers et al., 2017). The injection can however be easily performed by any staff trained for mouse ICSI. As any highly technical gesture, there is a learning curve. We observed that our results improved with time (we had more births and mutations in the later injections). We believe that this might be partly related with skills improvement of the technical staff.

In our experience the use of “all in one” vectors (with the sgRNA and Cas9) and its direct injection into the male pronucleus saved much time and many operator dependent operations (such as RNA isolation and purification).

We note that some sgRNA did not permit to obtain any mutated animals (or very few) but the co-injection of two sgRNA always permitted to obtain several mutated pups. The use of two sgRNA located within the same gene has been described to mediate large deletions loosely delimited by the two sequences targeted by the sgRNAs (Chen et al., 2014; Han et al., 2014). In our experience such events are rare as we did evidenced them. Here, our main goal was to produce KO animals and the occurrence of such large deletion would be interesting as they would necessarily be deleterious and would have the added advantage of being easy to detect by a simple PCR. As we did not observe any disadvantage associated with the co-injection of two sgRNA we recommend this strategy.

4.4. Timing of CRISPR/Cas9 gene editing projects

Overall a CRISPR/Cas9 project can be divided into three steps: 1. target sequence selection, design of mutant alleles and preparation/ordering of plasmids/reagents, 2. microinjection or electroporation of zygotes and embryo transfer, and 3. genotyping of founder mice, back crossing and phenotyping. The first step only take a few hours of intellectual work and 2–3 weeks of production, usually by a service provider. In step 2 female mice are treated with hormones for superovulation starting three days before zygote microinjections. Upon the mating of superovulated females to males, zygotes are collected the next day, microinjected with CRISPR/Cas9 reagents and transferred into foster mothers, which give rise to birth after 3 weeks. Step 3 is the longest and usually takes approximately 6 months (Fig. 2). Biopsies can be taken from the pups derived from microinjected zygotes at the age of 10 days and used for genotyping. Initial genotyping will have to be done by double stranded sequencing of the target region, usually taking approximately 2 weeks. In many cases CRISPR/Cas9 microinjection permit the birth of mutant F0 founders that can be mated to WT animals when they reach fertility around 9 weeks. Considering that mice will reach reproductive age at 9 weeks, the expected timeframe to obtain the first mutated mature F1 animals is therefore of 24 weeks (2 gestation periods of 3 weeks and 2 periods of sexual maturation of 9 weeks) and mature F1 progeny will be available after 36 weeks (See Fig. 2). Mutated F0 pups recovered from manipulated zygotes are however not always successfully produced and a new series of injection must then be performed. It will usually take at least a month to reschedule the injection procedure.

5. Conclusion

Our genetic study on infertile subjects with flagellar defects (MMAF phenotype) permitted to identify some very plausible mutations located in 11 candidate genes in 63% of the analyzed subjects. This clearly confirms the interest of WES for both

pathological gene identification and diagnostic purpose in the context of male infertility. We also confirmed that the CRISPR/cas9 technique is very efficient to create KO and KI animals and is a promising strategy to study the function of candidate genes for infertility while permitting to convincingly confirm their pathogeny.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgments

This work was mainly supported by the following grants: The "MAS-Flagella" project financed by French ANR and the DGOS for the program PRTS 2014 to PR and CA and the Swiss National Science Foundation (Sinergia grant N° CRSII5_171007) to SN.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.mce.2018.03.002.

References

Alkuraya, F.S., 2016. Discovery of mutations for Mendelian disorders. Hum. Genet. 135, 615–623. https://doi.org/10.1007/s00439-016-1664-8.

Amiri-Yekta, A., Coutton, C., Kherraf, Z.-E., Karaouzène, T., Le Tanno, P., Sanati, M.H., Sabbaghian, M., Almadani, N., Gilani, S., Ali, M., Hosseini, S.H., Bahrami, S., Daneshipour, A., Bini, M., Arnoult, C., Colombo, R., Gourabi, H., Ray, P.F., 2016. Whole-exome sequencing of familial cases of multiple morphological abnormalities of the sperm flagella (MMAF) reveals new DNAH1 mutations. Hum. Reprod. 31, 2872–2880. https://doi.org/10.1093/humrep/dew262.

Barman, H.K., Rasal, K.D., Chakrapani, V., Ninawe, A.S., Vengayil, D.T., Asrafuzzaman, S., Sundaray, J.K., Jayasankar, P., October 2017. Gene editing tools: state-of-the-art and the road ahead for the model and non-model fishes. Transgenic Res. 26 (5), 577–589. https://doi.org/10.1007/s11248-017-0030-5.

Ben Khelifa, M., Coutton, C., Zouari, R., Karaouzène, T., E, D., Pernet-Gallay, K., Merdassi, G., Satre, V., Jouk, P.S., Hennebicq, S., Grunwald, D., Escalier, D., Pernet-Gallay, K., Jouk, P.S., Thierry-Mieg, N., Touré, A., Arnoult, C., Ray, P.F., 2014. Mutations in DNAH1, which encodes an inner arm heavy chain dynein, lead to male infertility due to multiple morphological abnormalities of the sperm flagella. Am. J. Hum. Genet. 94 (1), 94–104.

Capecchi, M.R., 1989. Altering the genome by homologous recombination. Science 244, 1288–1292.

Capecchi, M.R., 2005. Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. Nat. Rev. Genet. 6, 507–512. https://doi.org/10.1038/nrg1619.

Chen, X., Xu, F., Zhu, C., Ji, J., Zhou, X., Feng, X., Guang, S., 2014. Dual sgRNA-directed gene knockout using CRISPR/Cas9 technology in Caenorhabditis elegans. Sci. Rep. 4. https://doi.org/10.1038/srep07581.

Coutton, C., Zouari, R., Abada, F., Khelifa, M.B., Merdassi, G., Triki, C., Escalier, D., Hesters, L., Mitchell, V., Levy, R., Sermondade, N., Boitrelle, F., Vialard, F., Satre, V., Hennebicq, S., Jouk, P.-S., Arnoult, C., Lunardi, J., Ray, P.F., 2012. MLPA and sequence analysis of DPY19L2 reveals point mutations causing globozoospermia. Hum. Reprod. 27, 2549–2558. https://doi.org/10.1093/humrep/des160.

Coutton, C., Escoffier, J., Martinez, G., Arnoult, C., Ray, P.F., 2015. Teratozoospermia: spotlight on the main genetic actors in the human. Hum. Reprod. Update 21, 455–485. https://doi.org/10.1093/humupd/dmv020.

Coutton, C., Vargas, A.S., Amiri-Yekta, A., Kherraf, Z.-E., Ben Mustapha, S.F., Le Tanno, P., Wambergue-Legrand, C., Karaouzène, T., Martinez, G., Crouzy, S., Daneshipour, A., Hosseini, S.H., Mitchell, V., Halouani, L., Marrakchi, O., Makni, M., Latrous, H., Kharouf, M., Deleuze, J.-F., Boland, A., Hennebicq, S., Satre, V., Jouk, P.-S., Thierry-Mieg, N., Conne, B., Dacheux, D., Landrein, N., Schmitt, A., Stouvenel, L., Lorès, P., El Khouri, E., Bottari, S.P., Fauré, J., Wolf, J.-P., Pernet-Gallay, K., Escoffier, J., Gourabi, H., Robinson, D.R., Nef, S., Dulioust, E., Zouari, R., Bonhivers, M., Touré, A., Arnoult, C., Ray, P.F., 2018. Mutations in CFAP43 and CFAP44 cause male infertility and flagellum defects in Trypanosoma and human. Nat. Commun. 9. https://doi.org/10.1038/s41467-017-02792-7.

Dieterich, K., Soto Rifo, R., Karen Faure, A., Hennebicq, S., Amar, B.B., Zahi, M., Perrin, J., Martinez, D., Sèle, B., Jouk, P.-S., Ohlmann, T., Rousseaux, S., Lunardi, J., Ray, P.F., 2007. Homozygous mutation of AURKC yields large-headed polyploid spermatozoa and causes male infertility. Nat. Genet. 39, 661–665. https://doi.org/10.1038/ng2027.

Dieterich, K., Zouari, R., Harbuz, R., Vialard, F., Martinez, D., Bellayou, H., Prisant, N., Zoghmar, A., Guichaoua, M.R., Koscinski, I., Kharouf, M., Noruzinia, M., Nadifi, S., Sefiani, A., Lornage, J., Zahi, M., Viville, S., Sele, B., Jouk, P.-S., Jacob, M.-C., Escalier, D., Nikas, Y., Hennebicq, S., Lunardi, J., Ray, P.F., 2009. The Aurora Kinase C c.144delC mutation causes meiosis I arrest in men and is frequent in the North African population. Hum. Mol. Genet. 18, 1301–1309. https://doi.org/10.1093/hmg/ddp029.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., Sander, J.D., 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat. Biotechnol. 31, 822–826. https://doi.org/10.1038/nbt.2623.

Han, J., Zhang, J., Chen, L., Shen, B., Zhou, J., Hu, B., Du, Y., Tate, P.H., Huang, X., Zhang, W., 2014. Efficient in vivo deletion of a large imprinted lncRNA by CRISPR/Cas9. RNA Biol. 11, 829–835. https://doi.org/10.4161/rna.29624.

Harbuz, R., Zouari, R., Pierre, V., Ben Khelifa, M., Kharouf, M., Coutton, C., Merdassi, G., Abada, F., Escoffier, J., Nikas, Y., Vialard, F., Koscinski, I., Triki, C., Sermondade, N., Schweitzer, T., Zhioua, A., Zhioua, F., Latrous, H., Halouani, L., Ouafi, M., Makni, M., Jouk, P.-S., Sèle, B., Hennebicq, S., Satre, V., Viville, S., Arnoult, C., Lunardi, J., Ray, P.F., 2011. A recurrent deletion of DPY19L2 causes infertility in man by blocking sperm head elongation and acrosome formation. Am. J. Hum. Genet. 88, 351–361. https://doi.org/10.1016/j.ajhg.2011.02.007.

Hryhorowicz, M., Lipiński, D., Zeyland, J., Słomski, R., 2017. CRISPR/Cas9 immune system as a tool for genome engineering. Arch. Immunol. Ther. Exp. 65, 233–240. https://doi.org/10.1007/s00005-016-0427-5.

Hsu, P.D., Lander, E.S., Zhang, F., 2014. Development and applications of CRISPR-Cas9 for genome engineering. Cell 157, 1262–1278. https://doi.org/10.1016/j.cell.2014.05.010.

Inui, M., Miyado, M., Igarashi, M., Tamano, M., Kubo, A., Yamashita, S., Asahara, H., Fukami, M., Takada, S., 2014. Rapid generation of mouse models with defined point mutations by the CRISPR/Cas9 system. Sci. Rep. 4, 5396. https://doi.org/10.1038/srep05396.

Iyer, V., Shen, B., Zhang, W., Hodgkins, A., Keane, T., Huang, X., Skarnes, W.C., 2015. Off-target mutations are rare in Cas9-modified mice. Br. J. Pharmacol. 12, 479, 479. https://doi.org/10.1038/nmeth.3408.

Kaneko, T., 2017. Genome editing in mouse and rat by electroporation. Methods Mol. Biol. Clifton NJ. 1630, 81–89. https://doi.org/10.1007/978-1-4939-7128-2_7.

Kherraf, Z.-E., Christou-Kent, M., Karaouzene, T., Amiri-Yekta, A., Martinez, G., Vargas, A.S., Lambert, E., Borel, C., Dorphin, B., Aknin-Seifer, I., Mitchell, M.J., Metzler-Guillemain, C., Escoffier, J., Nef, S., Grepillat, M., Thierry-Mieg, N., Satre, V., Bailly, M., Boitrelle, F., Pernet-Gallay, K., Hennebicq, S., Fauré, J., Bottari, S.P., Coutton, C., Ray, P.F., Arnoult, C., 2017. SPINK2 deficiency causes infertility by inducing sperm defects in heterozygotes and azoospermia in homozygotes. EMBO Mol. Med. 9, 1132–1149. https://doi.org/10.15252/emmm.201607461.

Kuscu, C., Arslan, S., Singh, R., Thorpe, J., Adli, M., 2014. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. Nat. Biotechnol. 32, 677–683. https://doi.org/10.1038/nbt.2916.

Ma, H., Marti-Gutierrez, N., Park, S.-W., Wu, J., Lee, Y., Suzuki, K., Koski, A., Ji, D., Hayama, T., Ahmed, R., Darby, H., Dyken, C.V., Li, Y., Kang, E., Park, A.-R., Kim, D., Kim, S.-T., Gong, J., Gu, Y., Xu, X., Battaglia, D., Krieg, S.A., Lee, D.M., Wu, D.H., Wolf, D.P., Heitner, S.B., Belmonte, J.C.I., Amato, P., Kim, J.-S., Kaul, S., Mitalipov, S., 2017. Correction of a pathogenic gene mutation in human embryos. Nature 548, 413. https://doi.org/10.1038/nature23305.

Midic, U., Hung, P.-H., Vincent, K.A., Goheen, B., Schupp, P.G., Chen, D.D., Bauer, D.E., VandeVoort, C.A., Latham, K.E., 2017. Quantitative assessment of timing, efficiency, specificity and genetic mosaicism of CRISPR/Cas9-mediated gene editing of hemoglobin beta gene in rhesus monkey embryos. Hum. Mol. Genet. 26, 2678–2689. https://doi.org/10.1093/hmg/ddx154.

Pierre, V., Martinez, G., Coutton, C., Delaroche, J., Yassine, S., Novella, C., Pernet-Gallay, K., Hennebicq, S., Ray, P.F., Arnoult, C., 2012. Absence of Dpy19l2, a new inner nuclear membrane protein, causes globozoospermia in mice by preventing the anchoring of the acrosome to the nucleus. Dev. Camb. Engl. 139, 2955–2965. https://doi.org/10.1242/dev.077982.

Ray, P.F., Toure, A., Metzler-Guillemain, C., Mitchell, M.J., Arnoult, C., Coutton, C., 2017. Genetic abnormalities leading to qualitative defects of sperm morphology or function. Clin. Genet. 91, 217–232. https://doi.org/10.1111/cge.12905.

Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., Siksnys, V., 2011. The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. Nucleic Acids Res. 39, 9275–9282. https://doi.org/10.1093/nar/gkr606.

Seruggia, D., Montoliu, L., 2014. The new CRISPR-Cas system: RNA-guided genome engineering to efficiently produce any desired genetic alteration in animals. Transgenic Res. 23, 707–716. https://doi.org/10.1007/s11248-014-9823-y.

Sha, Yan-Wei, Wang, Xiong, Xu, Xiaohui, Su, Zhi-Ying, Cui, Yuanqing, Mei, Li-Bin, Huang, Xian-Jing, Chen, Jie, He, Xue-Mei, Ji, Zhi-Yong, Bao, Hongchu, Yang, Xiaoyu, Li, Ping, Li, Lin, 2017. Novel mutations in CFAP44 and CFAP43 cause multiple morphological abnormalities of the sperm flagella (MMAF). Reprod. Sci. 1933719117749756 https://doi.org/10.1177/1933719117749756.

Shah, S.A., Erdmann, S., Mojica, F.J.M., Garrett, R.A., 2013. Protospacer recognition motifs: mixed identities and functional diversity. RNA Biol. 10, 891–899. https://doi.org/10.4161/rna.23764.

Soumillon, M., Necsulea, A., Weier, M., Brawand, D., Zhang, X., Gu, H., Barthès, P., Kokkinaki, M., Nef, S., Gnirke, A., Dym, M., de Massy, B., Mikkelsen, T.S., Kaessmann, H., 2013. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. Cell Rep. 3, 2179–2190. https://doi.org/10.1016/j.celrep.2013.05.031.

Tang, S., Wang, X., Li, W., Yang, X., Li, Z., Liu, W., Li, C., Zhu, Z., Wang, L., Wang, J., Zhang, L., Sun, X., Zhi, E., Wang, H., Li, H., Jin, L., Luo, Y., Wang, J., Yang, S., Zhang, F., 2017. Biallelic mutations in CFAP43 and CFAP44 cause male infertility with multiple morphological abnormalities of the sperm flagella. Am. J. Hum. Genet. 100, 854–864. https://doi.org/10.1016/j.ajhg.2017.04.012.

Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., Jaenisch, R., 2013. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell 153, 910–918. https://doi.org/10.1016/j.cell.2013.04.025.

Wefers, B., Bashir, S., Rossius, J., Wurst, W., Kühn, R., 2017. Gene editing in mouse zygotes using the CRISPR/Cas9 system. Methods San Diego Calif 121–122, 55–67. https://doi.org/10.1016/j.ymeth.2017.02.008.

Yen, S.-T., Zhang, M., Deng, J.M., Usman, S.J., Smith, C.N., Parker-Thornburg, J., Swinton, P.G., Martin, J.F., Behringer, R.R., 2014. Somatic mosaicism and allele complexity induced by CRISPR/Cas9 RNA injections in mouse zygotes. Dev. Biol. 393, 3–9. https://doi.org/10.1016/j.ydbio.2014.06.017.

Supplementary table 1: List of homozygous variants identified in FLAGC and FLAGF.

Gene	Nucleotide change	Amino-acid change	gnomAD frequency
FLAGC	c.2353_2354delTT	p.Leu785MetfsTer5 (p.L785X)	0
	c.1023+1G>A	Splicing variant	0
	c.2279T>A	p.Ile760Asn (p.I760N)	0
FLAGF	c.1084C>T	p.Arg362Ter (p.R362*)	$4.69x10^{-5}$
	c.2112delG	p.Arg704SerfsTer7 (R704X)	0
	c.2324A>T	p.Asp775Val (p.D775V)	0

Supplementary table 2: Generation of F0 mice after CRISPR/cas9 injection and number of homozygous (hmz) or heterozygous (htz) mutated (mut) pups in target region 1 (TR1) and target region 2 (TR2).

	Cfap43		*Cfap44*			*FlagC*	*FlagF*		*FlagF-HA*		Total	Av./Inj.
	Inj. 1	Inj. 2	Inj. 1	Inj. 2	Inj. 3	Inj. 1	Inj. 1	Inj. 2	Inj. 1	Inj. 2		
Injected zygotes						97	130	166	135	167	695	139
Transferred embryos						66	66	118	94	135	479	95.8
Newborns	12	9	7	10	9	8	3	14	10	20	102	10.2
Pups with 1 htz TR1 mut.	0	2	0	0	1	1	1	1	4	4	15	1.5
Pups with ≥2 TR1 mut.	0	0	0	0	0	1	0	1	1	8	10	1
Pups with 1 TR1 hmz mut.	0	0	0	0	0	0	1	0	0	0	1	0.1
Pups with 1 htz TR2 mut.	0	2	0	0	0	1	0	1	/	/	4	0.4
Pups with ≥ 2 TR2 mut.	0	0	0	0	0	0	0	0	/	/	0	0
Pups with 1 TR2 hmz mut	0	1	0	0	0	0	0	0	/	/	1	0.1
Total No of mut. pups	0	5	0	0	1	3	2	3	5	12	31	
% mut. Pups per live pup	0.0	55.6	0.0	0.0	11.1	37.5	66.7	21.4	50	60	30.4	
Total No of *de novo* indels	0	6	0	0	1	4	3	3	6	20	43	

Supplementary table 3: Generation of F1 mice from CRISPR/cas9 created F0 and number of pups presenting the parental mutation F0 or a new mutation Mut. 1 or Mut. 2.

Litter 1 and 2 for a same gene are generated with the same parents. The new mutations mut. 1 and mut.2 are identical from one litter to another. A maximum of two different "new" mutations (plus the parental mutation) were identified the 3 analyzed couples.

	Cfap43			*Cfap44*				*FlagC*			Total	% mut.
	Litter 1	Litter 2		Litter 1	Litter 2	Litter 3		Litter 1	Litter 2			
Newborn	3	12		6	11	10		9	10		61	
Pups with mut. F0	2	3		1	3	4		4	3		20	
% pups with F0 mut.	66,7	25,0		16,7	27,3	40,0		44,4	30,0			32,8
Pups with mut. 1	0	3		1	2	0		4	1		11	18,0
Pups with mut. 2	0	0		1	1	1		0	0		3	4,9
Total with non parental mut	0	3		2	3	1		4	1		14	
% pups with non parental mut	0,0	25,0		33,3	27,3	10,0		44,4	10,0			23,0
Total Nb mut. pups	2	6		3	6	5		8	4		34	
% overall mut. pups	66,7	50,0		50,0	54,5	50,0		88,9	40,0			55,7

Supplementary table 4: List of all mutations identified in F0 and F1 animals and position in regard to the PAM 1st nucleotide.

*indicates the mutations selected for generation of F1.†indicates the F0 homozygous mutations. Position refers to the position in regards to the PAM 1st nucleotide (**C**GG).ie: in "CAGCTGACTCA*TATA*CCGA**CGG"**delTATA (italic) is in position -5 in regards to the PAM (in bold). F0, F1 indicates if the mutation was first identified in F0 or F1 animals. "Target" indicates if the indels occurred in the first or second targeted region.

Gene Targeted	Position	Size	F0	F1	Mutation
Cfap43 Target1	NC	NC	2		Uncharacterized events
Cfap43 Target2	-4	-1	2		delG†
	-4	+4		1	insTCCA
	-6	-4	1		delAAGG*
	-10	-10	1		delTGTGAAGGGC
Cfap44 Target1	-4	+7	1		InsTCAGATA*
	-6	-3		1	delTAC
	-8	-4		1	delTATA*
Cfap44 Target2	/				
FlagC Target1	NC	NC	2		Uncharacterized events
	-4	+1		1	insT*
	-4	-3	1		delTAC
FlagC Target2	-13	-12	1		delAAACCCGTTAGC
FlagF Target1			2		Uncharacterized events
	-3	-1	2		delG†
	-6	-4	1		delTTCG
FlagF Target2	-4	+11	1		insAATGTACATAG
FlagF-HA	NC	NC	≥12		Uncharacterized events
		-	6		HA inserted
	-3	-2	1		delAG
	-3	-3	1		delAGC
	-4	-9	1		delCAGCTGGAA
	-5	-3	1		delGCA
	-9	-8	3		delTTCAGCAG
	-15	-15	1		delAATTAGTTCAGCAGC
Median	-4	-3			
Total			≥40	5	

Chapitre IV. Les techniques d'Assistance Médicale à la Procréation (AMP)

Les techniques d'assistance médicale à la procréation consistent à manipuler les gamètes males et/ou femelles pour procéder à une fécondation. Elles ne permettent pas nécessairement de corriger la ou les cause(s) à l'origine de l'infertilité du couple mais permettent de pallier certaines difficultés à concevoir. Dans les pays occidentaux, la majorité des couples infertiles ont recours aux techniques d'assistance médicale à la procréation (AMP) pour obtenir une grossesse. On compte actuellement environ cinq millions de personnes devant leur existence à ces techniques. En Europe et aux USA, les techniques d'AMP sont responsables de 1-3% des naissances. En France en 2006, on a recensé 119 000 cas de recours aux techniques d'AMP comprenant les inséminations, les FIV (fécondation in vitro), les ICSI (micro-injection intra-cytoplasmique de spermatozoïdes) et les transferts d'embryons congelés **(Figure 40)**.

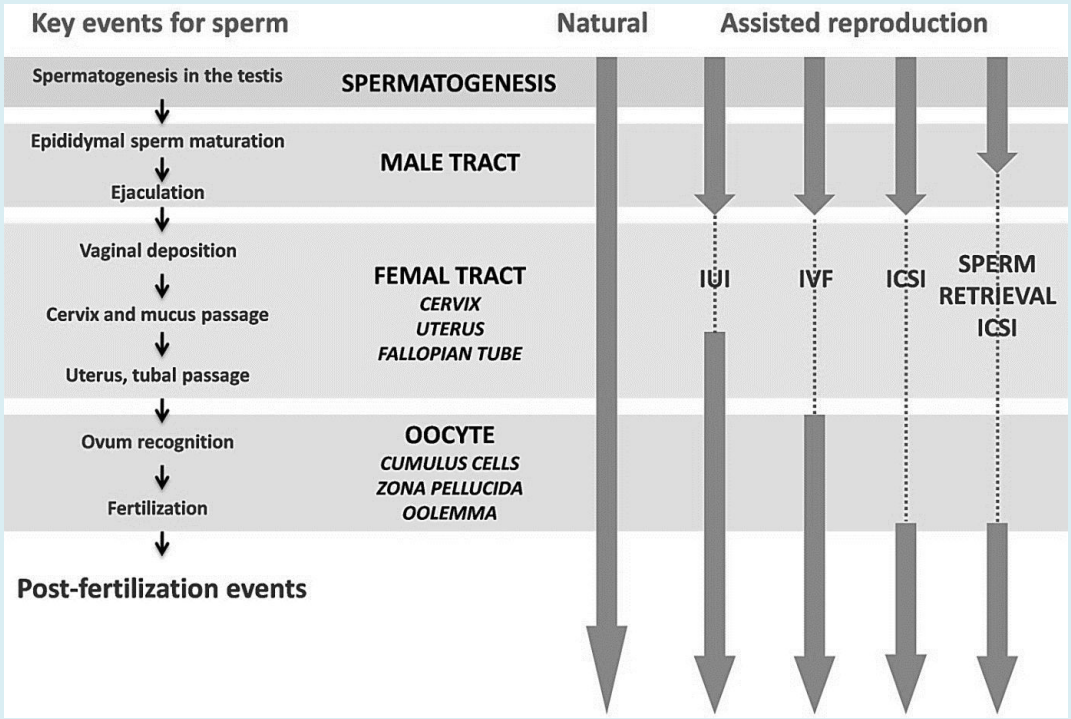


Figure 40. Procréation naturelle versus procréation médicalement assistée.
IUI : insémination artificielle intra-utérine ; IVF : fécondation in vitro ; Sperm retrieval ICSI : Extraction de spermatozoïdes testiculaires ou épididymaires en vue d'une ICSI (micro-injection intracytoplasmique) (adaptée de Sakkas et al., 2015).

1.1- L'Insémination Artificielle (IA)

L'insémination artificielle consiste à introduire des spermatozoïdes préalablement sélectionnés dans le tractus génital féminin en utilisant un cathéter qui franchit le col utérin. L'insémination est réalisée de façon synchrone avec l'ovulation soit le lendemain du pic spontané de la LH soit 34-36 h après le déclenchement artificiel de l'ovulation. Cette technique ne peut être envisagée que si les trompes utérines sont perméables et qu'un nombre suffisant de spermatozoïdes mobiles peut être obtenu.

1.2- La Fécondation In Vitro (FIV)

La fécondation in vitro (FIV) consiste à reproduire la fécondation et une partie du développement embryonnaire précoce au laboratoire en dehors du tractus génital féminin. Les principales indications de cette technique sont la stérilité tubaire et les infertilités idiopathiques. Elle se déroule en quatre étapes successives : 1) la stimulation ovarienne, 2) le recueil des gamètes, 3) la FIV proprement dite et 4) le transfert embryonnaire dans l'utérus.

1.2.1- La stimulation ovarienne

La stimulation ovarienne a pour objectif de permettre la maturation simultanée de plusieurs follicules afin d'augmenter les chances de fécondation et de grossesse. Les stimulateurs de l'ovulation ont beaucoup évolué dans la pratique clinique, du citrate du clomifène (inducteur de sécrétion de la FSH) aux gonadotrophines (urinaires humaines ou recombinantes). Plusieurs schémas thérapeutiques existent. En vue d'une FIV, le protocole associe généralement une gonadotrophine à un agoniste ou antagoniste de la GnRH qui en mettant l'hypophyse au repos permettent de prévenir l'ovulation prématurée. Le monitoring de la stimulation ovarienne est indispensable pour la surveillance du bon déroulement de la croissance des follicules ovariens. Il est basé sur le contrôle échographique des ovaires et le dosage plasmatique du 17-β-estradiol (E2).

1.2.2- Le recueil des gamètes

Habituellement, lorsque la taille des follicules recrutés atteint 17-18 mm avec un taux plasmatique de 17-β-estradiol de l'ordre de 1000 pg/mL l'ovulation est déclenchée. L'injection de l'hCG (human chorionic gonadotropin) reproduit le pic préovulatoire et induit la maturation finale des ovocytes. Le recueil des ovocytes matures est alors réalisé par ponction ovarienne du liquide folliculaire par voie transvaginale échoguidée juste avant

l'ovulation soit 34-36 h après l'injection de l'hCG. Le sperme est recueilli par masturbation au laboratoire, le jour de la ponction des follicules, il est ensuite préparé techniquement afin de sélectionner les meilleurs spermatozoïdes en ce qui concerne la morphologie et la mobilité.

1.2.3- La FIV proprement dite

La FIV consiste à mettre en culture les spermatozoïdes sélectionnés avec les ovocytes recueillis. La fécondation survient entre 16h et 20h après la mise en culture, caractérisée par l'apparition de deux pronucléi, mâle et femelle, suivie par la première division cellulaire mitotique de l'ovocyte fécondé (zygote) aboutissant à la formation de l'embryon.

1.2.4- Le transfert embryonnaire

Il s'agit de la dernière étape de la FIV. Les embryons obtenus sont cultivés *in vitro* habituellement pendant 48 h voir plus dans certains cas, avant qu'ils ne soient transférés dans la cavité utérine. Les bonnes pratiques recommandent le transfert d'un seul embryon afin de limiter le risque de grossesse multiple. Les autres embryons seront cryoconservés pour une utilisation ultérieure. La phase lutéale est souvent soutenue par l'administration de progestérone. Le premier dosage de β-hCG plasmatique a lieu deux semaines après la ponction folliculaire.

1.3- La FIV avec micro-injection intracytoplasmique du spermatozoïde

En 1992, Palermo et al., rapportèrent la première grossesse obtenue par l'injection d'un seul spermatozoïde dans le cytoplasme d'un ovocyte mature (Palermo G et al., 1992). Depuis, le traitement de l'infertilité masculine a connu un réel essor. Le spermatozoïde retenu est chargé dans une micropipette puis introduit dans le cytoplasme d'un ovocyte mature (en métaphase II) décoronisé et maintenu par une micropipette de contention (Figure 41).

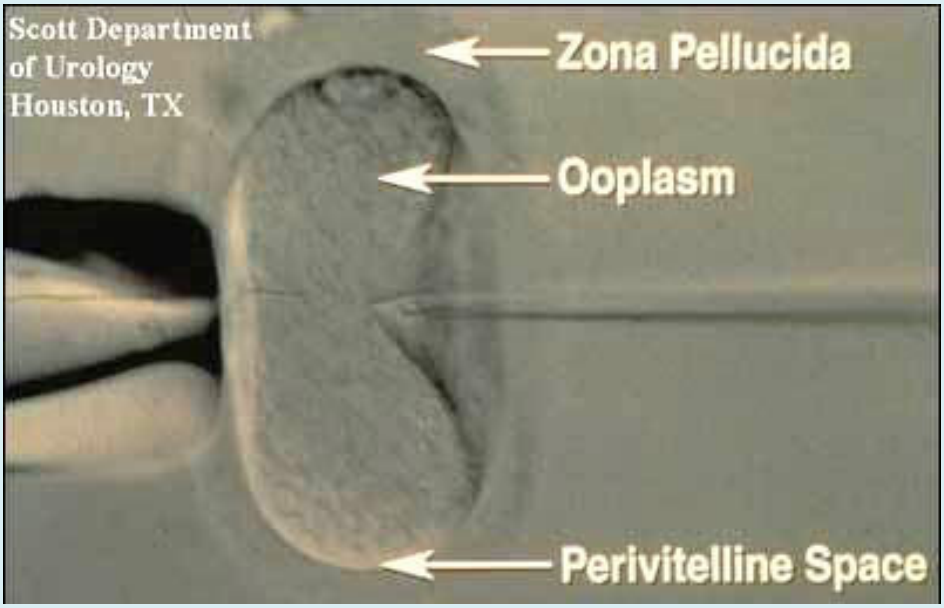


Figure 41. ICSI, intracytoplasmic sperm injection.
(D'après McKnight and McKenzie, 2016).

Cette technique peut être envisagée avec des spermatozoïdes épididymaires obtenus par PESA (percutaneous epididymal sperm aspiration) ou MESA (microsurgical epididymal sperm aspiration) en cas d'azoospermie obstructive (AO) ou avec des spermatozoïdes testiculaires obtenus par TESA (percutaneous testicular sperm aspiration), TESE (conventional testicular sperm extraction) ou Micro-TESE (microsurgical testicular sperm extraction) en cas d'azoospermie non-obstructive (ANO) **(Figure 42, Tableau 3)**.

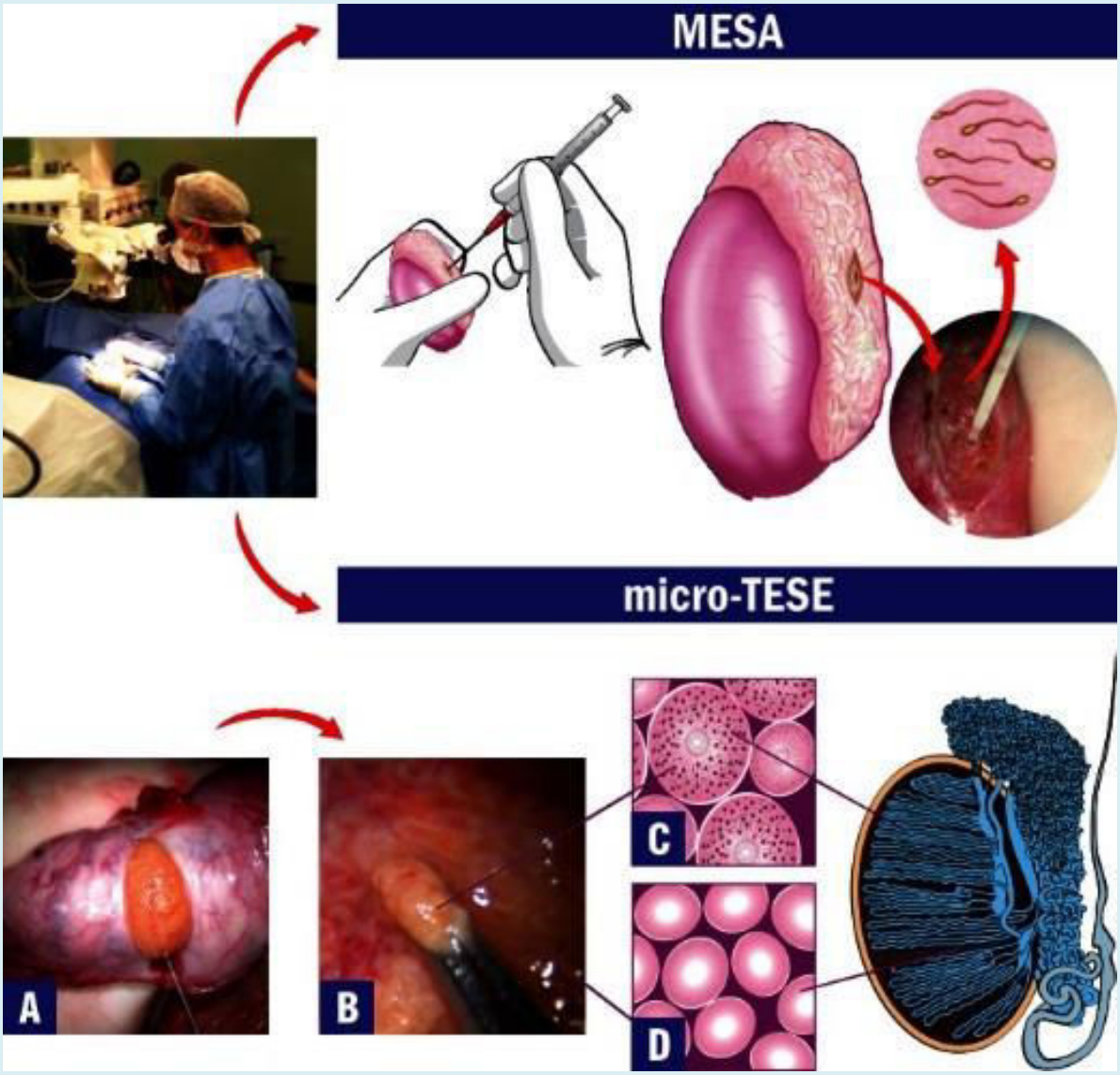


Figure 42. Comparaison des techniques d'extraction des spermatozoïdes épididymaires et testiculaires.
MESA : microsurgical epididymal sperm aspiration. Micro-TESE : microsurgical testicular sperm extraction (d'après Esteves et al., 2011a).

	Advantages	Disadvantages
PESA	Fast and low cost Minimal morbidity, repeatable No microsurgical expertise required Few instruments and materials No surgical exploration	Few sperm retrieved Cryopreservation limited Fibrosis and obstruction at aspiration site Risk of hematoma/spermatocele
MESA	Large number of sperm retrieved Excellent chance of sperm cryopreservation Reduced risk of hematoma	Surgical exploration required Increased cost and time-demanding Microsurgical instruments and expertise required Postoperative discomfort
TESA	Fast and low cost Repeatable No microsurgical expertise required Few instruments and materials No surgical exploration Minimal/mild postoperative discomfort	Relatively low success rate in NOA Few sperm retrieved in NOA Cryopreservation limited Risk of hematoma/testicular atrophy
TESE	No microsurgical expertise required Fast and repeatable	Relatively low success rate in NOA Relatively few sperm retrieved in NOA Risk of testicular atrophy (with multiple biopsies) Postoperative discomfort
Micro-TESE	Higher success rates in NOA Larger number of sperm retrieved Relatively higher chance of sperm cryopreservation [2] Low risk of complications	Surgical exploration required Increased cost and time-demanding Microsurgical instruments and expertise required Postoperative discomfort

Tableau 3. Comparaison des techniques d'extraction des spermatozoïdes épididymaires et testiculaires.
PESA: percutaneous epididymal sperm aspiration; MESA: microsurgical epididymal sperm; TESA: percutaneous testicular sperm aspiration; TESE: conventional testicular sperm extraction; Micro-TESE: microsurgical testicular sperm extraction (d'après Esteves et al., 2011b).

Malgré le succès des techniques d'AMP, la moitié des couples sortent du parcours de soin sans enfant. Une partie de ces échecs est expliquée par l'altération de la gamétogenèse et la production de gamètes incompatibles avec la fécondation et/ou le développement embryonnaire. Bien que la gamétogenèse occupe une place centrale dans la biologie humaine, sa physiologie n'a été que partiellement explorée et sa pathogénie moléculaire reste à ce jour insuffisamment connue.

PARTIE 2 : ETUDE EXPERIMENTALE

Objectifs

L'objectif de cette thèse est l'identification et la caractérisation moléculaire et physiopathologique de nouvelles causes génétiques entrainant des anomalies spermatogéniques sévères et une infertilité masculine. La partie expérimentale de cette thèse est subdivisée en deux parties.

La première partie est centrée sur une forme rare d'azoospermie non obstructive caractérisée par un blocage spermatogénique post-méiotique. Cette partie a comme objectif principal, la caractérisation moléculaire et physiopathologique du déficit du gène *SPINK2* chez l'Homme et la souris transgénique knock-out. Une mutation perte de fonction homozygote du gène *SPINK2* a été identifiée par séquençage exomique chez deux frères atteints nés de parents consanguins. La compréhension de la fonction de ce gène et de la pathogénie moléculaire et cellulaire associée pourrait nous ouvrir des voies thérapeutiques intéressantes chez les patients porteurs de mutations dans ce gène spécifique de la spermiogénèse.

La seconde partie de la thèse a comme objectif principal l'identification de causes génétiques responsables d'une forme rare et sévère de tératozoospermie monomorphe caractérisée par une mosaïque d'anomalies morphologiques du flagelle spermatique (phénotype MMAF) entrainant une asthénozoospermie et une infertilité masculine primaire isolée. Une étude antérieure réalisée par notre équipe a permis d'identifier des mutations homozygotes délétères dans le gène *DNAH1* chez un tiers des sujets testés ce qui suggère que d'autres gènes sont également impliqués. Grâce au séquençage exomique nous espérons identifier ces gènes et optimiser le rendement du diagnostic génétique dans le phénotype MMAF.

Chapitre I. L'azoospermie non-obstructive

1- Article 2 : SPINK2 deficiency causes infertility by inducing sperm defects in heterozygotes and azoospermia in homozygotes

Kherraf ZE, Christou-Kent M, Karaouzène T, Amiri-Yekta A, Martinez G, Vargas AS, Lambert E, Borel C, Dorphin B, Aknin-Seifer I, Mitchell MJ, Metzler-Guillemain C, Escoffier J, Nef S, Grepillat M, Thierry-Mieg N, Satre V, Bailly M, Boitrelle F, Pernet-Gallay K, Hennebicq S, Fauré J, Bottari SP, Coutton C, Ray PF & Arnoult C.

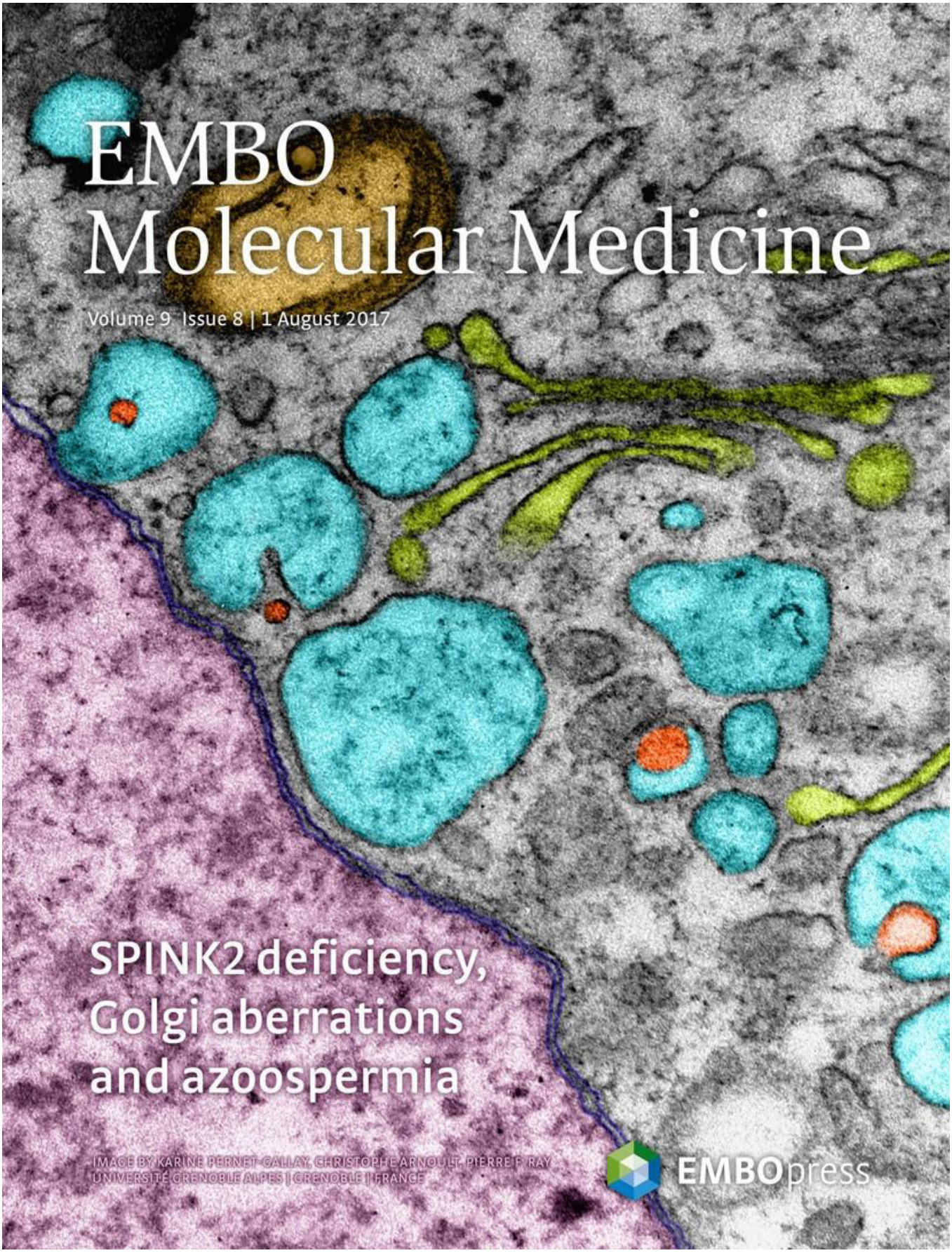
EMBO Molecular Medicine. Mai 2017.

Contexte et principaux résultats de l'étude :

L'azoospermie non obstructive (ANO) traduit un trouble quantitatif sévère de la spermatogenèse et à ce titre est suspectée d'avoir une forte composante génétique. Cependant, jusqu'à présent, très peu de gènes ont été associés à la survenue des ANOs. Les récentes avancées technologiques apportées par le séquençage haut débit ont stimulé la recherche dans ce domaine. Grace à l'approche de séquençage exomique en particulier, l'identification de nouveaux variants pathogéniques est devenue plus rapide et efficace.

Lors de mon travail de thèse j'ai analysé les résultats de séquençage exomique réalisé pour deux frères atteints d'ANO primaire et issus de parents apparentés. Cette analyse m'a permis d'identifier une mutation homozygote localisée à proximité d'un site accepteur d'épissage dans le gène *SPINK2* qui code pour un inhibiteur de sérine-protéases décrit comme étant fortement exprimé dans le testicule. Afin de confirmer la responsabilité de ce gène dans la survenue de l'infertilité masculine et de comprendre sa fonction, nous avons étudié des souris KO pour *Spink2*. Nous avons observé que les souris knock-out (KO) homozygotes miment parfaitement le phénotype d'azoospermie observé chez les patients. Nous avons montré que la protéine codée par ce gène est exprimée dans l'acrosome à partir du stade de spermatide ronde et que son absence induit un stress cellulaire qui perturbe l'appareil de Golgi et empêche la biogenèse de l'acrosome et la poursuite de la spermiogénèse. Nous avons également pu observer que l'haploinsuffisance de ce gène chez l'homme et chez la souris était associé à un taux élevé d'anomalies morphologiques des spermatozoïdes et une baisse de la mobilité progressive conduisant à une hypofertilité variable. Ces résultats montrent pour la première fois que l'oligo-tératozoospermie et l'azoospermie peuvent constituer un continuum pathologique dû à une même pathogénie moléculaire.

La caractérisation moléculaire et physiopathologique de cette cause génétique nous ouvre plusieurs voies thérapeutiques qui pourraient permettre de rétablir une spermatogenèse fonctionnelle chez les patients concernés et leur offrir une chance de conception intraconjugale.


EMBO
Molecular Medicine
Volume 9 Issue 8 | 1 August 2017
SPINK2 deficiency, Golgi aberrations and azoospermia
IMAGE BY KARINE PERNET-GALLAY, CHRISTOPHE ARNOULT, PIERRE F. RAY
UNIVERSITÉ GRENOBLE ALPES | GRENOBLE | FRANCE
EMBOpress

Published online: May 29, 2017

Research Article

TRANSPARENT PROCESS OPEN ACCESS

EMBO Molecular Medicine

SPINK2 deficiency causes infertility by inducing sperm defects in heterozygotes and azoospermia in homozygotes

Zine-Eddine Kherraf[1,†], Marie Christou-Kent[1,†], Thomas Karaouzene[1], Amir Amiri-Yekta[1,2,3], Guillaume Martinez[1], Alexandra S Vargas[1], Emeline Lambert[1], Christelle Borel[4], Béatrice Dorphin[5], Isabelle Aknin-Seifer[6], Michael J Mitchell[7], Catherine Metzler-Guillemain[7], Jessica Escoffier[1], Serge Nef[4], Mariane Grepillat[1], Nicolas Thierry-Mieg[8], Véronique Satre[1,9], Marc Bailly[10,11], Florence Boitrelle[10,11], Karin Pernet-Gallay[12], Sylviane Hennebicq[1,13], Julien Fauré[2,12], Serge P Bottari[1,14], Charles Coutton[1,9], Pierre F Ray[1,2,‡,*] & Christophe Arnoult[1,‡]

Abstract

Azoospermia, characterized by the absence of spermatozoa in the ejaculate, is a common cause of male infertility with a poorly characterized etiology. Exome sequencing analysis of two azoospermic brothers allowed the identification of a homozygous splice mutation in *SPINK2*, encoding a serine protease inhibitor believed to target acrosin, the main sperm acrosomal protease. In accord with these findings, we observed that homozygous *Spink2* KO male mice had azoospermia. Moreover, despite normal fertility, heterozygous male mice had a high rate of morphologically abnormal spermatozoa and a reduced sperm motility. Further analysis demonstrated that in the absence of Spink2, protease-induced stress initiates Golgi fragmentation and prevents acrosome biogenesis leading to spermatid differentiation arrest. We also observed a deleterious effect of acrosin overexpression in HEK cells, effect that was alleviated by SPINK2 coexpression confirming its role as acrosin inhibitor. These results demonstrate that SPINK2 is necessary to neutralize proteases during their cellular transit toward the acrosome and that its deficiency induces a pathological continuum ranging from oligoasthenoteratozoospermia in heterozygotes to azoospermia in homozygotes.

Keywords azoospermia; exome sequencing; genetics; infertility; spermatogenesis

Subject Categories Genetics, Gene Therapy & Genetic Disease; Urogenital System

DOI 10.15252/emmm.201607461 | Received 13 December 2016 | Revised 14 April 2017 | Accepted 26 April 2017 | Published online 29 May 2017

EMBO Mol Med (2017) 9: 1132–1149

Introduction

The World Health Organization estimates that 50 million couples worldwide are confronted with infertility. Assisted reproduction technologies (ART) initiated 35 years ago by Nobel Prize Winner Robert Edwards have revolutionized the practice of reproductive medicine, and it is now estimated that approximately 15% of couples in Western countries seek assistance from reproductive clinics for infertility or subfertility. Despite technological breakthroughs and advances, approximately half of the couples concerned still fail to achieve a successful pregnancy even after repeated treatment cycles. Alternative treatment strategies should therefore be

1 Genetic Epigenetic and Therapies of Infertility, Institute for Advanced Biosciences, Inserm U1209, CNRS UMR 5309, Université Grenoble Alpes, Grenoble, France
2 CHU de Grenoble, UF de Biochimie Génétique et Moléculaire, Grenoble, France
3 Department of Genetics, Reproductive Biomedicine Research Center, Royan Institute for Reproductive Biomedicine, ACECR, Tehran, Iran
4 Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva 4, Switzerland
5 Laboratoire d'Aide Médicale à la Procréation, Centre AMP 74, Contamine-sur-Arve, France
6 Laboratoire de Biologie de la Reproduction, Hôpital Nord, Saint Etienne, France
7 Aix Marseille Univ, INSERM, GMGF, Marseille, France
8 Univ. Grenoble Alpes / CNRS, TIMC-IMAG, Grenoble, France
9 CHU de Grenoble, UF de Génétique Chromosomique, Grenoble, France
10 Department of Reproductive Biology and Gynaecology, Poissy General Hospital, Poissy, France
11 EA 7404 GIG, Université de Versailles Saint Quentin, Montigny le Bretonneux, France
12 Grenoble Neuroscience Institute, INSERM 1216, Grenoble, France
13 CHU de Grenoble, UF de Biologie de la procréation, Grenoble, France
14 CHU de Grenoble, UF de Radioanalyses, Grenoble, France
*Corresponding author. Tel: +33 4 76 76 55 73; E-mail: pray@chu-grenoble.fr
†These authors contributed equally to this work
‡These authors contributed equally to this work as senior authors

envisaged to improve ART success rate, especially for patients impervious to usual assisted reproductive technologies. Improvement in treatment efficiency essentially depends upon an accurate diagnosis and the characterization of the molecular etiology of the defect. These efforts to better characterize infertility subtypes should first be concentrated on the most severe defects since they generally have a poor prognosis and affected patients would benefit the most from new treatments. Moreover, the most severe phenotypes are more likely to be caused by monogenic defects which are easier to identify. As such, the genetic exploration of non-obstructive azoospermia (NOA), the absence of spermatozoa in the ejaculate due to a defect in spermatogenesis, should be considered a priority. NOA is a common cause of infertility found in approximately 10% of the couples assessed for infertility. Although a genetic etiology is likely to be present in most cases of azoospermia, only a few defective genes have so far been associated with this pathology accounting for a minority of cases. At present, only chromosomal abnormalities (mainly 47XXY, Klinefelter syndrome identified in 14% of cases) and microdeletions of the Y chromosome are routinely diagnosed, resulting in a positive genetic diagnosis in < 20% of azoospermia cases (Tuttelmann *et al*, 2011). The evolution of sequencing technologies and the use of whole-exome or whole-genome sequence (WES/WGS) analysis paves the way to a great improvement in our ability to characterize the causes of genetically heterogeneous pathologies such as NOA.

Spermatogenesis can be subdivided into three main steps: (i) multiplication of diploid germ cells; (ii) meiosis, with the shuffling of parental genes and production of haploid cells; and (iii) spermiogenesis, the conversion of round spermatids into one of the smallest and most specialized cells in the body, the spermatozoa. NOA is expected to be mainly caused by failures in steps 1 and 2, and it is indeed what has been observed in a majority of cases so far. Very recently, defects in six genes were linked to azoospermia in man. Most of these genes code for meiosis-controlling proteins such as TEX11, TEX15, SYCE1, or MCM8, and the absence of the functional proteins induces a blockage of meiosis (Tuttelmann *et al*, 2011; Maor-Sagie *et al*, 2015; Okutman *et al*, 2015; Yang *et al*, 2015; Yatsenko *et al*, 2015). Another WES analysis of two consanguineous families identified likely causal mutations in *TAF4B* and *ZMYND15* (Ayhan *et al*, 2014). Study of *Taf4b* KO mice showed that homozygous mutant males are subfertile with extensive pre-meiotic germ cell loss due to altered differentiation and self-renewal of the spermatogonial stem cell pool, thus illustrating that pre-meiotic block induces NOA. More surprisingly, *ZMYND15* codes for a spermatid-specific histone deacetylase-dependent transcriptional repressor and its absence in mice induced a significant depletion of late-stage spermatids (Yan *et al*, 2010) suggesting that NOA can also be induced by post-meiotic defects.

Here, WES analysis of two brothers with NOA led to the identification of a homozygous truncating mutation in the *SPINK2* gene coding for a Kazal family serine protease inhibitor. Studying KO mice, we observed that homozygous KO animals also suffered from azoospermia thus confirming the implication of SPINK2 in NOA. Furthermore, we observed that SPINK2 is expressed from the round-spermatid stage onwards thus confirming that post-meiotic anomalies can result in NOA. We suggest that SPINK2 is necessary to neutralize the action of acrosomal proteases shortly after their synthesis and before they can be safely stored in the acrosome where they normally remain dormant until their release during the acrosome reaction. We also show that in the absence of SPINK2, protease-induced stress initiates Golgi fragmentation contributing to the arrest of spermatid differentiation and their shedding from the seminiferous epithelium. The characterization of the molecular pathophysiology of this defect opens several novel therapeutic perspectives which may allow the restoration of a functional spermatogenesis.

Results

Medical assessment of two brothers with defective sperm production

Two French brothers (Br1 and Br2), born from second cousin parents (Fig 1A), and their respective wives sought medical advice from infertility clinics in France (Chatellerault, Tours, Poissy, and Grenoble) between 2005 and 2014 after 2 years of unsuccessful attempts to spontaneously conceive. Analyses of their ejaculates (Fig 1B; $n = 5$) evidenced the absence of spermatozoa for the first brother (Br1) and a very low concentration (0–200,000/ml, mean 126,000/ml $n = 5$) for the second (Br2). Moreover, all spermatozoa were immotile and presented an abnormal morphology (pin-shaped head devoid of acrosome; detached flagella) and were not suitable for *in vitro* fertilization (IVF) with intracytoplasmic sperm injection (ICSI). Interestingly, ejaculates of both brothers presented a significant concentration of germ cells ($8.6 \times 10^6 \pm 6.2 \times 10^6$/ml and $9.0 \times 10^6 \pm 7.0 \times 10^6$/ml for Br-1 and Br-2, respectively) likely corresponding to spermatids. As Br1 and Br2 both present a severe default of sperm production with a high number of spermatids in the ejaculate, we believe that they present the same phenotype, likely caused by the same genetic defect. A normal karyotype was observed for both brothers (46,XY), and no deletions of the Y

Figure 1. Azoospermia in two consanguineous brothers.

A Genetic tree of the studied family showing affected brothers Br1 and Br2 illustrating the consanguinity of the parents (P1 and P2).

B Comparisons of ejaculate volume ($n = 5$) and spermograms ($n = 5$) of brothers Br1 and Br2 with those of fertile controls ($n = 35$) evidence the absence of mature sperm and the presence of round cells in the ejaculates. Data represent mean ± SEM. *P*-values are $P = 4 \times 10^{-4}$ (a), $P = 0.6$ (b, non-significant), and $P = 4 \times 10^{-5}$ (c); statistical differences were assessed using *t*-test.

C, D Testis sections from a fertile control and (D) patient Br1 stained with periodic acid–Schiff (PAS). The lumen of tubules from the control is large and mature sperm are present (C), whereas the lumen of most of seminiferous tubules from patient Br1 is filled with non-condensed and early condensed round spermatids and no mature sperm are observed. Scales bars, 100 μm.

E, F In the fertile control (E) seminiferous tubule cross sections, spermatogonia (Sg), spermatocytes (Sc) and spermatids (RS) are regularly layered, whereas the different types of spermatogenic cells are disorganized in patient Br1 (F). Scales bars, 100 μm.

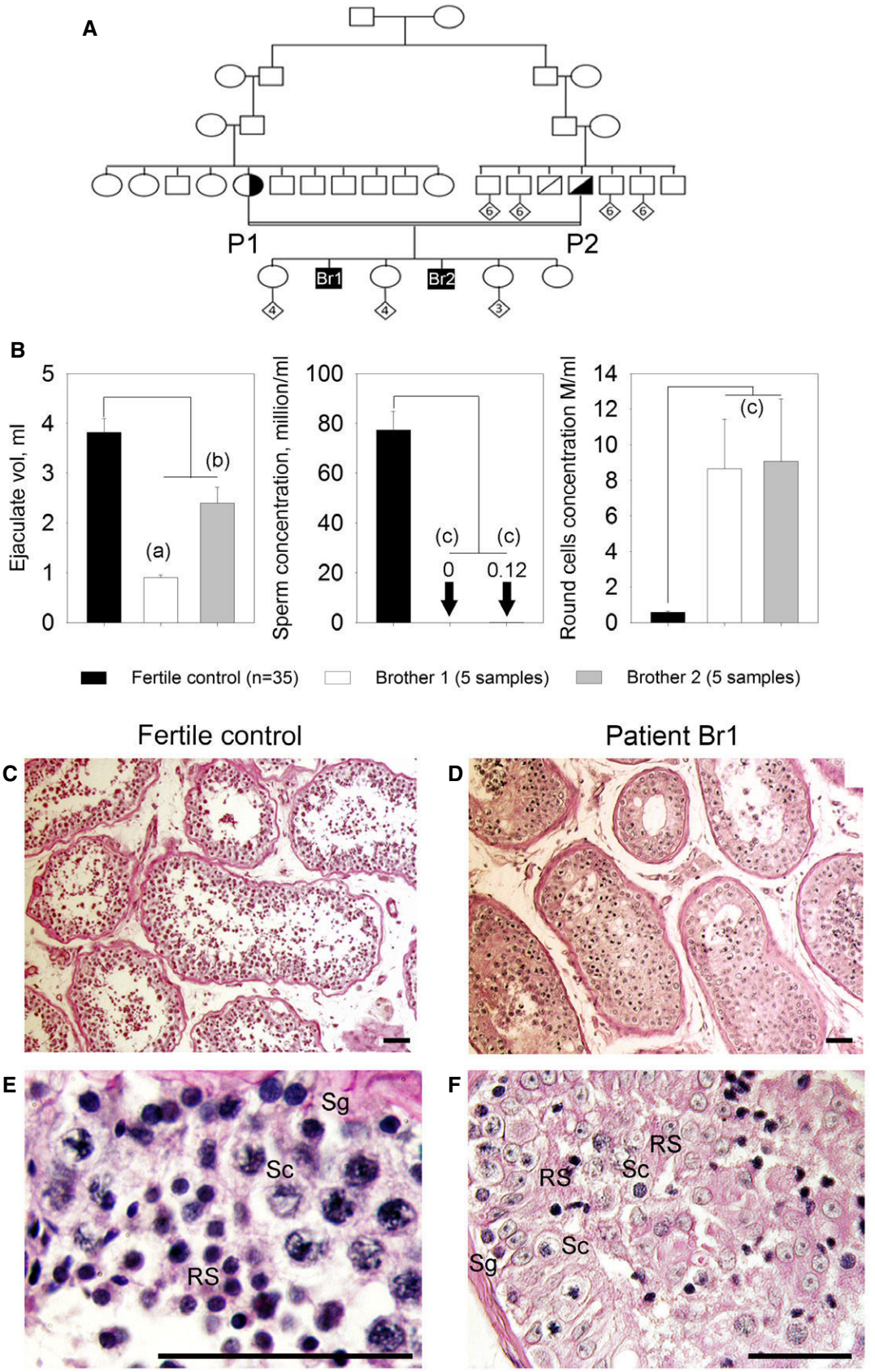


Figure 1.

chromosome were observed at the AZF loci. Testis sperm extraction was carried out twice for Br1 in 2008 and 2014. Each time the recovery was unsuccessful (although a few spermatozoa were observed in fixed dilacerated testicular tissues) suggesting a diagnosis of post-meiotic NOA. Histological analysis of seminiferous tubules obtained from Br1 biopsies showed: (i) a disorganization of the structure of the tubules; (ii) that the lumen of the seminiferous tubules were filled with immature germ cells, an indication of intense desquamation of the germinal epithelium; and (iii) a reduced number of round spermatids, with an overrepresentation of early round spermatids (Fig 1C–F). Brother Br2 has only had spermograms for diagnostic purposes which did not show any ICSI-compatible spermatozoa and has not been able to attempt ART.

Whole-exome sequencing identifies a homozygous truncating mutation in *SPINK2*

Since the brothers were married to unrelated women, we excluded the possibility of a contributing female factor and focused our research on the brothers. Given the familial history of consanguinity, we postulated that their infertility was likely caused by a common homozygous mutation. We proceeded with WES to identify a possible genetic defect(s) which could explain the observed azoospermia. After exclusion of common variants, both bothers carried a total of 121 identical missense heterozygous variants (none appearing as obvious candidate) and only five identical homozygous variants common to both brothers (Appendix Table S1). Among these different genes, only the Chr4:57686748G>C *SPINK2* variant was described to be predominantly expressed in human testis (Appendix Fig S1A) as well as in mouse testis (Appendix Fig S1B). The mutation was validated by Sanger sequencing in both brothers (homozygous) and their parents (heterozygous) (Fig 2A). *SPINK2* thus appeared as the best candidate to explain the human condition. The variant Chr4:57686748G>C was not present in > 121,000 alleles analyzed in the ExAC database (http://exac.broadinstitute.org) and could have an effect on RNA splicing. *SPINK2* is located on chromosome 4 and contains four exons (Fig 2B). The gene codes for a Kazal type 2 serine protease inhibitor also known as an acrosin–trypsin inhibitor. The Ensembl expression database (www.ensembl.org) predicts the presence of four transcripts. We studied the expression of the different transcripts in human testis by RT–PCR, and only one band was present corresponding to NM_021114, ENST00000248701, which codes for a protein of 9.291 kDa consisting of 84 amino acids (Fig 2C). All nucleotide sequences herein refer to this transcript. The identified mutation, c.56-3C>G, is located three nucleotides before exon 2 and may create a new splice acceptor site, leading to a frameshift and premature stop codon in exon 2 and the generation of an abnormal transcript (T1) and/or to the skipping of exon 2 (44 nt) giving rise to an early stop codon at the beginning of exon 3 and the generation of another abnormal transcript (T2) (Fig 2B). To validate these hypotheses, RT–PCR was performed on testicular extract from Br1. Two bands were observed (Fig 2C) and sequenced after isolation of each band following gel electrophoresis. Sequence analysis demonstrated that the bands corresponded to T1 and T2, demonstrating that both abnormal transcripts were present in the patient's testis (Fig 2D). Since the protease inhibitor and binding sites of the protein are coded mostly by exon 3, it is expected that the truncated proteins corresponding to T1 and T2 transcripts are not functional (Appendix Fig S2). Sequencing of Br1's transcripts therefore confirms that the identified splice variant abrogates the production of a full-length protein thereby confirming its role as a deleterious mutation.

Importance of *SPINK2* variants as a cause for human infertility: sequence analysis of a cohort of infertile men with an altered spermatogenesis

We sequenced *SPINK2* whole coding sequences of 611 patients affected by azoo- or oligozoospermia (210 patients with azoospermia, 393 subjects with oligozoospermia and 8 with unspecified cause). Only one variant, identified in patient 105 (P105), was not described in ExAC and was likely deleterious (Appendix Table S2). This variant, c.1A>T (Fig EV1A), abrogates the *SPINK2* start codon and was present heterozygously in P105, a man with oligozoospermia. An alternate start site could potentially be used in the middle of exon 3 allowing the synthesis of a truncated protein of 2 kDa lacking the reactive site and disulfide bonds, both known to be crucial for SPINK2 function (Fig EV1B). However, overexpression of the mutated gene in HEK cells did not produce any portion of the SPINK2 protein indicating that the putative alternative start site is not functional and that the alteration of the initial start site does not permit the synthesis of any part of the SPINK2 protein. This was evidenced by transfecting HEK293 cells with a plasmid containing the full human *SPINK2* ORF sequence with the c.1A>T mutation and a C-terminus DDK-tagged. Extracted proteins were loaded onto a 20% acrylamide gel and detected with anti-DDK or anti-SPINK2 antibodies (Fig EV1C). P105 and his wife, born from non-consanguineous parents, experienced a 5-year period of infertility before giving birth to a healthy boy conceived spontaneously. They sought medical advice 2 years after their son's birth to initiate a second pregnancy. Sperm analysis resulted in the diagnosis of oligozoospermia associated with a reduced percentage of progressive motile spermatozoa (Table EV1). The patient's sperm morphology was assessed with Harris–Shorr staining using the modified David's classification and showed that 34–39% of sperm had a normal morphology ($n = 2$). The main defects observed were abnormal acrosome (34–39%) and defective neck–head junction (40–46%), defects that are similar to those observed in patient Br2.

This analysis indicates that *SPINK2* defects are extremely rare with an allelic frequency of approximately 1/1,200 in the cohort of infertile men analyzed. The rarity of *SPINK2* variants and the fact that P2, the father of Br1 and Br2, also harboring a heterozygous mutation, presents in a milder phenotype than P105 could indicate that *SPINK2* haploinsufficiency induces a milder phenotype of oligozoospermia with an incomplete penetrance on infertility.

Homozygous *Spink2* KO mice have azoospermia due to a spermiogenesis blockade at the round-spermatid stage

In order to confirm that the absence of SPINK2 leads to azoospermia, homozygous *Spink2* KO ($^{-/-}$) mice were obtained and their reproductive phenotype was studied. We first performed qRT–PCR on *Spink2*$^{+/+}$ and S*pink2*$^{-/-}$ testis mRNA extracts to validate the absence of *Spink2* mRNA and thus of protein. Contrary to what was observed in WT littermates, we observed no *Spink2* amplification in KO males, confirming Spink2 deficiency (Appendix Fig S3). Males

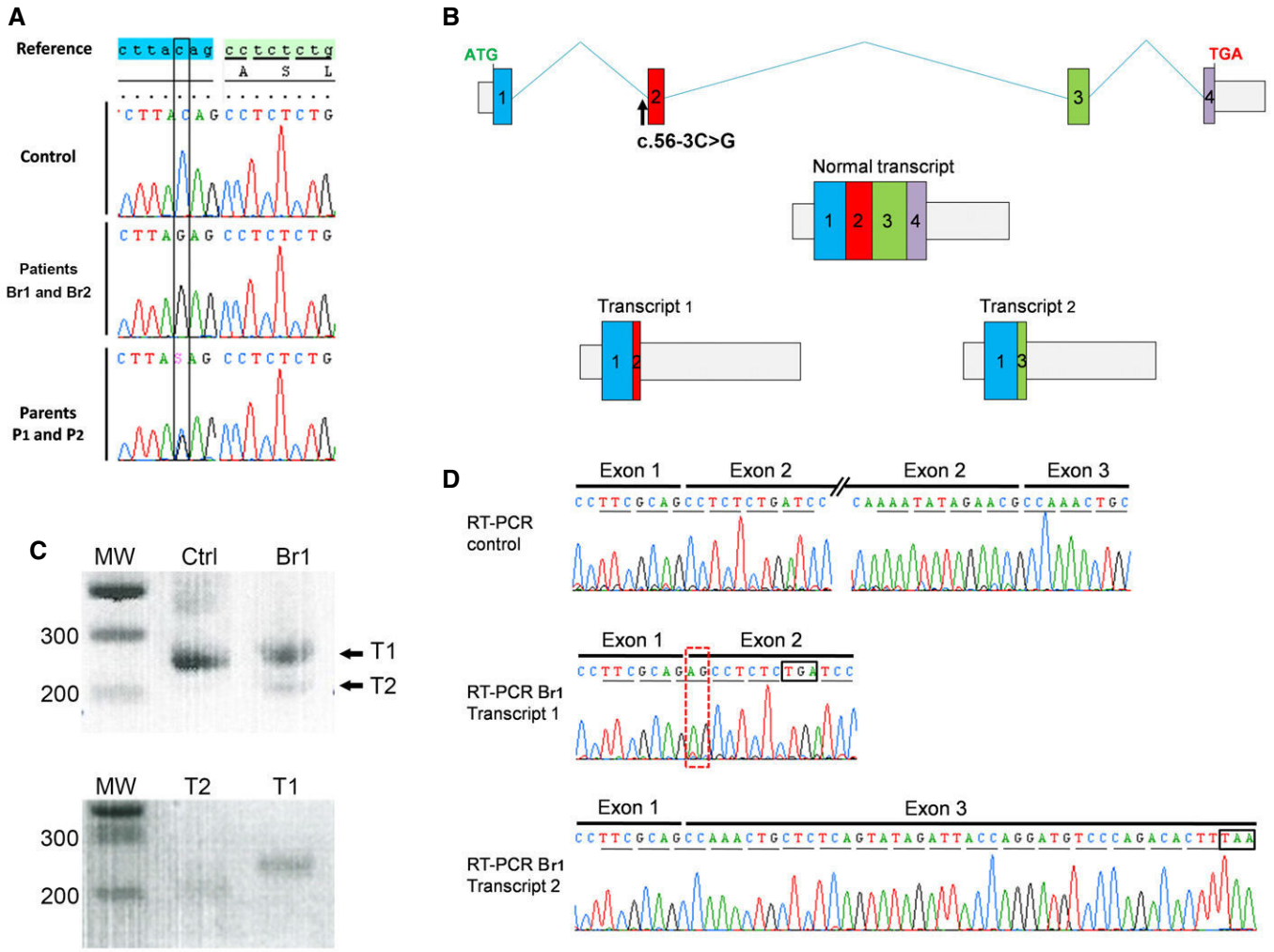


Figure 2. Identification of a *SPINK2* variant (c.56-3C>G) by exome sequencing and its consequences on splicing and translation.

A The identified variant, homozygous in patients 1 and 2 and heterozygous in their parents, is located three nucleotides before exon 2 and creates an AG that immediately precedes the original AG splice acceptor site.

B If recognized during splicing, this new acceptor site is expected to add two nucleotides (AG) at the beginning of exon 2, inducing a frameshift leading to a stop codon 3 amino acids later (transcript 1). The non-recognition of the abnormal acceptor site is expected to induce the skipping of exon 2 (transcript 2). The first stop codon can be observed 15 codons after the mis-inserted exon 3.

C RT–PCR of mRNA extracts from fertile control (Ctrl) and the brother Br1. Results show one band for Ctrl. The sequencing of this band showed that it corresponds to transcript NM_021114. For Br1, two bands were present, named T1 and T2. Bottom gel shows T1 and T2 after gel isolation.

D Transcripts T1 and T2 were collected and sequenced: T1 showed the insertion of an additional AG (red-dashed rectangle) leading to a premature stop codon (black box), whereas transcript T2 showed that exon 2 had been excised; these two transcripts correspond to the expected transcripts 1 and 2 from panel (B). Stop codons are shown in black boxes.

were completely infertile, whereas no reproductive defects were observed in females (Fig 3A1). Homozygous KO mice had comparatively smaller sized testes and a testis/body weight ratio half that of their wild-type (WT) littermates [3.63 ± 0.21 in WT and 1.77 ± 0.03 in KO (Fig 3A2)]. Furthermore, there was a complete absence of spermatozoa in *Spink2*$^{-/-}$ caudal epididymis (Fig 3A3) which only contained round cells likely corresponding to round spermatids and multinucleated cells, known as symblasts. Histological studies of KO seminiferous tubules stained with periodic acid–Schiff (PAS) revealed the presence of germ cells up to the early round-spermatid stage but condensed and elongated spermatids and mature spermatozoa were completely absent, contrary to WT (Fig 3B1 and C1). The lumen of the seminiferous tubules of Spink2$^{-/-}$ males contained round cells and symblasts (Fig EV2A and B), a result in agreement with observations of the cellular content of the cauda epididymis, which showed the presence of round cells only (Fig EV2C). In contrast to what was observed in WT (Fig 3B2), sections of caudal epididymis confirmed the absence of spermatozoa and the presence of symblasts and round cells (Fig 3C2). Comparing PAS staining of *Spink2*$^{+/+}$ and *Spink2*$^{-/-}$ seminiferous tubules, we noticed that contrary to WT, *Spink2*$^{-/-}$ round spermatids did not contain an acrosomal vesicle, suggesting

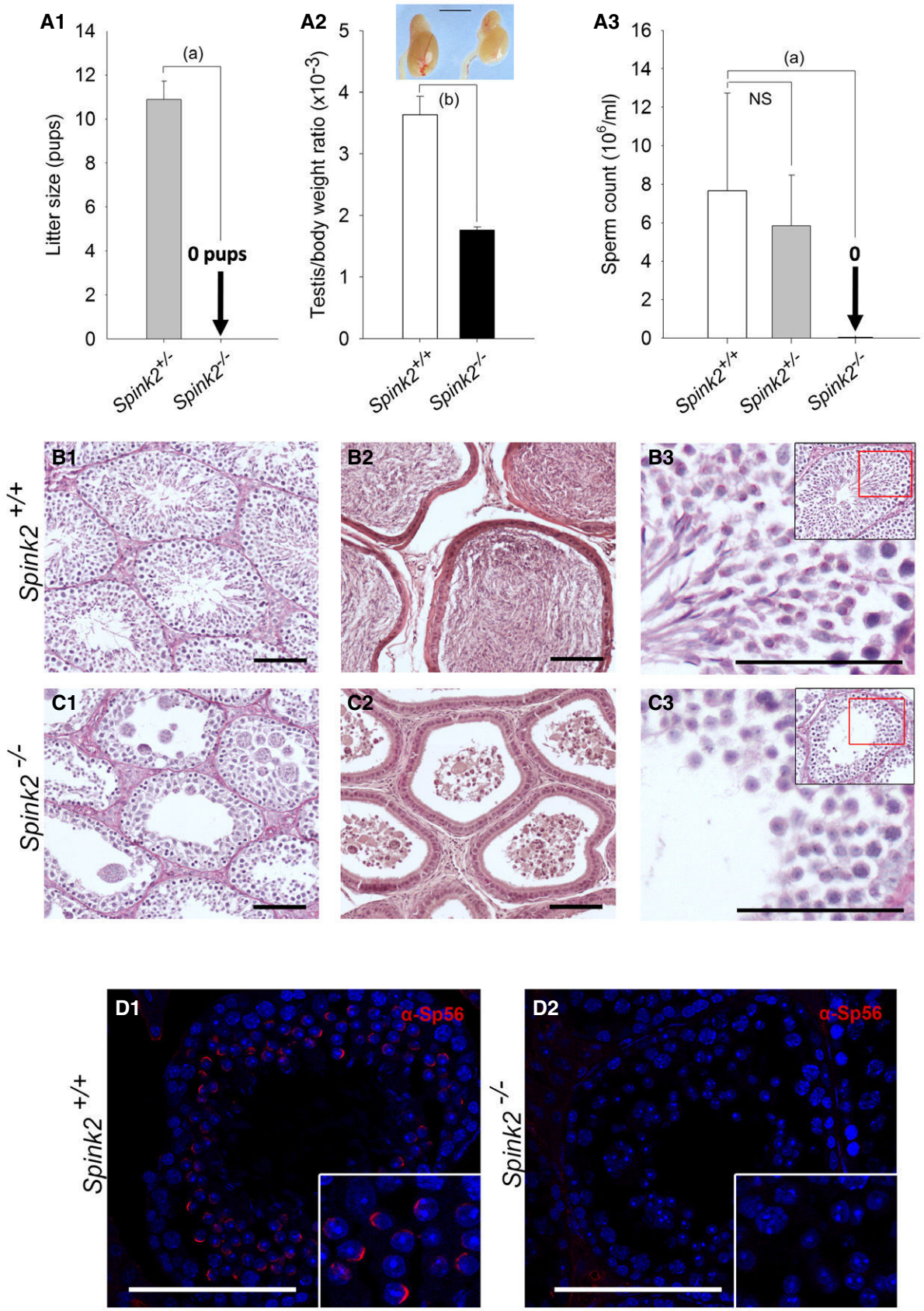

A1
Litter size (pups)
(a)
0 pups
Spink2+/-
Spink2-/-
A2
Testis/body weight ratio (x10-3)
(b)
Spink2+/+
Spink2-/-
A3
Sperm count (10^6/ml)
(a)
NS
0
Spink2+/+
Spink2+/-
Spink2-/-
B1
B2
B3
Spink2 +/+
C1
C2
C3
Spink2 -/-
D1
α-Sp56
Spink2 +/+
D2
α-Sp56
Spink2 -/-

Figure 3.

Figure 3. *Spink2*$^{-/-}$ males are infertile and azoospermic, and spermatogenesis presents a post-meiotic blockade.

A1 Litter size of *Spink2*$^{-/-}$ *and Spink2*$^{+/-}$ males mated with wild-type females (n = 5).

A2 Testis/body weight ratio for WT and *Spink2*$^{-/-}$ mice (n = 6) and morphology and size of wild-type and *Spink2*$^{-/-}$ testes of male siblings. Scale bar, 5 mm.

A3 Sperm concentrations from the cauda epididymis of wild-type, *Spink2*$^{+/-}$, *and Spink2*$^{-/-}$ male testes (n = 10).

B, C Histological comparisons of testis and epididymis from WT and *Spink2*$^{-/-}$ mice. (B1, C1) Periodic acid–Schiff (PAS) staining of seminiferous tubule cross sections shows complete spermatogenesis in WT (B1) contrary to *Spink2*$^{-/-}$ mice (C1), where condensed, elongated spermatids and mature sperm are absent. (B2, C2) Sections of epididymis stained with eosin/hematoxylin. In the lumen of tubules from WT mice, mature sperm are present (B2), whereas only round cells and multinucleated symblasts occupy the lumen of tubules from *Spink2*$^{-/-}$ mice (C2). (B3, C3) Enlargement of seminiferous tubule sections stained with PAS evidences deep pink staining in round spermatids, which corresponds to the acrosome in WT mice (B3), whereas round spermatids from *Spink2*$^{-/-}$ mice present no deep pink staining, indicating that the acrosome is not formed (C3). Scale bars, 100 μm.

D1, D2 Immunofluorescence experiments using an anti-Sp56 antibody (red staining) confirm the presence of the acrosome in seminiferous tubule sections from WT contrary to those from *Spink2*$^{-/-}$ mice, where no staining is observed. Scale bars, 100 μm.

Data information: Data represent mean ± SEM. *P*-values are $P = 1 \times 10^{-5}$ (a) and $P = 1 \times 10^{-4}$ (b); statistical differences were assessed using *t*-test. NS, not statistically significant.

that the absence of Spink2 prevents acrosome biogenesis (Fig 3B3 and C3). This point was confirmed by immunofluorescent staining using the Sp56 antibody, a specific marker of the acrosome (Kim *et al*, 2001) (Fig 3D1 and D2). We then identified the spermatogonia using an anti-PLZF antibody (Zhang *et al*, 2014) (Fig EV3A and B) and observed no significant difference in the median number of spermatogonia per tubule (n = 30) between *Spink2*$^{+/+}$ and *Spink2*$^{-/-}$ mice (Fig EV3C). These results indicate that the absence of Spink2 does not impact spermatogonial survival but leads to an early arrest of round-spermatid differentiation. Overall, the *Spink2*$^{-/-}$ mouse phenotype perfectly mimics the human condition and confirms that SPINK2 deficiency is involved in human azoospermia.

SPINK2 is an acrosomal protein

In order to further investigate the molecular pathogeny of this SPINK2-dependent azoospermia, we determined the localization of SPINK2 in human and mouse testis. We first verified the specificity of a SPINK2 antibody through Western blot (WB) and immunofluorescence (IF) experiments on HEK293 cells overexpressing human SPINK2. In Western blots, the SPINK2 antibody recognized three bands of less than 17 kDa weight, likely corresponding to oligomeric complexes (Appendix Fig S4A). No bands appeared in non-transfected cells. Moreover, the overexpressed SPINK2 featured a DDK-tag which was recognized by an anti-DDK-tag antibody revealing three bands of identical molecular mass (Appendix Fig S4B). No bands were observed when the primary antibody was omitted. SPINK2 expression was also studied by IF and confocal microscopy. Transfected cells displayed a cytoplasmic staining, whereas no staining was observed in non-transfected cells (Appendix Fig S4C). Taken together, these results demonstrate the specificity of this antibody in WB and IF experiments. Next, the localization of SPINK2 was determined by IF in human and mouse seminiferous tubule cross sections and in mature sperm (Fig EV4). In mouse, SPINK2 was present in the acrosomal vesicle from the beginning of the acrosome's biogenesis at the round-spermatid stage as indicated by a colocalization with Sp56, a marker of the acrosome (Fig EV4A and B). In accordance with the results shown in Fig 3D2, no SPINK2 staining was observed in *Spink2*$^{-/-}$ testis cross sections (Fig EV4C). A similar localization was observed for SPINK2 in human seminiferous tubule sections (Fig EV4D). Finally, we observed that SPINK2 remains present in the acrosome of human and mouse mature spermatozoa (Fig EV4E and F).

Ultrastructure of *Spink2*$^{-/-}$ round spermatids shows that fusion of proacrosomal vesicles is hampered and that the Golgi apparatus is fragmented

We showed that SPINK2 is located in the acrosome and that its absence prevents acrosome biogenesis. To understand the reasons for the absence of acrosome biogenesis, we performed transmission electronic microscopy (EM) to study the ultrastructure of round spermatids from *Spink2*$^{-/-}$ males (Fig 4). In wild-type round spermatids, proacrosomal vesicles generated by the Golgi apparatus docked in a specialized area of the nuclear envelope (NE) and fused together to form a giant acrosomal vesicle (Fig 4A). Contrary to WT, in *Spink2*$^{-/-}$, the proacrosomal vesicles generated by the Golgi apparatus of round spermatids were mostly unable to fuse (Fig 4B2, white arrowheads), likely explaining the absence of acrosome biogenesis. Moreover, the Golgi apparatus from *Spink2*$^{-/-}$ animals produced abnormal proacrosomal vesicles of irregular sizes (Fig 4B2) and showed a considerable disorganization with a decreased proportion of flattened membrane stacks (Fig 4B2) displaying shorter lengths (Fig 4C). Acrosome biogenesis is dependent on the simultaneous synthesis of vesicles by the Golgi apparatus and the modification of the nuclear envelope (NE) facing the Golgi apparatus, with tight apposition of both nuclear membranes and aggregation of a nuclear dense lamina (NDL) on the nuclear side of the inner nuclear membrane (Kierszenbaum *et al*, 2003). In *Spink2*$^{-/-}$ round spermatids, the densification of the NE appears to occur normally and the NDL is clearly visible in EM (Fig 4B2). Using IF, the modification of the NE facing the Golgi apparatus was followed with an anti-Dpy19l2 antibody. We indeed had previously shown that Dpy19l2 participates in linking the acrosome to the nucleus and that it is located in the nuclear membrane facing the forming acrosome (Pierre *et al*, 2012) and is thus a component of this specialized area of the nuclear envelope. In costaining experiments using anti-Dpy19l2 and anti-GM130 antibodies to stain the nuclear envelope facing the acrosomal vesicle (evidenced by the NDL in EM) and the Golgi apparatus, respectively, we found that in WT round spermatids, the Golgi apparatus is either located immediately in front of the NDL in the early phase of acrosome biogenesis or, at a slightly later stage, lies adjacent to it (Fig 4D1 and D2). In contrast to WT, the Golgi apparatus of *Spink2*$^{-/-}$ round spermatids was positioned randomly around the nucleus, often found on the opposite side of the NDL (Fig 4D3–D6) indicative of a disruption of the polarity of the NDL and of the Golgi apparatus, which should both be located at the apical face of the round spermatid.

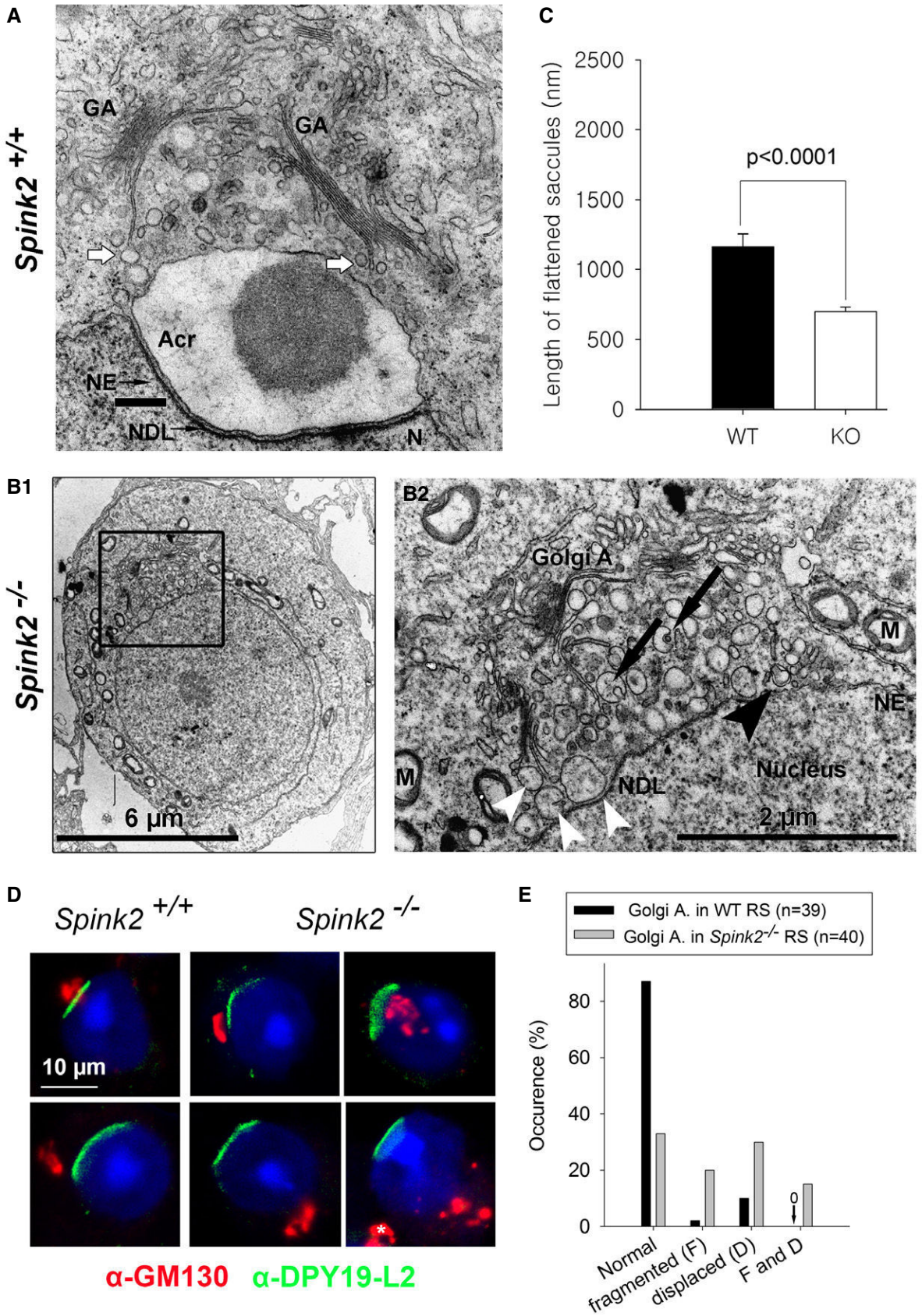

A
Spink2 +/+
GA
GA
Acr
NE
NDL
N
C
Length of flattened saccules (nm)
2500
2000
1500
1000
500
0
p<0.0001
WT
KO
B1
Spink2 -/-
6 µm
B2
Golgi A
M
NE
Nucleus
M
NDL
2 µm
D
Spink2 +/+
Spink2 -/-
10 µm
α-GM130
α-DPY19-L2
E
Golgi A. in WT RS (n=39)
Golgi A. in Spink2-/- RS (n=40)
Occurence (%)
80
60
40
20
0
Normal
fragmented (F)
displaced (D)
F and D

Figure 4.

Figure 4. Lack of Spink2 prevents the fusion of proacrosomal vesicles and induces a disorganization of the Golgi apparatus.

A Partial section of a WT round spermatid observed by EM showing the early biogenesis of the acrosome (Acr) due to the continuous formation and aggregation of small vesicles (white arrows) coming from the Golgi apparatus (GA). The nuclear envelope (NE) facing the acrosome has a specific organization and is associated with the nuclear dense lamina (NDL). N, nucleus. Scale bar, 400 nm.

B Ultrastructure of the Golgi apparatus in *Spink2*$^{-/-}$ round spermatid observed by EM. (B1) Ultrastructure of a *Spink2*$^{-/-}$ round spermatid observed at low magnification. The black box corresponds to the Golgi apparatus and is enlarged in (B2). (B2) In the absence of Spink2, vesicles do not aggregate at the nuclear envelope although modification of the NE and formation of the NDL occur. Unfused vesicles of different sizes accumulate in the cytoplasm with very few docking on the nuclear envelope (white arrowhead). Moreover, the GA shows disorganization with strong decrease or absence of stacks of flattened membranes. Finally, microautophagy-like structures and vesicles with a double membrane (black arrowhead) are observed around the GA (black arrows). M, mitochondria. Scale bars, 6 μm (B1) and 2 μm (B2).

C The length of flattened saccules is statistically reduced in *Spink2*$^{-/-}$ round spermatids (WT saccules, n = 74; and KO saccules, n = 136). Data represent mean ± SEM; the statistical difference was assessed with t-test, P-value as indicated.

D Absence of Spink2 induces Golgi apparatus fragmentation and mislocalization. (D1, D2) IF experiments using an anti-Dpy19l2 antibody marking the specific NE facing the NDL (green staining) and an anti-GM130 antibody marking the cis-Golgi (red staining) show that the Golgi apparatus (GA) is a compact structure and located either in front of the NDL or close to it in WT round spermatids (normal). (D3–D6) In contrast, similar double staining of round spermatids from *Spink2*$^{-/-}$ mice shows that only one-third of GA are compact and normally placed (D3) and the other GA are either displaced (D4), fragmented (D5), or both (D6). In panel (D6), white asterisk corresponds to a GA belonging to a different cell.

E Quantification of the morphology and the relative localization of the GA and Dpy19l2 staining in WT (n = 40) and *Spink2*$^{-/-}$ (n = 39) round spermatids.

Disjunction of the Golgi apparatus and of the NDL was also observed in EM (Fig EV5A). Moreover, anti-GM130 staining in *Spink2*$^{-/-}$ round spermatids appeared disseminated and punctuated, confirming the disorganization of the Golgi apparatus and indicating a fragmentation of the organelle (Fig 4D4–D6).

Interestingly, EM observations of *Spink2*$^{-/-}$ round spermatids showed the presence of multivesicular bodies, a known biomarker of microautophagy (Li *et al*, 2012) (Fig EV5B). These latter structures strongly suggest that the absence of Spink2 activates an uncharacterized self-degradation pathway. Visual signs of the initial events of microautophagy occurring at the Golgi apparatus level are the engulfment of vesicles (Fig 4B2, black arrows) and the presence of already engulfed vesicles (Fig 4B2, black arrowhead). We note that the thorough examination of round spermatids on EM images did not reveal any detectable signs of morphological hallmarks of apoptosis such as chromatin condensation, fragmentation of the plasma membrane, and the presence of apoptotic bodies. Moreover, no differences in DNA damage were observed between WT and *Spink2*$^{-/-}$ round spermatids when assessed by terminal deoxynucleotidyl transferase (TdT)-mediated deoxyuridine triphosphate (dUTP)-nick-end labeling (TUNEL) test (Appendix Fig S5). Altogether, these results suggest that the absence of Spink2 at the round-spermatid stage does not activate the apoptotic pathway.

Rescue of acrosin-induced cell proliferation defects by coexpression with SPINK2

During spermatid differentiation, several enzymes, involved in sperm penetration through the protective layers surrounding the oocytes, accumulate in the acrosomal vesicle. Among these different enzymes, several proteases have been described to play a key role, including acrosin, believed to be the main acrosomal protease (Liu & Baker, 1993). Acrosin, a trypsin-like protease, is synthesized in the reticulum as a zymogen (proacrosin), transits through the Golgi apparatus, and accumulates in the acrosomal vesicle. Autoactivation of acrosin is pH-dependent and occurs at a pH > 6 (Meizel & Deamer, 1978) leading to sequential N-ter and C-ter cleavages of the proacrosin (46 kDa), eventually giving active forms of acrosin with lower weights of 20–34 kDa (Baba *et al*, 1989; Zahn *et al*, 2002). Since the pH of both the endoplasmic reticulum and the Golgi apparatus is greater than 6 (Rivinoja *et al*, 2012), we postulated that Spink2, as a serine peptidase inhibitor, prevents acrosin autoactivation in these cellular compartments, thus preventing cellular stress induced by uncontrolled protease activation. Such stress would cause cellular defects including Golgi apparatus destabilization and defective acrosome biogenesis leading to spermatid differentiation arrest. To test this hypothesis, heterologous expressions of human C-terminus DDK-tagged proacrosin (ACR), SPINK2, or both were carried out in HEK293 cells and the kinetics of cell proliferation were followed using xCELLigence Real-Time Cell Analysis (RTCA) technology for the different conditions. It is worth noting that no members of the SPINK family are reported to be expressed in HEK293 cells. Analyses of kinetics showed that proacrosin expression quickly led to cell proliferation arrest and detachments in contrast to what was observed in the control condition (Fig 5A and B). Interestingly, cells showed a normal proliferation when SPINK2 was coexpressed with proacrosin (Fig 5A and B), therefore demonstrating that cell stress and damages induced by the proacrosin were prevented by SPINK2 coexpression. The presence of the different overexpressed proteins was verified in the different conditions by Western blotting using the SPINK2 antibody (Fig 5C), an anti-acrosin (Fig 5D), and the anti-DDK (Fig 5E) antibodies. In extracts of HEK293 cells transfected with proacrosin only and revealed with an anti-acrosin antibody (Fig 5D), two bands were present at around 46 and 34 kDa. The latter (red arrowhead) likely corresponds to the active form of acrosin resulting from the cleavage of proacrosin upon autoactivation. This band was not present when acrosin was coexpressed with SPINK2 or in non-transfected cells (control). Moreover, a closer inspection of the band around 46 kDa in the extracts of cells transfected with proacrosin only or proacrosin + SPINK2 shows that this band is of lower MW and was less intense in "acrosin" extract compared to "acrosin + SPINK2" cell extract, showing the process of successive cleavages occurring during proacrosin autoactivation (Zahn *et al*, 2002). Similar results were obtained with the anti-DDK antibody (Fig 5E). It is worth noting that anti-DDK antibody immunodecorates the zymogen form only and not the active form of acrosin because the C-terminus containing the DDK-tag is cleaved upon autoactivation. Western blot results thus demonstrate that coexpression of proacrosin with SPINK2 prevented its autoactivation. We can thus conclude that in the absence of a serine peptidase inhibitor, proacrosin can autoactivate and induces a cellular stress leading to

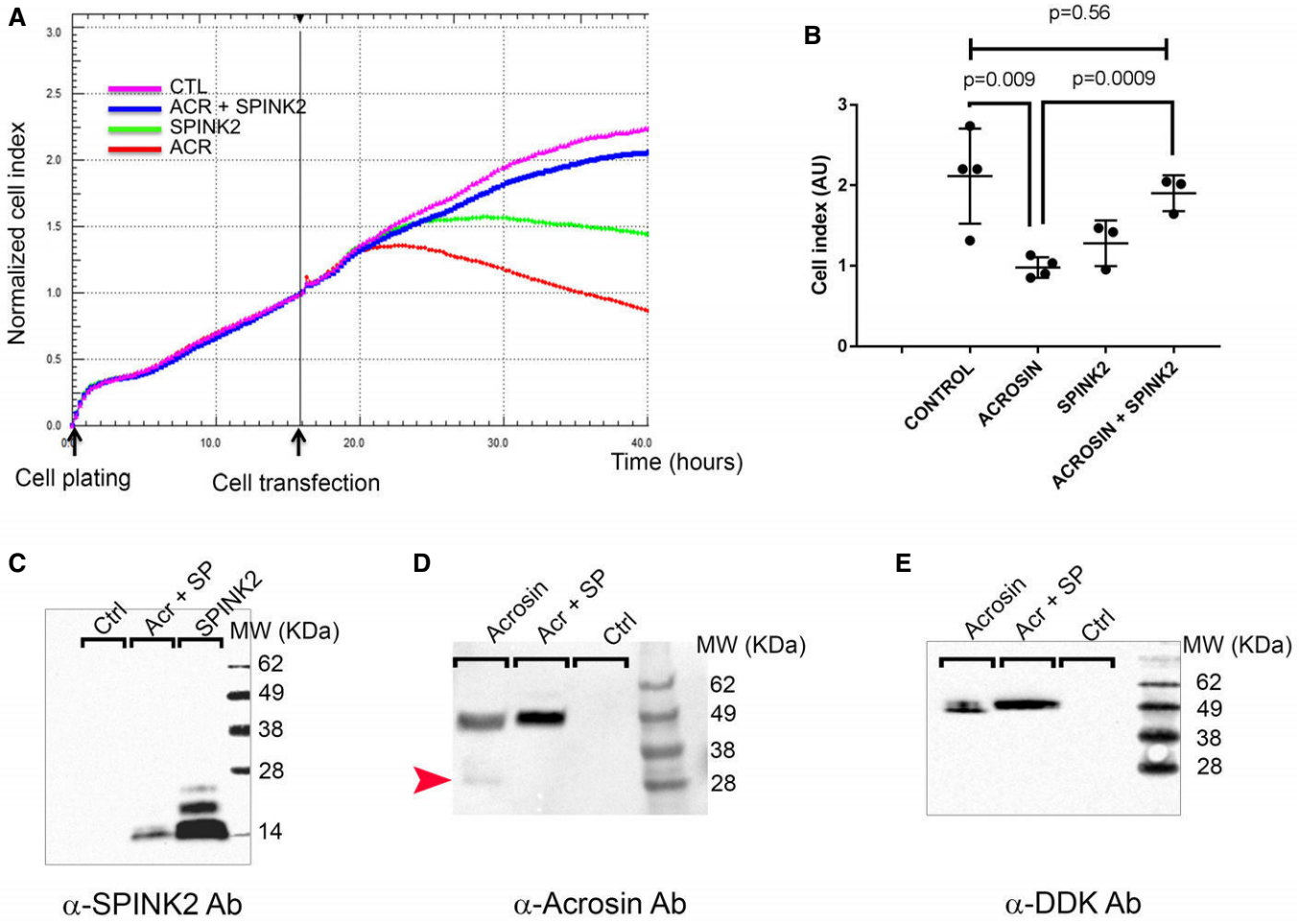


Figure 5. Heterologous expression of proacrosin in HEK293 cells induces acrosin activation and cell proliferation arrest, a phenotype rescued by SPINK2 coexpression.

A Representative kinetics of HEK293 cell proliferation measured with Real-Time Cell Analysis (RTCA) technology in different conditions as indicated. Each point corresponds to the mean of four technical replicates measured simultaneously. Black arrows indicate the time of cell plating (t = 0 h) and introduction of the different plasmids in the cell chambers (t = 16 h).

B Scatter plots showing the mean and SD of the cell index measured at 40 h after plating (corresponding to cell proliferation and detachment) in different transfection conditions and measured for three independent biological replicates. Statistical differences were assessed using t-test, P-values as indicated.

C Western blot using an anti-SPINK2 antibody showing the expression of SPINK2 in cell extracts of HEK293 cells transfected with different plasmids containing SPINK2 (SP) or acrosin and SPINK2.

D Representative Western blot using an anti-acrosin antibody. In extracts of HEK293 cells transfected with proacrosin only (lane "acrosin"), two bands were observed, one at around 34 kDa and corresponding to the active form of acrosin (red arrowhead) and one at 46 kDa and corresponding to the zymogen form, whereas in extracts of HEK293 cells transfected with proacrosin and SPINK2 (lane "Acr + SP"), only the zymogen form was observed. Equal protein loading was verified by stain-free gel technology (Taylor & Posch, 2014) and Western blots against tubulin (Appendix Fig S6). Note that the zymogen form in lane "acrosin" has a slightly lower mass and that the band is less intense than that in lane "Acr + SP".

E Representative Western blot using an anti-DDK antibody showing the expression of the proacrosin zymogen form in HEK293 cells transfected with different plasmids as indicated. Note that once more, the zymogen form in the lane "acrosin" has a slightly lower mass and that the band is less intense than that in lane "Acr + SP". Similarly, equal protein loading was verified by stain-free gel technology (Appendix Fig S6).

cell proliferation arrest and cell detachment, a phenotype similar to that observed in round spermatids from *Spink2*$^{-/-}$ males.

SPINK2 haploinsufficiency induces sperm defects with incomplete penetrance in man

Only one additional subject, P105, was identified with a *SPINK2* heterozygous deleterious variant, and we cannot be sure that this variant is the cause of the patient's oligozoospermia. Two arguments could in fact suggest that *SPINK2* haploinsufficiency is not deleterious: (i) Br1 and Br2's father is SPINK2 heterozygous and has conceived six children spontaneously, and unfortunately, we could not obtain sperm samples to characterize this man's sperm parameters; and (ii) because heterozygous *Spink2*$^{+/-}$ male mice are fertile, they did not produce litters of reduced size (Fig 3A). We however carried out a detailed characterization of *Spink2*$^{+/-}$ and *Spink2*$^{+/+}$

sperm parameters to address the question of the impact of SPINK2 haploinsufficiency on mouse spermatogenesis. Heterozygous males displayed a significant increase in teratozoospermia (Fig 6A). Abnormal spermatozoa showed non-hooked heads, isolated heads, or a malformed base of the head (Fig 6B). Moreover, sperm motility of heterozygous males was impaired with lower total and progressive motility (Fig 6C). We note that the observed defects are very similar to those observed in the heterozygous patient P105 (Table EV1). We can therefore conclude that in mice, *SPINK2* haploinsufficiency induces asthenoteratospermia with no alteration of reproductive fitness, whereas in man it leads to oligoteratozoospermia with variable expressivity and infertility with an incomplete penetrance.

Discussion

SPINK family emerges as an important family for human genetic diseases

SPINK proteins are serine protease inhibitors containing one or several Kazal domains which interact directly with the catalytic domains of proteases blocking their enzymatic activity (Rawlings *et al*, 2004). The Kazal domain structure contains three disulfide bonds which are highly conserved. Different SPINK proteins are specifically expressed in different tissues and inhibit a number of serine proteases, such as secreted trypsin in the pancreas, acrosin in sperm, or kallikrein in the skin. Downregulation of the activity of different SPINK proteins leads to severe pathologies such as chronic pancreatitis and Netherton syndrome. In the pancreas, trypsin is produced as an inactive zymogen to prevent cell damage, yet the trypsinogen is occasionally able to autoactivate. This protease activity is then blocked by SPINK1. Chronic pancreatitis can be triggered by mutations of *SPINK1* that decrease or suppress its trypsin inhibitor function, leading to cell distress (Chen *et al*, 2000; Witt *et al*, 2000). In the skin, kallikrein-related peptidases are controlled by SPINK5 and unopposed kallikrein-peptidase activity due to *SPINK5* deficiency leads to Netherton syndrome, a severe skin disease (Furio & Hovnanian, 2014). SPINK6 and SPINK9 are also expressed in the skin, and altered expression levels are associated with atopic dermatitis or psoriasis (Redelfs *et al*, 2016). The other members of the SPINK family, including SPINK2, have not yet been associated with a human pathology. Here, we have clearly demonstrated that the absence of SPINK2 induces azoospermia, a severe infertility phenotype, emphasizing the importance of this family in human pathologies.

Role of SPINK2 during spermiogenesis

We have shown that SPINK2 is located in the acrosomal vesicle in round spermatids and remains present in mature spermatozoa, suggesting that this protein is necessary for spermiogenesis and sperm survival. SPINK proteins are known to control protease activities in different tissues (Witt *et al*, 2000; Rawlings *et al*, 2004; Ohmuraya *et al*, 2012; Furio & Hovnanian, 2014) and since SPINK2 is located in the acrosome, it very likely neutralize acrosomal proteases before their release prior fertilization. Several proteases have been described to be present in the acrosome (Arboleda & Gerton, 1987; Kohno *et al*, 1998; Cesari *et al*, 2004). Among these, acrosin (Acr) was the first to be described and is the acrosomal protein which has been the most studied. Acrosin is present in the acrosome as a zymogen called proacrosin (Huang-Yang & Meizel, 1975) which is predicted to be activated during the acrosome reaction (Brown & Harrison, 1978) upon a rise in acrosomal pH to 7 which induces pH-dependent proacrosin autoactivation (Baba *et al*, 1989). Before the acrosome reaction, at least two mechanisms prevent autoactivation: The first is the acrosomal acidic pH which is below 5, which blocks autoactivation of proacrosin (Meizel & Deamer, 1978); and the second is the presence in the sperm of a non-fully characterized proacrosin conversion inhibitor of 12 kDa which has been purified from boar acrosome (Kennedy *et al*, 1982). The presented results strongly suggest that this protein is in fact SPINK2. Proacrosin is however produced in the endoplasmic reticulum and transits through the Golgi apparatus, two cellular compartments with a pH of approximately 7 and 6.5, respectively. In these compartments, autoactivation of proacrosin is thus possible and would result in the release of active acrosin within these apparatuses. We therefore believe that SPINK2, which transits through the same cellular compartments, quenches this premature protease activity and prevents the described cascade of events leading to azoospermia. This hypothesis is supported by heterologous expression experiments: We have indeed demonstrated that proacrosin expression in HEK293 cells induces (i) autoactivation of proacrosin and (ii) cellular proliferation arrest and cell detachment. Moreover, cellular toxicity of proacrosin expression is prevented by SPINK2 coexpression, showing the ability of SPINK2 to inhibit acrosin activity.

One of the most striking effects of SPINK2 deficiency is the fragmentation of the Golgi apparatus, a key organelle for protein processing and translocation, in particular for membrane proteins. The notable strong desquamation of the germinal epithelium may be due to severe changes in membrane protein composition resulting from a defective Golgi apparatus function.

Impact of SPINK2 deficiency

We have shown that the absence of SPINK2 in round spermatids leads to several subcellular defects targeting the process of proacrosomal vesicle formation by the Golgi apparatus. The observed abnormalities include the disorganization and delocalization of the Golgi apparatus, the presence of vesicles of various sizes, and the absence of proacrosomal vesicle fusion. The absence of SPINK2 likely allows proacrosin autoactivation within the reticulum and the Golgi apparatus compartments, leading to the above-described subcellular defects. It was previously shown that transgenic expression of porcine proacrosin in mice led to post-meiotic cell death and oligozoospermia, supporting the hypothesis that unbalanced expression of acrosin/Spink2 is deleterious (O'Brien *et al*, 1996). Interestingly, we have demonstrated that the cell responds to this stress by activating a microautophagy-like pathway: First, we showed that larger vacuoles engulfed small vacuoles, likely leading to the observed heterogeneity in vacuole size in the vicinity of the Golgi apparatus; and secondly, multivesicular bodies, a hallmark of microautophagy (Li *et al*, 2012), were clearly observed within Spink2$^{-/-}$ round spermatids, whereas they were never observed in WT. Furthermore, the lack of various SPINK proteins induces autophagy-induced cell death in regenerating Hydra (Chera *et al*, 2009)

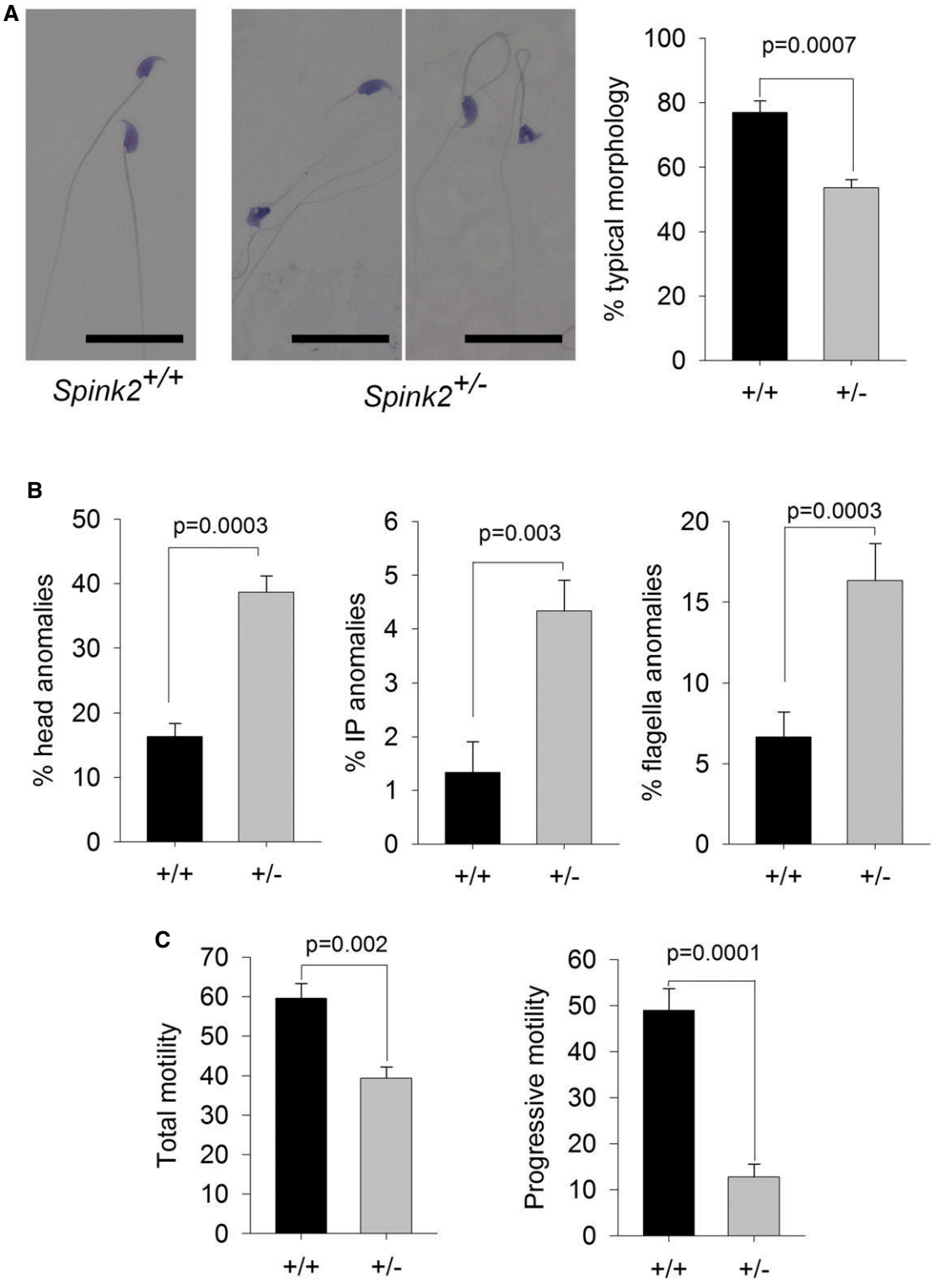


Figure 6. Sperm from *Spink2*$^{+/-}$ heterozygous mice exhibit morphological defects and low motility.

A Light microscopy analysis of sperm from *Spink2*$^{+/-}$ heterozygous mice reveals the presence of numerous non-typical forms of sperm. Scale bars, 25 μm. Graph on the right shows the mean ± SD percentage of defective sperm in WT (n = 3) and *Spink2*$^{+/-}$ mice (n = 3).

B Anomalies were observed in the head and the mid- and principle pieces in WT and *Spink2*$^{+/-}$ mice (n = 3).

C Total and progressive sperm motility were strongly decreased in *Spink2*$^{+/-}$ heterozygous mice (n = 5) in comparison with WT sperm (n = 5).

Data information: n represents the number of biological replicates, and for each replicate, more than 100 sperm were assessed per condition. Data are presented as mean ± SD. Statistical differences were assessed using t-test, P-values as indicated.

and was also described in newborn mice when *Spink3* (orthologue of *Spink1*) is mutated (Ohmuraya *et al*, 2005). Based on these results, it has been postulated that SPINK1/Spink3 could have the dual function of protease inhibitor and negative regulator of autophagy (Ohmuraya *et al*, 2012). Our results show that the absence of SPINK2 induces a microautophagy-like pathway in germ cells thereby further supporting this hypothesis.

Oligozoospermia and azoospermia is a continuum correlated with *SPINK2* haploinsufficiency

We observed that in man, the presence of a homozygous *SPINK2* mutation leads to azoospermia while a heterozygous mutation can induce oligozoospermia suggesting that *SPINK2* haploinsufficiency can result in oligozoospermia. In mice, we showed that the complete absence of the protein leads to azoospermia. We also showed that heterozygous animals have terato-astheno-zoospermia but with no obvious decrease in sperm number and no impact on fertility. A previous study carried out in a different mouse hypomorphic mutant line showed that a significant inactivation of *Spink2* (likely in excess of 90%) led to a reduction by half of sperm number within the epididymis and a five-fold increase in morphologically abnormal spermatozoa. Male mice also exhibited a reduced fertility and produced litters of reduced size with an average of 5.19 pups by litter compared to 8.56 in controls (Lee *et al*, 2011). These results and the results presented therein thus confirm that in mice, the severity of the phenotype is dependent on *Spink2* expression levels and that there is a phenotypic continuum ranging from (i) azoospermia in the complete absence of the protein (ii) to teratozoospermia and oligozoospermia associated with subfertility when only a fraction of the protein is present and finally (iii) to astheno-teratozoospermia with no impact on fertility when half of the protein is present. These observations in mice strongly support the notion that SPINK2 heterozygous mutations in man will impact spermatogenesis with a variable effect on fertility. We identified only one heterozygous mutation out of 611 analyzed patients indicating that *SPINK2* variants are very rare, likely because heterozygous variants underwent a strong negative selection during evolution. This hypothesis is supported by data from the ExAC database which indicate that the *SPINK2* gene has a high probability of loss of function intolerance (pLI = 0.72).

The testis is the organ which expresses the highest number of tissue-specific transcripts ($n > 500$) (Feig *et al*, 2007; Dezso *et al*, 2008) and altered spermatogenesis has been observed in knockout mouse models for more than 388 genes (Massart *et al*, 2012). It is therefore expected that NOA is genetically highly heterogeneous and that few patients carry causal defects on the same gene. Due to the involvement of the corresponding proteins in multiple phases of spermatogenesis, the causes of azoospermia are numerous and involve genes controlling spermatogonial self-renewal, meiosis, and spermiogenesis. Here, we have confirmed that alterations of spermiogenesis do not only lead to teratozoospermia as described several times previously (Dieterich *et al*, 2007; Harbuz *et al*, 2011; Ben Khelifa *et al*, 2014) but also to azoospermia. The vast majority of patients with an altered spermatogenesis can be treated with IVF or by the direct injection of a sperm into the oocyte (ICSI). Most patients with NOA however cannot benefit from ICSI-IVF treatments. Identifying the genetic defects responsible for NOA and characterizing their molecular pathogeny will provide a basis for the development of therapeutic solutions tailored to the patient. In this particular case, we have shown that SPINK2 deficiency can induce azoospermia and demonstrated that unrestricted acrosomal protease activity induces the arrest of spermiogenesis. Moreover, we provided evidence that this process activates a microautophagy-like pathway. As we have shown that the pool of undifferentiated spermatogonia is not affected, we can envisage a method of treatment targeting protease activity using a protease inhibitor, as is done for chronic pancreatitis caused by SPINK1 deficiency (Kambhampati *et al*, 2014).

Materials and Methods

Patients and biological samples

Human sperm were obtained from patients consulting for diagnosis or assisted reproductive techniques at the fertility center of the Grenoble University Hospital. All patients signed an informed consent for use of part of their samples in research programs respecting the WMA declaration of Helsinki. The samples were then stored in the CRB Germetheque (certification under ISO-9001 and NF-S 96-900) following a standardized procedure. Consent for CRB storage was approved by the CPP Sud-Ouest of Toulouse (coordination of the multisite CRB Germetheque). The storage and transfer authorization number for the CRB Germetheque is AC2009-886. The scientific and ethical board of the CRB Germetheque approved the transfer of the semen samples for this study. Additional DNA samples from patients with azoospermia and oligozoospermia were obtained from the CHU of Grenoble, Saint Etienne, and Marseille. All patients gave their informed consent for the anonymous use of their leftover samples. Brothers Br1 and Br2 are French citizens from a traveling group originating from Romania but whose recent ancestors lived in Spain and the south of France. Subject P105 is also a French citizen with eastern ascendants (from Russia).

Exome sequencing and bioinformatic analysis

Genomic DNA was isolated from saliva using Oragene saliva DNA collection kit (DNA Genotek Inc., Ottawa, Canada). Exome capture was performed using NimbleGen SeqCap EZ Kit version 2 (Roche). Sequencing was conducted on an Illumina HiSeq 2000 instrument with paired-end 76-nt reads. Sequence reads were aligned to the reference genome (hg19) using MAGIC (SEQC/MAQC-III Consortium, 2014). Duplicate reads and reads that mapped to multiple locations in the exome were excluded from further analysis. Positions with sequence coverage below 10 on either the forward or reverse strand were excluded. Single nucleotide variations (SNV) and small insertions/deletions (indels) were identified and quality-filtered using in-house scripts. The most promising candidate variants were identified using an in-house bioinformatics pipeline. Variants with a minor allele frequency > 5% in the NHLBI ESP6500 or in 1000 Genomes Project phase 1 datasets, or > 1% in ExAC, were discarded. We also compared these variants to an in-house database of 56 control exomes. All variants present in a homozygous state in this database were excluded. We used Variant Effect Predictor (VEP) to predict the impact of the selected variants. We only retained variants impacting splice donor/acceptor sites or

causing frameshift, in-frame insertions/deletions, stop gain, stop loss, or missense variants except those scored as "tolerated" by SIFT (sift.jcvi.org) and as "benign" by PolyPhen-2 (genetics.bwh.harvard.edu/pph2). All steps from sequence mapping to variant selection were performed using the ExSQLibur pipeline (https://github.com/tkaraouzene/ExSQLibur). Our datasets were obtained from subjects who have consented to the use of their individual genetic data for biomedical research, but not for unlimited public data release. Therefore, we submitted it to the European Genome-phenome Archive, through which researchers can apply for access of the raw data under the accession number EGAD00001003326.

Sanger sequencing

Sanger sequencing of the four *SPINK2* exons and intron borders was carried out using the primers described in Appendix Table S3. Thirty-five cycles of PCR amplification were carried out with a hybridization temperature of 60°C. Sequencing reactions were performed using BigDye Terminator v3.1 (Applied Biosystems). Sequence analyses were carried out on ABI 3130XL (Applied Biosystems). Sequences were analyzed using seqscape software (Applied Biosystems).

RT–PCR and quantitative real-time PCR

Total RNA from various tissues including testes from three WT and homozygous KO mice was extracted using the mirVana™ PARIS™ Kit (Life Technologies®) according to the manufacturer's protocol. Human cDNAs were obtained from Life Technologies® mRNA.

Reverse transcription was carried out with 5 µl of extracted RNA (~500 ng). Hybridization of the oligo dT was performed by incubating for 5 min at 65°C and quenching on ice with the following mix: 5 µl RNA, 3 µl of poly-T-oligo primers (dT) 12–18 (10 mM; Pharmacia), 3 µl of the four dNTPs (0.5 mM, Roche Diagnostics) and 2.2 µl of H_2O. Reverse transcription then was carried out for 30 min at 55°C after the addition of 4 µl of 5× buffer, 0.5 µl RNase inhibitor, and 0.5 µl Transcriptor reverse transcriptase (Roche Diagnostics). One microliter of the obtained cDNA mix was used for the subsequent PCR. Primers are described in Appendix Table S4.

A specific region of the transcript was amplified using a StepOnePlus™ Real-Time PCR System (Life Technologies®) with Power SYBR® Green PCR Master Mix (Life Technologies®) according to the manufacturer's protocol. PCR without template was used as a negative control to verify experimental results. The sequence for oligonucleotide primers used and their product sizes are summarized in Appendix Table S5.

After amplification, the specificity of the PCR was determined by both melt-curve analysis and gel electrophoresis to verify that only a single product of the correct size was present. Quantification of the fold change in gene expression was determined by the relative quantification method ($2^{-\Delta\Delta C_T}$) using the beta-actin gene as a reference. Data are shown as the average fold increase ± standard error of the mean.

Primary antibodies

SPINK2 rabbit polyclonal antibody was from Sigma-Aldrich (HPA026813) and used at 1/1,000 for Western blot analysis. Sperm protein Sp56 and Golgi matrix protein GM130 (610822) mouse monoclonal antibodies were from QED Bioscience Inc. (used at 1/800 and 1/200, respectively). Promyelocytic leukemia zinc finger protein PLZF rabbit polyclonal antibody (Sc-22839) was from Santa Cruz Biotechnology. Dpy19l2 antibodies were produced in rabbit as polyclonal antibodies raised against RSKLREGSSDRPQSSC and CTGQARRRWSAATMEP peptides corresponding to amino acids 6–21 and 21–36 of the N-terminus of mouse Dpy19l2 (Pierre *et al*, 2012). DDK antibody was from OriGene (TA50011) or Sigma-Aldrich (FLAG® M2F1804) and used at 1/10,000 for Western blot analysis. Acrosin antibody was previously described (Gallo *et al*, 1991) and is a gift from Denise Escalier.

Western blot analysis

HEK293 cells were lysed in 25 mM Tris pH 7.4, 5 mM EDTA, 1% Triton X-100, and complete protease inhibitor cocktail (Roche) and were then centrifuged. After centrifugation at 20,000 *g* for 15 min at 4°C, the soluble supernatant was conserved and subjected to SDS–PAGE. The protein concentration from supernatants was quantified by the bicinchoninic acid assay (BCA assay) using bovine serum albumin as a standard. Sample concentrations were adjusted and mixed with 1× high-SDS sample buffer (4% SDS, 62 mM Tris–HCl pH 6.8, 0.1% bromophenol blue, 15% glycerol, 5% ß-mercaptoethanol) and separated using 4–20% SDS mini-PROTEAN® TGX Stain-Free™ Precast Gels (Bio-Rad) or 10 and 20% polyacrylamide–SDS gels and transferred into PVDF membranes (Millipore, 0.2 µm) using Trans-Blot® Turbo™ Blotting System and Midi Transfer Packs (Bio-Rad). The membranes were blocked in 5% non-fat dry milk in PBS/0.1% Tween and incubated for 1 h at room temperature with the primary antibody, followed by 45-min incubation with a species-matched horseradish peroxidase-labeled secondary antibody (1/10,000) (Jackson ImmunoResearch). Immunoreactivity was detected using chemiluminescence detection kit reagents (Luminata; Millipore) and a ChemiDoc Station (Bio-Rad).

Real-time cell analysis

The growth, proliferation, and adhesion kinetics of HEK293 cells were determined using RTCA technology (ACEA Biosciences, San Diego, CA, USA). Fifty microliters of DMEM supplemented with 10% HI-FBS and 50 µg/ml gentamicin (cell culture medium) was loaded in each well of the E-plate 96 (gold-microelectrode array integrated E-plate; ACEA Biosciences). E-plate 96 was then connected to the system to obtain background impedance readings. Around 1.5×10^4 cells in 50 µl were added to the wells containing 50 µl of culture medium. The E-plates were placed on the RTCA SP Station located in a 37°C, 5% CO_2 tissue culture incubator for continuous impedance recording. The cell index values measured by continuous impedance recordings every 5 min are proportional to the number of adherent cells. After 16–17 h, cells were transfected as described below, and for each of the conditions, four replicates were done. The assay was conducted for 40 h.

Mice

All animal procedures were run according to the French guidelines on the use of animals in scientific investigations with the approval of

the local ethical committee (Grenoble-Institut des Neurosciences—ethical committee, study agreement number 004). Mice were euthanized by cervical dislocation.

The Spink2tm1.1 (KOMP)Vlcg mouse strain used for this research project was created from ES cell clone Spink2_AG5_M7, generated by Regeneron Pharmaceuticals, Inc. and made into live mice by the KOMP Repository (www.komp.org) and the Mouse Biology Program (www.mousebiology.org) at the University of California Davis. The methods used to create the VelociGene targeted alleles have been published (Valenzuela *et al*, 2003). They were then reared by the Mouse Clinical Institute—MCI—located in Strasbourg as part of the "knockout mouse project". The colony used in this study was initiated from two couples consisting of heterozygous females and males. Mice were housed with unlimited access to food and water and were sacrificed after 8 weeks of age (the age of sexual maturity).

Genotyping

DNA for genotyping was isolated from tail biopsies. Tail biopsies (2 mm in length) were digested in 200 µl of DirectPCR Lysis Reagent (Tail) (Viagen Biotech Inc, CA, USA) and 0.2 mg of proteinase K for 12–15 h at 55°C followed by 1 h at 85°C for proteinase K inactivation. The DNA was directly used for PCRs. Multiplex PCR was done for 35 cycles, with an annealing temperature of 58°C, and an elongation time of 60 s at 72°C. PCR products were separated by 2% agarose gel electrophoresis. Genotypes were determined according to the migration pattern. Primers are described in Appendix Table S6.

Phenotypic analysis of mutant mice

To test fertility, pubescent *Spink2*$^{-/-}$ males (8-week-old) were mated with WT females.

To determine sperm concentration, sperm samples were collected from the cauda epididymis and vas deferens of 8-week-old males, and sperm number was determined using a hemocytometer under a light microscope.

Sperm motility analysis

Experiments were performed on a CASA CEROS v.12 (Hamilton Thorne Biosciences, Beverly, MA, USA) using Leja double-chamber slides (Leja Products B.V., the Netherlands) for standard count with 100 µm depth. After epididymal extraction, sperm cells were allowed to swim for 10 min at 37°C and then were immediately analyzed. At least 150 cells were analyzed per sample with the following parameters: acquisition rate: 60 Hz; number of frames: 45; minimum contrast: 50; minimum cell size: 5; low static-size gate: 0.3; high static-size gate: 1.95; low static-intensity gate: 0.5; high static-intensity gate: 1.3; minimum elongation gate: 0; maximum elongation gate: 87; and magnification factor: 0.7. The motility parameters measured were curvilinear velocity (VCL), straight-line velocity (VSL), average path velocity (VAP), and amplitude of lateral head displacement (ALH). Motile sperm were defined by VAP > 1 and progressive sperm were defined by VAP > 30 and VSL/VAP > 0.7.

Histological analysis

To analyze testicular integrity, testes from adult *Spink2*$^{+/+}$ and *Spink2*$^{-/-}$ mice were fixed by immersion in 4% paraformaldehyde (PFA) for 14 h, embedded in paraffin, and sectioned (4 µm). For histological analysis, after being deparaffinized slides were stained with hematoxylin and eosin or by the PAS technique. The colored sections were digitized at ×40 magnification through an Axioscope microscope (Zeiss, Germany) equipped with a motorized X–Y-sensitive stage. For sperm morphology analysis, sperm were washed twice in PBS and then displayed over slides, dried at room temperature, and then fixed in 75% ethanol for Harris–Shorr staining. At least 100 sperm cells were analyzed per sample.

Testicular germ cell dissociation

C57BL/6 male or *Spink2* KO mice (8-week-old) were euthanized by cervical dislocation. The testes were surgically removed and placed in PBS (at room temperature). The tunica albuginea was removed from the testes with sterile forceps and discarded. Then, the testes were incubated in 1 mg/ml of collagenase solution in EKRB cell buffer containing (in mM) 2 $CaCl_2$, 12.1 glucose, 10 HEPES, 5 KCl, 1 $MgCl_2$, 6 Na-lactate, 150 NaCl, 1 NaH_2PO_4, and 12 $NaHCO_3$ pH 7, and agitated horizontally at a maximum of 120 rpm for 30 min at 25°C. The dispersed seminiferous tubules were then washed with PBS and cut thinly. Cells were dissociated by gentle pipetting filtered through a 100-µm filter and then pelleted by centrifugation at 500 *g* for 7 min. Cells were resuspended in 1 ml PBS, fixed with 4% PFA solution, washed with PBS, and finally layered onto polylysine-coated slides.

Immunohistochemistry

Mice were anesthetized by intraperitoneal injection of a ketamine/xylazine cocktail (87.5 mg/kg ketamine and 12.5 mg/kg xylazine) and sacrificed through intracardiac perfusion of PFA (4%). The testes and epididymides were removed and fixed for a further 8 h before paraffin embedding and sectioning. Mature sperm cells were obtained for analysis through mechanical dilaceration of the epididymis. Sperm cells were fixed in 4% PFA for 1 min and washed in PBS before being spotted onto poly-L-lysine-pre-coated slides. Spermatogenic cells of the round-spermatid stage were purified by unit gravity sedimentation from a spermatogenic cell suspension obtained from sexually mature males as described in Yassine *et al* (2015).

For immunofluorescence experiments, heat-induced antigen retrieval was performed by boiling slides immersed in either 0.01 M sodium citrate buffer–0.05% Tween-20, pH 6.0, or 10 mM Tris base–1 mM EDTA solution–0.05% Tween-20, pH 9.0, for 15–25 min. Sections were blocked in 2% goat serum–0.1% Triton X-100 for 1 h at RT and incubated with primary antibodies overnight at 4°C. The slides were then washed and incubated with secondary antibody (DyLight 549-conjugated goat anti-mouse IgG or DyLight 488-conjugated goat anti-rabbit IgG, Jackson ImmunoResearch) and Hoechst 33342 for 2 h at RT, rinsed, and mounted with Dako mounting medium (Life Technology). Images were taken by confocal microscopy (Zeiss LSM 710) and processed using Zen 2009 software.

Electron microscopy (EM)

Adult male mice were anesthetized and fixed by intracardiac injection with 2% glutaraldehyde and 2.5% PFA in 0.1 M cacodylate, pH 7.2. For morphological analysis, samples were fixed with 2.5% glutaraldehyde in 0.1 M cacodylate buffer pH 7.4 over 24 h at room temperature. Samples were then washed with buffer and post-fixed with 1% osmium tetroxide and 0.1 M cacodylate pH 7.2 for 1 h at 4°C. After extensive washing with water, cells were further stained with 1% uranyl acetate pH 4 in water for 1 h at 4°C before being dehydrated through graded alcohol (30%–60%–90%–100%–100%–100%) and infiltrate with a mix of 1/1 epon/alcohol 100% for 1 h and several baths of fresh epon (Flukka) during 3 h. Finally, samples were embedded in a capsule full of resin that was left to polymerize over 72 h at 60°C. Ultrathin sections of the samples were cut with an ultramicrotome (Leica), and the sections were post-stained with 5% uranyl acetate and 0.4% lead citrate before being observed with an electron microscope at 80 kV (JEOL 1200EX). Images were acquired with a digital camera (Veleta; SIS, Olympus), and morphometric analysis was performed with iTEM software (Olympus).

Cell culture and transfection

Mycoplasma-free HEK293 cells were a gift from A. Andrieux from Grenoble Neuroscience Institute and grown in Dulbecco's modified Eagle's medium supplemented with 10% FBS (Invitrogen, France) and 50 μg/ml gentamicin (Sigma) in a 37°C, 5% CO_2 cell culture incubator and transiently transfected with Cter-DDK-tagged human acrosin (RC214256; OriGene, Rockville, MD, USA) and/or human *SPINK2* (RC205388; OriGene) and/or human c.1A>T mutated *SPINK2*-containing pCMV6 plasmids, using JetPRIME Transfection Reagent (Polyplus, France) according to the manufacturer's instructions. For immunochemistry experiments, transfected cells were fixed with 4% PFA 2 days after transfection.

DNA strand breaks

Sections were permeabilized using a 0.1% (v/v) Triton X-100 and 0.1% (w/v) sodium citrate in 1× PBS for 2 min and labeled by terminal deoxynucleotidyl transferase-mediated deoxy-UTP nick-end labeling (TUNEL) according to the Roche protocol of the *In Situ* Cell Detection Kit (Roche Diagnostics, Mannheim, Germany). Nuclei were counterstained in a 0.5 μg/ml Hoechst solution for 3 min, washed in PBS for 3 min, and mounted with DAKO mounting medium.

Statistical analyses

n represents the number of biological replicates. For sperm analyses, for each replicate, more than 100 sperm were assessed per condition. Statistical analyses were performed with SigmaPlot 10 and GraphPad Prism 7. *t*-Tests were used to compare WT and KO samples. Data represent mean ± SEM or SD, as indicated. Statistical tests with a two-tailed *P*-value ≤ 0.05 were considered significant.

Expanded View for this article is available online.

The paper explained

Problem
Infertility concerns one in seven couples and is usually addressed by performing *in vitro* fertilization (IVF) often by injecting spermatozoa directly into the oocytes by intracytoplasmic sperm injection (ICSI). Some men have a non-obstructive azoospermia (NOA), caused by a deficient spermatogenesis, and have no spermatozoa in the ejaculate. In some cases, a testicular biopsy can be performed in hope of finding some mature spermatozoa that will be used for ICSI, but most men with NOA will not be able to have biological children. It is believed that most cases of NOA are caused by a genetic factor, but a diagnosis is obtained for only approximately 20% of patients.

Results
We performed exome sequencing on two brothers with NOA and identified a homozygous mutation in the *SPINK2* gene coding for a serine protease inhibitor believed to target the acrosin, the main protease of the acrosome, a large vesicle located to the anterior part of the spermatozoa and containing an enzyme mix necessary to perforate the zona pellucida of the oocyte to achieve fertilization. Mouse study allowed to observe that homozygous KO male also had NOA, confirming the human diagnostic. Germ cells could go through meiosis but were blocked at the round-spermatid stage. We further observed that in the round spermatids, in the absence of SPINK2, the acrosin could autoactivate during its transit through the endoplasmic reticulum and the Golgi apparatus leading to a disorganization of the Golgi and its inability to form the acrosome and a block at the round-spermatid stage. We further demonstrate that the presence of a heterozygous SPINK2 mutation was also deleterious leading to the production of sperm with variable levels of anomalies.

Impact
We identified a new gene leading to male infertility permitting to improve the diagnostic efficiency for NOA patients. We demonstrate that whole-exome sequencing is an efficient technique to identify new infertility genes and to realize a diagnostic for affected men. We showed that the control of proteases by antiproteases, and in particular by SPINK2, is critical during spermiogenesis and demonstrate that the SPINK gene family is involved not only in pancreatitis or skin disease but also in male infertility.

Acknowledgements
We thank the GIN electron microscopy platform and Anne Bertrand, and the IAB microscopy platform and Alexei Grichine and Jacques Mazzega for their technical help. We thank Myriam Dridi for her work on HEK cells and antibody validation, Jean Pascal Hograindleur for his help for CASA experiments and Denise Escalier for her generous gift of human anti-acrosin antibody. This work was mainly supported by the French research agency (ANR) within the 2009 Genopat program for the ICG2I project "Identification and characterization of genetic causes of male infertility" to PR and CA. Support was also obtained from the Fondation Maladies Rares (FMR) for the project R16070CC, "Identification of genetic causes of human NOA".

Author contributions
PFR and CA designed the study, supervised all laboratory work, and wrote the manuscript. They have full access to all of the data in the study and take responsibility for the integrity of the data and its accuracy. All authors read, corrected, and made a significant contribution to the manuscript. Z-EK, TK, AA-Y, CB, MG, NT-M, and CC produced and analyzed the genetic data, and Z-EK, MC-K, AA-Y, and ASV performed immunohistochemistry (IF) experiments. SPB, JE and EL performed Western blot experiments and real-time cell

analyses. Z-EK and GM performed sperm analyses; KP-G and ASV performed the electron microscopy; and BD, IA-S, MM, CM-G, SN, VS, MB, FB, JF, and SH provided clinical samples and data and supplied biological materials.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Arboleda CE, Gerton GL (1987) Studies of three major proteases associated with guinea pig sperm acrosomes. *J Exp Zool* 244: 277 – 287

Ayhan O, Balkan M, Guven A, Hazan R, Atar M, Tok A, Tolun A (2014) Truncating mutations in TAF4B and ZMYND15 causing recessive azoospermia. *J Med Genet* 51: 239 – 244

Baba T, Michikawa Y, Kawakura K, Arai Y (1989) Activation of boar proacrosin is effected by processing at both N- and C-terminal portions of the zymogen molecule. *FEBS Lett* 244: 132 – 136

Ben Khelifa M, Coutton C, Zouari R, Karaouzene 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

Brown CR, Harrison RA (1978) The activation of proacrosin in spermatozoa from ram bull and boar. *Biochim Biophys Acta* 526: 202 – 217

Cesari A, Sanchez JJ, Biancotti JC, Vazquez-Levin MH, Kaiser G, Palma GA, Alberio R, Vincenti AE, Fornes MW (2004) Immunolocalization of bovine sperm protease BSp120 by light and electron microscopy during capacitation and the acrosome reaction: its role in *in vitro* fertilization. *Mol Reprod Dev* 69: 411 – 418

Chen JM, Mercier B, Audrezet MP, Ferec C (2000) Mutational analysis of the human pancreatic secretory trypsin inhibitor (PSTI) gene in hereditary and sporadic chronic pancreatitis. *J Med Genet* 37: 67 – 69

Chera S, Buzgariu W, Ghila L, Galliot B (2009) Autophagy in Hydra: a response to starvation and stress in early animal evolution. *Biochim Biophys Acta* 1793: 1432 – 1443

Dezso Z, Nikolsky Y, Sviridov E, Shi W, Serebriyskaya T, Dosymbekov D, Bugrim A, Rakhmatulin E, Brennan RJ, Guryanov A *et al* (2008) A comprehensive functional analysis of tissue specificity of human gene expression. *BMC Biol* 6: 49

Dieterich K, Soto RR, Faure AK, Hennebicq S, Ben Amar B, Zahi M, Perrin J, Martinez D, Sele B, Jouk PS *et al* (2007) Homozygous mutation of AURKC yields large-headed polyploid spermatozoa and causes male infertility. *Nat Genet* 39: 661 – 665

Feig C, Kirchhoff C, Ivell R, Naether O, Schulze W, Spiess AN (2007) A new paradigm for profiling testicular gene expression during normal and disturbed human spermatogenesis. *Mol Hum Reprod* 13: 33 – 43

Furio L, Hovnanian A (2014) Netherton syndrome: defective kallikrein inhibition in the skin leads to skin inflammation and allergy. *Biol Chem* 395: 945 – 958

Gallo JM, Escalier D, Grellier P, Precigout E, Albert M, David G, Schrevel J (1991) Characterization of a monoclonal antibody to human proacrosin and its use in acrosomal status evaluation. *J Histochem Cytochem* 39: 273 – 282

Harbuz R, Zouari R, Pierre V, Ben Khelifa M, Kharouf M, Coutton C, Merdassi G, Abada F, Escoffier J, Nikas Y *et al* (2011) A recurrent deletion of DPY19L2 causes infertility in man by blocking sperm head elongation and acrosome formation. *Am J Hum Genet* 88: 351 – 361

Huang-Yang YH, Meizel S (1975) Purification of rabbit testis proacrosin and studies of its active form. *Biol Reprod* 12: 232 – 238

Kambhampati S, Park W, Habtezion A (2014) Pharmacologic therapy for acute pancreatitis. *World J Gastroenterol* 20: 16868 – 16880

Kennedy WP, Swift AM, Parrish RF, Polakoski KL (1982) Proacrosin conversion inhibitor. Purification and initial characterization of a boar sperm protein which prevents the conversion of proacrosin into acrosin. *J Biol Chem* 257: 3095 – 3099

Kierszenbaum AL, Rivkin E, Tres LL (2003) Acroplaxome, an F-actin-keratin-containing plate, anchors the acrosome to the nucleus during shaping of the spermatid head. *Mol Biol Cell* 14: 4628 – 4640

Kim KS, Cha MC, Gerton GL (2001) Mouse sperm protein sp56 is a component of the acrosomal matrix. *Biol Reprod* 64: 36 – 43

Kohno N, Yamagata K, Yamada S, Kashiwabara S, Sakai Y, Baba T (1998) Two novel testicular serine proteases, TESP1 and TESP2, are present in the mouse sperm acrosome. *Biochem Biophys Res Commun* 245: 658 – 665

Lee B, Park I, Jin S, Choi H, Kwon JT, Kim J, Jeong J, Cho BN, Eddy EM, Cho C (2011) Impaired spermatogenesis and fertility in mice carrying a mutation in the Spink2 gene expressed predominantly in testes. *J Biol Chem* 286: 29108 – 29117

Li WW, Li J, Bao JK (2012) Microautophagy: lesser-known self-eating. *Cell Mol Life Sci* 69: 1125 – 1136

Liu DY, Baker HW (1993) Inhibition of acrosin activity with a trypsin inhibitor blocks human sperm penetration of the zona pellucida. *Biol Reprod* 48: 340 – 348

Maor-Sagie E, Cinnamon Y, Yaacov B, Shaag A, Goldsmidt H, Zenvirt S, Laufer N, Richler C, Frumkin A (2015) Deleterious mutation in SYCE1 is associated with non-obstructive azoospermia. *J Assist Reprod Genet* 32: 887 – 891

Massart A, Lissens W, Tournaye H, Stouffs K (2012) Genetic causes of spermatogenic failure. *Asian J Androl* 14: 40 – 48

Meizel S, Deamer DW (1978) The pH of the hamster sperm acrosome. *J Histochem Cytochem* 26: 98 – 105

O'Brien DA, Welch JE, Goulding EH, Taylor AA Jr, Baba T, Hecht NB, Eddy EM (1996) Boar proacrosin expressed in spermatids of transgenic mice does not reach the acrosome and disrupts spermatogenesis. *Mol Reprod Dev* 43: 236 – 247

Ohmuraya M, Hirota M, Araki M, Mizushima N, Matsui M, Mizumoto T, Haruna K, Kume S, Takeya M, Ogawa M *et al* (2005) Autophagic cell death of pancreatic acinar cells in serine protease inhibitor Kazal type 3-deficient mice. *Gastroenterology* 129: 696 – 705

Ohmuraya M, Sugano A, Hirota M, Takaoka Y, Yamamura K (2012) Role of intrapancreatic SPINK1/Spink3 expression in the development of pancreatitis. *Front Physiol* 3: 126

Okutman O, Muller J, Baert Y, Serdarogullari M, Gultomruk M, Piton A, Rombaut C, Benkhalifa M, Teletin M, Skory V *et al* (2015) Exome sequencing reveals a nonsense mutation in TEX15 causing spermatogenic failure in a Turkish family. *Hum Mol Genet* 24: 5581 – 5588

Pierre V, Martinez G, Coutton C, Delaroche J, Yassine S, Novella C, Pernet-Gallay K, Hennebicq S, Ray PF, Arnoult C (2012) Absence of Dpy19l2, a new inner nuclear membrane protein, causes globozoospermia in mice by preventing the anchoring of the acrosome to the nucleus. *Development* 139: 2955 – 2965

Rawlings ND, Tolle DP, Barrett AJ (2004) Evolutionary families of peptidase inhibitors. *Biochem J* 378: 705 – 716

Redelfs L, Fischer J, Weber C, Wu Z, Meyer-Hoffert U (2016) The serine protease inhibitor of Kazal-type 9 (SPINK9) is expressed in lichen simplex chronicus, actinic keratosis and squamous cell carcinoma. *Arch Dermatol Res* 308: 133 – 137

Rivinoja A, Pujol FM, Hassinen A, Kellokumpu S (2012) Golgi pH, its regulation and roles in human disease. *Ann Med* 44: 542 – 554

SEQC/MAQC-III Consortium (2014) A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat Biotechnol* 32: 903 – 914

Taylor SC, Posch A (2014) The design of a quantitative western blot experiment. *Biomed Res Int* 2014: 361590

Tuttelmann F, Werny F, Cooper TG, Kliesch S, Simoni M, Nieschlag E (2011) Clinical experience with azoospermia: aetiology and chances for spermatozoa detection upon biopsy. *Int J Androl* 34: 291 – 298

Valenzuela DM, Murphy AJ, Frendewey D, Gale NW, Economides AN, Auerbach W, Poueymirou WT, Adams NC, Rojas J, Yasenchak J *et al* (2003) High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. *Nat Biotechnol* 21: 652 – 659

Witt H, Luck W, Hennies HC, Classen M, Kage A, Lass U, Landt O, Becker M (2000) Mutations in the gene encoding the serine protease inhibitor, Kazal type 1 are associated with chronic pancreatitis. *Nat Genet* 25: 213 – 216

Yan W, Si Y, Slaymaker S, Li J, Zheng H, Young DL, Aslanian A, Saunders L, Verdin E, Charo IF (2010) Zmynd15 encodes a histone deacetylase-dependent transcriptional repressor essential for spermiogenesis and male fertility. *J Biol Chem* 285: 31418 – 31426

Yang F, Silber S, Leu NA, Oates RD, Marszalek JD, Skaletsky H, Brown LG, Rozen S, Page DC, Wang PJ (2015) TEX11 is mutated in infertile men with azoospermia and regulates genome-wide recombination rates in mouse. *EMBO Mol Med* 7: 1198 – 1210

Yassine S, Escoffier J, Nahed RA, Pierre V, Karaouzene T, Ray PF, Arnoult C (2015) Dynamics of Sun5 localization during spermatogenesis in wild type and Dpy19l2 knock-out mice indicates that Sun5 is not involved in acrosome attachment to the nuclear envelope. *PLoS One* 10: e0118698

Yatsenko AN, Georgiadis AP, Ropke A, Berman AJ, Jaffe T, Olszewska M, Westernstroer 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

Zahn A, Furlong LI, Biancotti JC, Ghiringhelli PD, Marijn-Briggiler CI, Vazquez-Levin MH (2002) Evaluation of the proacrosin/acrosin system and its mechanism of activation in human sperm extracts. *J Reprod Immunol* 54: 43 – 63

Zhang T, Murphy MW, Gearhart MD, Bardwell VJ, Zarkower D (2014) The mammalian Doublesex homolog DMRT6 coordinates the transition between mitotic and meiotic developmental programs during spermatogenesis. *Development* 141: 3662 – 3671

Expanded View Figures

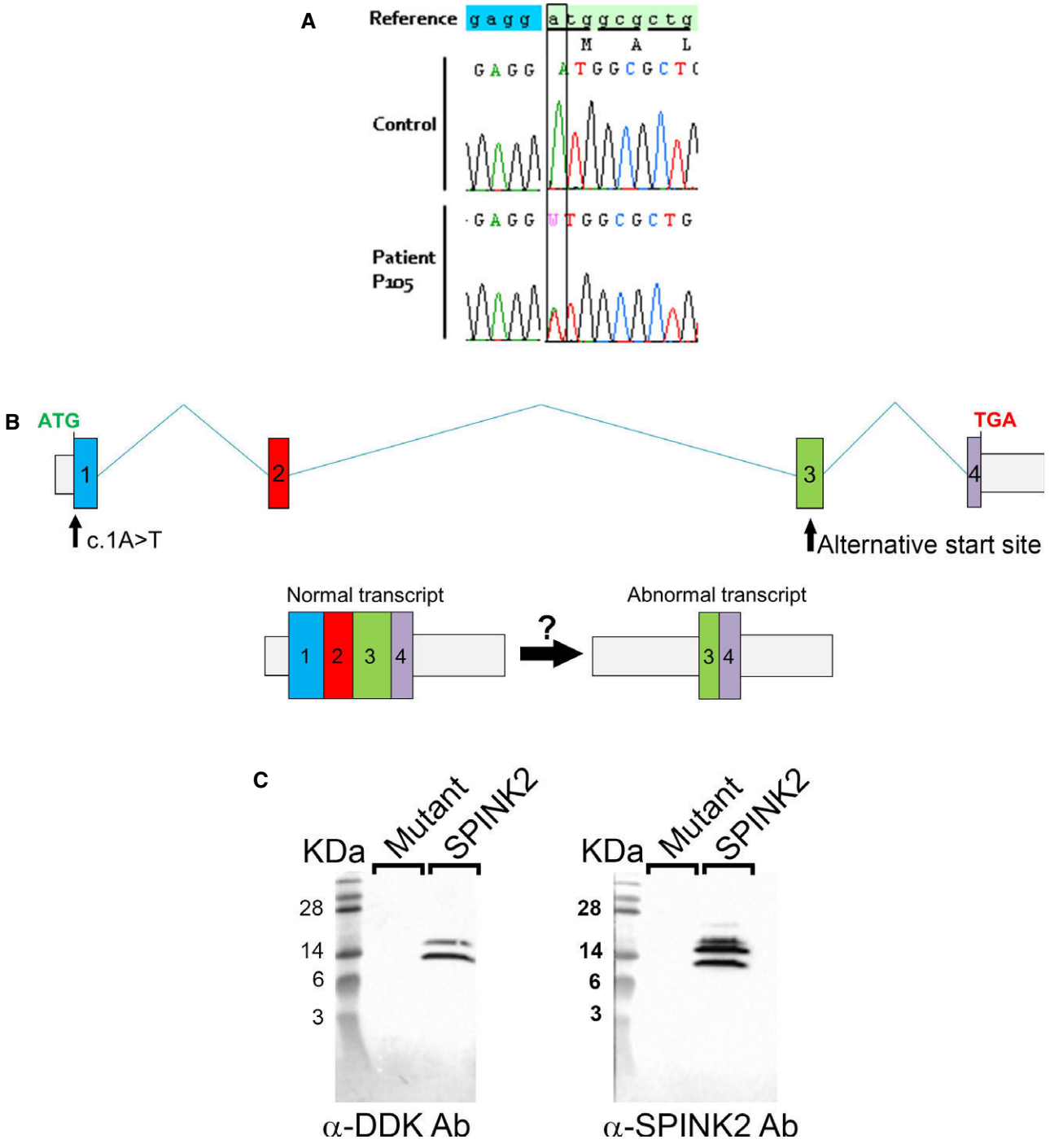


Figure EV1. Genetic analyses of the c.1A>T variant identified in patient P105.

A The c.1A>T variant, heterozygous in patient P105, abrogates the original start codon.

B An alternative start codon is present in the middle of exon 3 and may initiate translation to potentially produce a protein containing the last 27 amino acids of the wild-type protein.

C Western blot of HEK293 cell extracts transfected with C-terminus DDK-tagged SPINK2 or c.1A>T SPINK2 mutant and revealed by anti-DDK or anti-SPINK2 antibodies shows the absence of the putative truncated SPINK2 form.

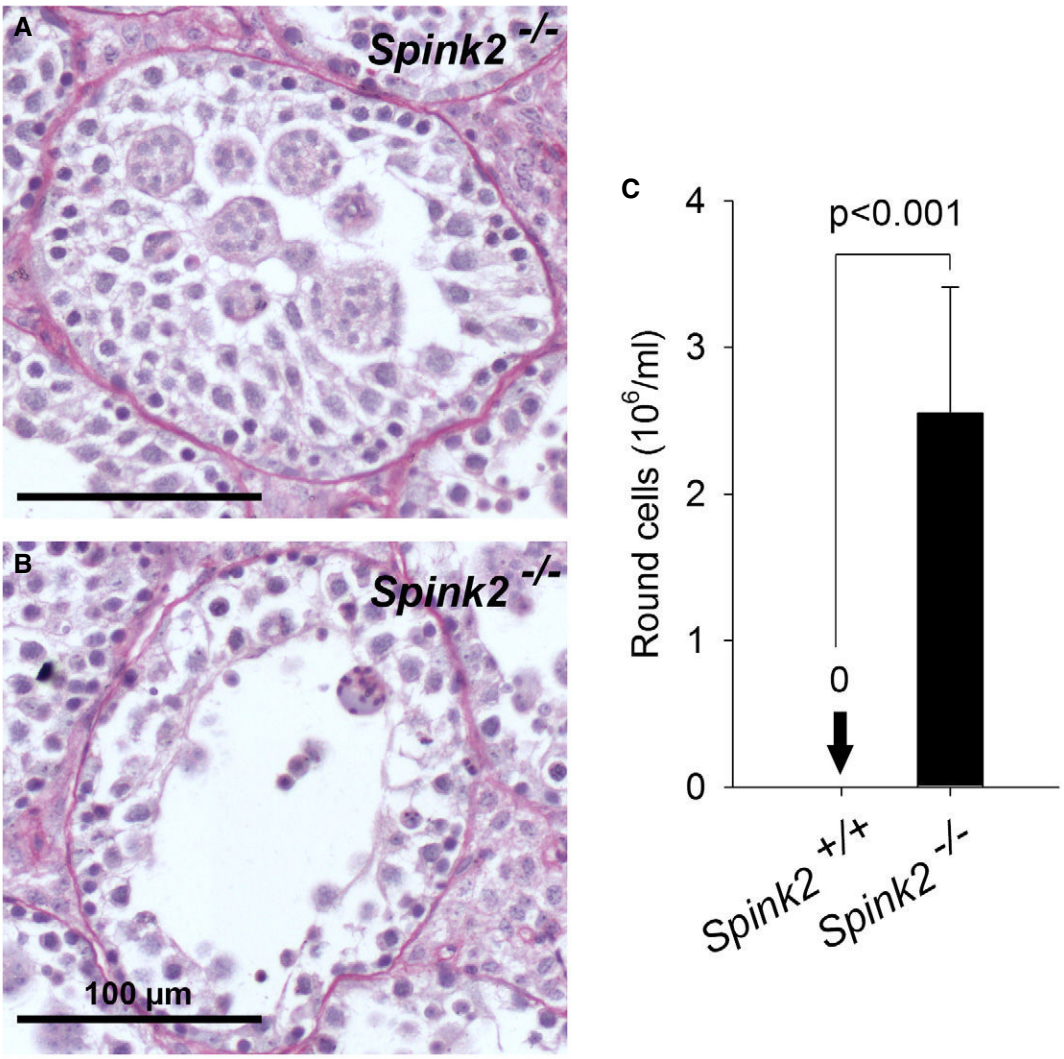


Figure EV2. Lack of Spink2 leads to germ cell desquamation and the presence of round cells in the cauda epididymis.

A, B Staining with hematoxylin and eosin of seminiferous tubule sections from *Spink2*$^{-/-}$ mice showed the presence of numerous round cells or symblasts within the lumen of tubules. Note that photograph (A) corresponds to a close-up of part of Fig 3C1 as it nicely illustrates the presence of symblasts in the tubule. Scale bar, 100 µm.

C Spermatocytograms of cauda epididymis extracts revealed the presence of only round cells at concentrations above 2 million/ml, whereas such cells were absent in WT extracts. Data are presented as mean ± SEM (n = 5).

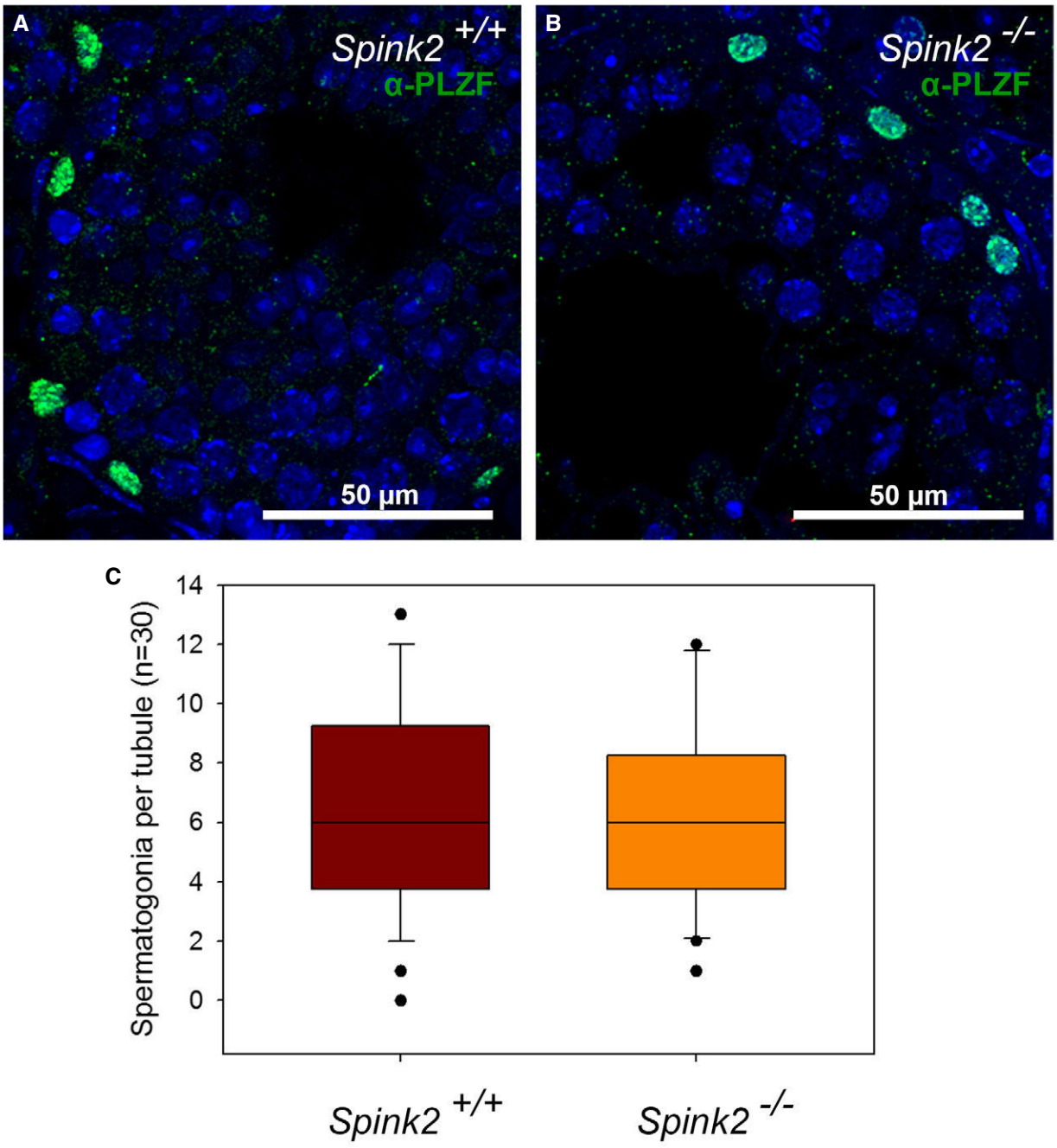


Figure EV3. Pool of non-differentiated spermatogonia is conserved in seminiferous tubules of *Spink2*$^{-/-}$ mice.

A, B Identification of spermatogonia by IF experiments using an anti-PLZF antibody in WT and *Spink2*$^{-/-}$ seminiferous tubule sections, respectively. Scale bars as indicated.

C Box plot showing the number of spermatogonia per seminiferous tubule section (n = 30 for each). No difference in the number of spermatogonia per seminiferous tubule section was observed. The line within the box marks the median, whiskers (error bars) above and below the box indicate the 90th and 10th percentiles, and black dots represent outliers.

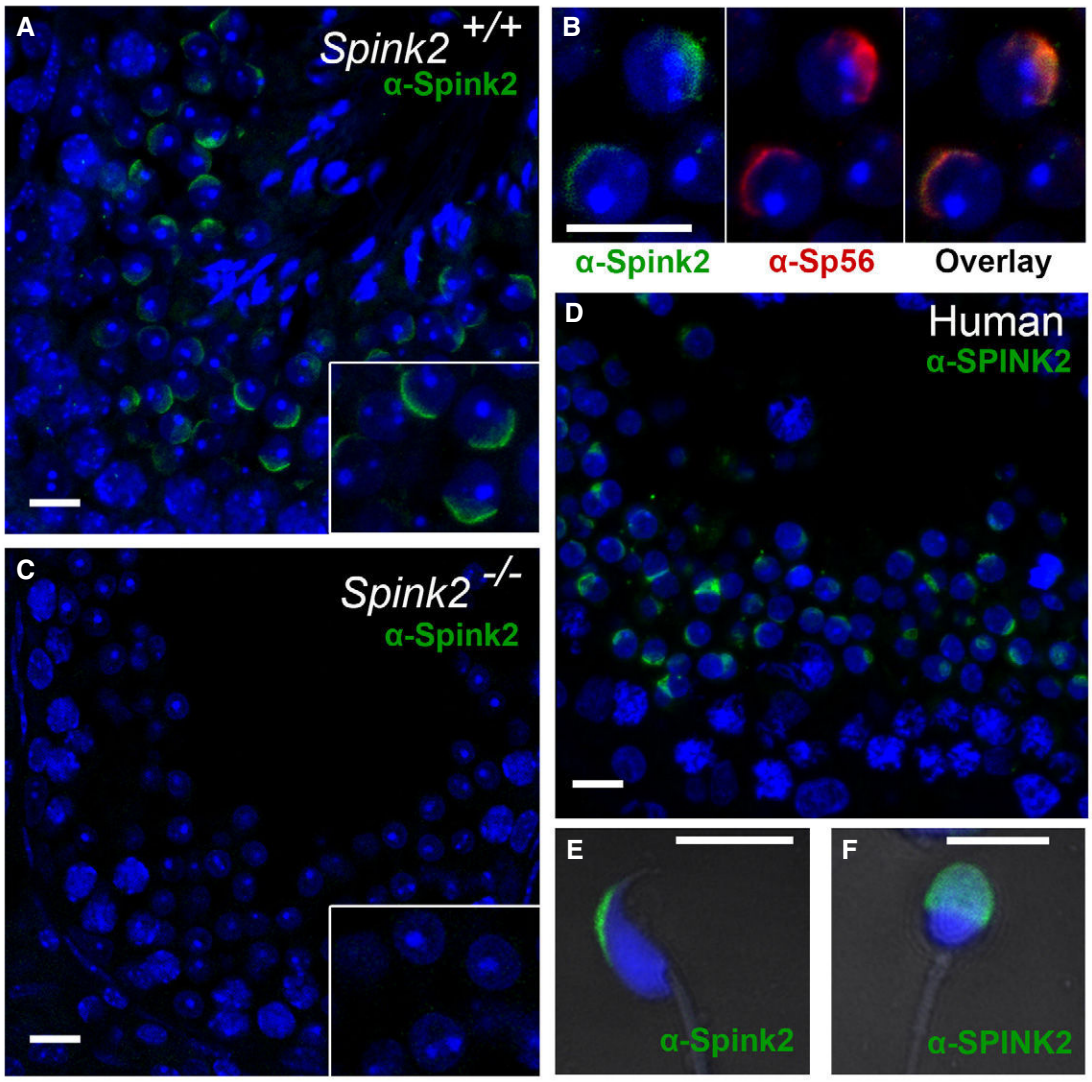


Figure EV4. SPINK2 is located in the acrosome of round spermatids and mature spermatozoa in mouse and human.

A Immunofluorescence experiments using an anti-SPINK2 antibody (green staining) reveals that Spink2 is located in the acrosome of WT mouse round spermatids. Scale bar, 10 μm.

B A double staining using an anti-SPINK2 antibody (green staining) and an anti-Sp56 antibody (red staining) shows that the two signals are colocalized. Scale bar, 10 μm.

C No Spink2 staining is observed in seminiferous tubule sections from *Spink2*$^{-/-}$ mice, stained with an anti-SPINK2 antibody. Scale bar, 10 μm.

D In sections of human seminiferous tubules, a similar localization of SPINK2 is observed in round spermatids, corresponding to the acrosome. Scale bar, 10 μm.

E IF experiments using an anti-SPINK2 antibody (green staining) reveal that Spink2 is located in the acrosome of WT epididymal mouse sperm. Scale bar, 10 μm.

F SPINK2 localization within the acrosome is conserved in ejaculated human sperm. No staining is observed in the midpiece or principle piece. Scale bar, 5 μm.

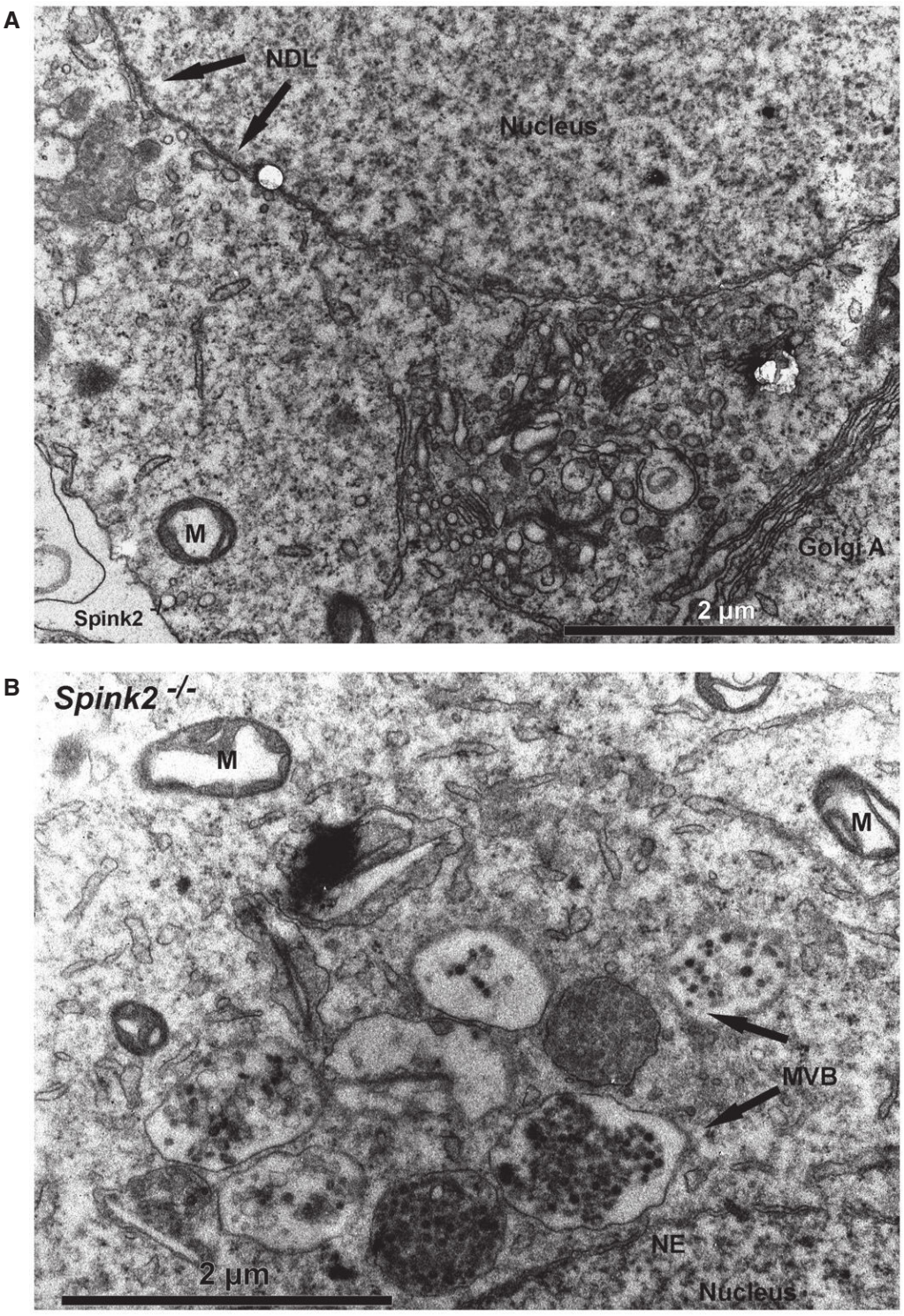


Figure EV5. Disjunction of the polarity of the Golgi apparatus and the nuclear dense lamina (NDL) in *Spink2*$^{-/-}$ round spermatids and presence of multivesicular bodies in the cytoplasm of *Spink2*$^{-/-}$ round spermatids.

A Partial section of a *Spink2*$^{-/-}$ round spermatid observed by EM showing the spatial disjunction of the Golgi apparatus and the NDL. NDL is evidenced by apposition of dense material on the nuclear side of the inner nuclear membrane. Another hallmark of this specialized area of the nuclear envelope is the tight apposition of both inner and outer nuclear membranes along the NDL. M, mitochondria. Scale bar, 2 µm.

B Partial section of a *Spink2*$^{-/-}$ round spermatid observed by EM showing structures corresponding to multivesicular bodies (MVB). M, mitochondria; NE, nuclear envelope; NDL, nuclear dense lamina. Scale bar, 2 µm.

Appendix - Table of content

Appendix Table S1. List of homozygous variants present in B1 and B2.

Appendix Table S2. List of *SPINK2* variants identified in azoo- and oligo- spermic patients.

Appendix Table S3. Sequence of the primer used for *SPINK2* sanger sequencing.

Appendix Table S4. Sequence of the primer used in RT-PCR and size of the amplified products

Appendix Table S5. Sequence of the primer used in real time RT-PCR and size of the amplified products

Appendix Table S6. Sequence of the primer used for genotyping and size of the amplified products.

Appendix Figure S1. Expression of *SPINK2* mRNA in different organs in human and mouse.

Appendix Figure S2. Impact on the protein structure of the c.206-3C>G *SPINK2* variant

Appendix Figure S3. Absence of *Spink2* expression in testicular extracts from KO mice.

Appendix Figure S4. Specificity of antibodies targeting SPINK2.

Appendix Figure S5. Absence of Spink2 does not lead to DNA fragmentation in spermatogenic cells.

Appendix Figure S6. Equal protein loading was verified by stain-free gel technology and Western blots against house keeping protein.

Appendix Table S1. List of homozygous variants present in B1 and B2.
Coordinates of all variations are based on the UCSC GRCh37/hg19 assembly. [&] No RefSeq transcript accession number currently available. Nucleotidic references

Gene	Variant coordinates	Transcript	cDNA Variation	Amino acid variation	Prediction	Testis expression
SPINK2	Chr4:57686748G>C	NM_001271718.1 NM_001271720.1 NM_001271722.1 NM_021114.3	Splice acceptor site	-	Damaging	High and predominant
GUF1	Chr4:44683156G>T	NM_021927.2	c.637G>T	p.Ser148Ile	Damaging	Low
ZCCHC5	ChrX:77913158A>G	NM_152694.2	c.1056A>G	p.Phe254Leu	Damaging	Low
ARR3	ChrX:69498482C>G	NM_004312.2 ENST00000374495[&]	c. 947C>G c.994C>G	p.Ser299Cys	Damaging	Low
FTHL17	ChrX:31089629TC>AA	NM_031894.2	c.542C>A	p.Glu148Leu	Damaging	low

Appendix Table S2. List of *SPINK2* variants identified in azoo- and oligo- spermic patients (RefSeq : NM_021114**).**

Coordinates of all variations are based on the UCSC GRCh37/hg19 assembly. [&] No RefSeq transcript accession number currently available. The 5'UTR was only sequenced in 363 patients and the whole coding sequence was sequenced in all 611 patients . When the variant was not identified in ExAC the number of ExAC controls indicated corresponds to the number of subject analyzed in ExAC for the nearest described variant. We note that the beginning of *SPINK2* exon1 is particularly poorly covered.

Position	Localisation of the variation	cDNA reference	Consequences on the protein	Nb of homozygous patients (%)	Nb of ExAC homozygous controls (%)	Nb of heterozygous patients (%)	Nb of ExAC heterozygous controls (%)
4:57687851 (rs115163565)	5' UTR	c.1-23 A>T	-	2/363 (0,55)	307/4980 (6,2)	32/363 (8,8)	1816/4980 (36,5)
4:57687849 (rs114591157)	5' UTR	c.1-21 G>C	-	3/363 (0,83)	318/5065 (6,3)	30/363 (8,3)	1852/5065 (36,0)
4:57687828	Exon 1	c.1 A>T	p.M1L	0/611	0/6054	1/611	0/6054

Appendix Table S3. Sequence of the primer used for *SPINK2* sanger sequencing.

Primer name	Primer sequence	Size of amplicon
SPINK2-hSeq-Ex1-F	GAGTGGCGCAGGTAACAGAC	245 bp
SPINK2-hSeq-Ex1-R	CTGGGGAACCGCCAGTAAC	
SPINK2-hSeq-Ex2-F	TGGCTAATGCCTCAAATTCC	339 bp
SPINK2-hSeq-Ex2-R	ACGCAGTCCTCAATGGTTTC	
SPINK2-hSeq-Ex3-F	ACACGGTGAAACCCTGTCTC	552 bp
SPINK2-hSeq-Ex3-R	CAGAGGTTGCAGTGAACCAA	
SPINK2-hSeq-Ex4-F	GTGGGGACTTTCACCCTCTT	423 bp
SPINK2-hSeq-Ex4-R	GCAAAAGCCAAGAAACAAGG	

Appendix Table S4. Sequence of the primer used in RT-PCR and size of the amplified products

Primer name	Primer sequence	Size of amplicon
SPINK2-hRT-F	CTGCTCCTGGCAGTTACCTT	221 bp
SPINK2-hRT-R	CAGGGTCCATTTCGAATGAT	

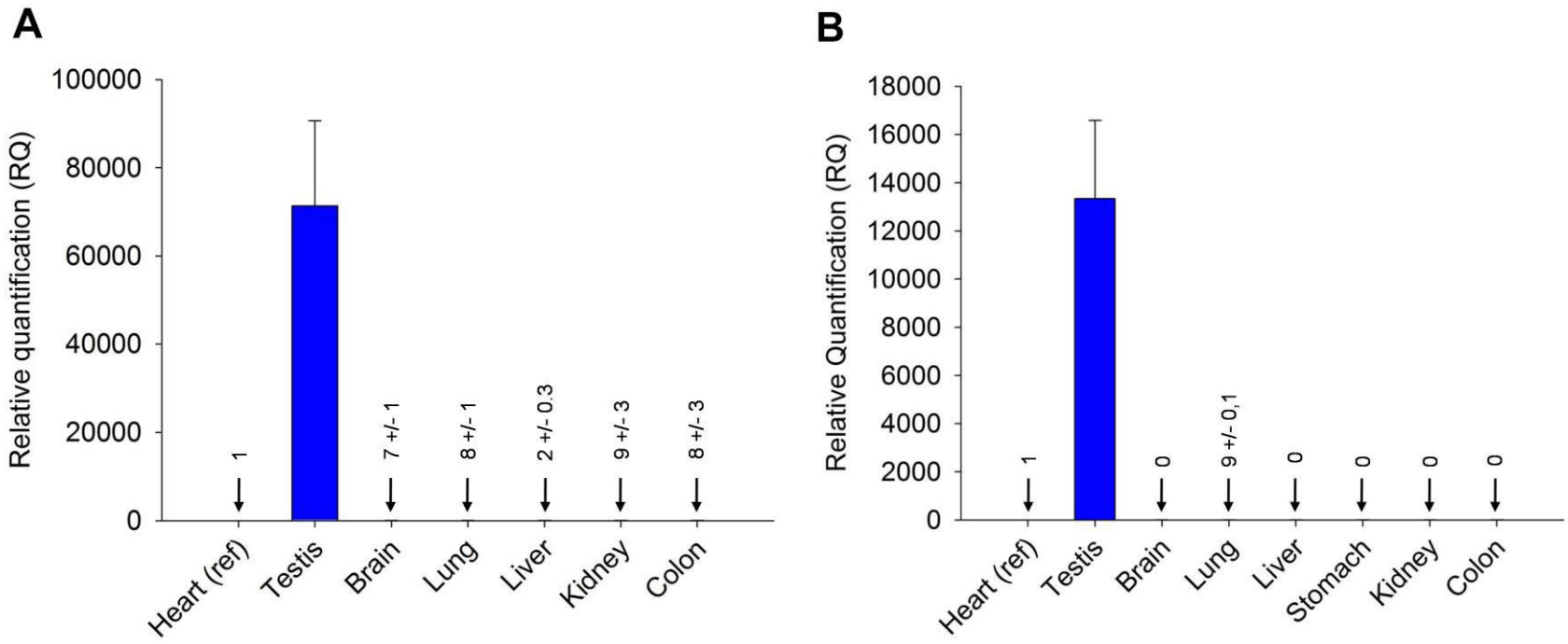
Appendix Table S5. Sequence of the primer used in real time RT-PCR and size of the amplified products

Primer name	Primer sequence	Size of amplicon
SPINK2-hqRT-F	CACTTTAACCCTGTGTGTGG	117 bp
SPINK2-hqRT-R	TCAGCAGGGTCCATTTCGAA	
Spink2-mqRT-F	CTCATGAGACTCTCGACTCTTCC	150 bp
Spink2-mqRT-R	TACATTCATTGCTGTAAGTGTTCATATC	
ACTB-hqRT-F	CCAACCGCGAGAAGATGA	97 bp
ACTB-hqRT-R	CCAGAGGCGTACAGGGATAG	
Actb-mqRT-F	ACCAGAGGCATACAGGGACA	104 bp
Actb-mqRT-R	CTAAGGCCAACCGTGAAAAG	

Appendix Table S6. Sequence of the primer used for genotyping and size of the amplified products.

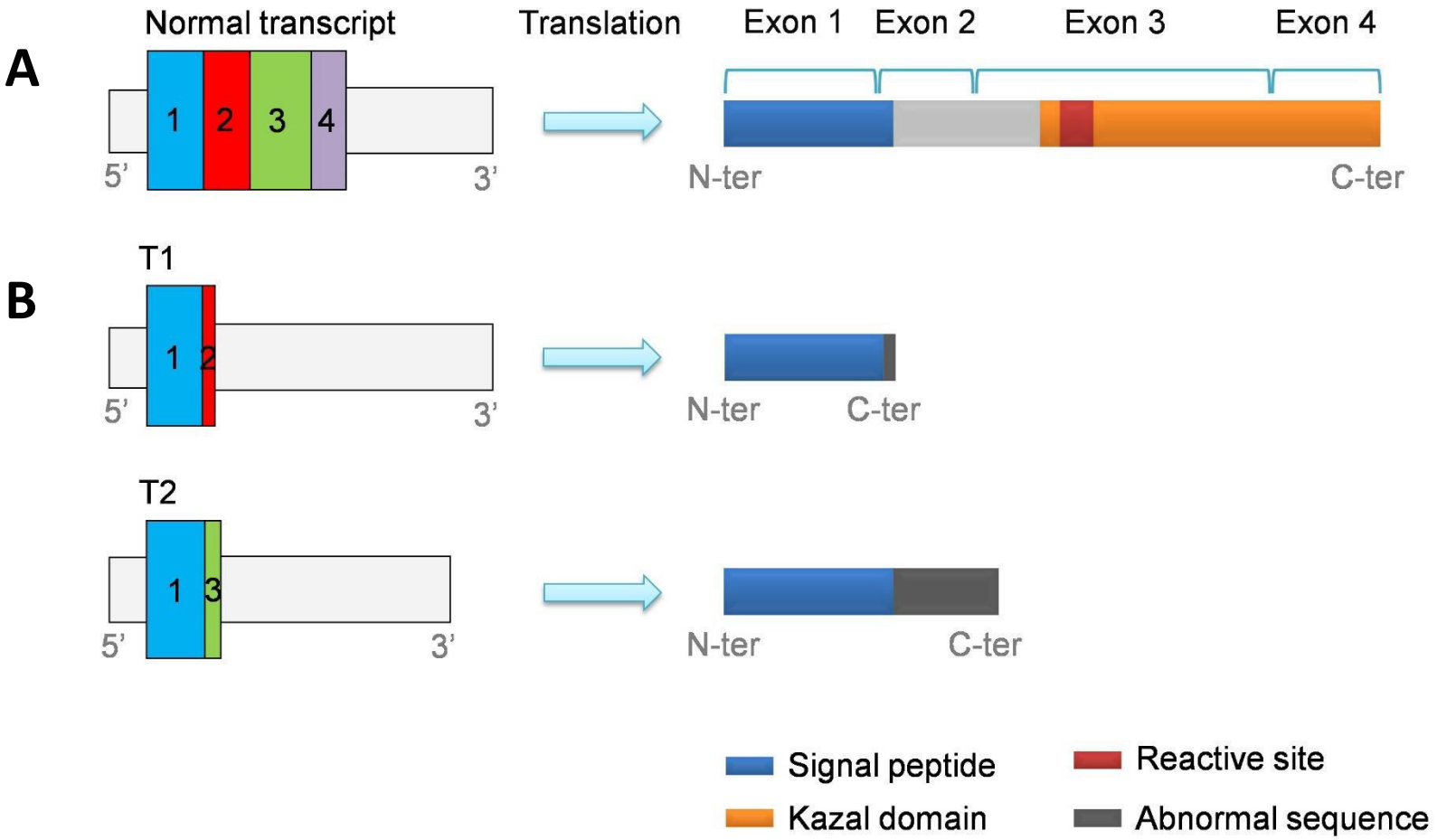
Primer name	Primer sequence	Size of amplicon
Spink2-WT-F	GCAATGGGCGTATCTCAAAT	174 bp
Spink2-WT-R	GGGACCTGATTTTCATGCAC	
Spink2-KO-F	CTCTTCCTCGCTCCCTTCTT	213 bp
Spink2-KO-R	GGGATTCTCCCAATCTCTCC	

Appendix Figure S1. Expression of *SPINK2* mRNA in different organs in human and mouse.
The expression pattern of SPINK2 transcript in various human **(A)** and mouse **(B)** tissues was determined by Quantitative real-time RT-PCR. Values were normalized to the expression level of beta-actin via the 2-ΔΔCT method. The 2-ΔΔCT value was set at 0 in heart, resulting in an arbitrary expression of 1. Data are presented as mean ± standard error mean (n=3). Statistical differences were assessed using t-test.

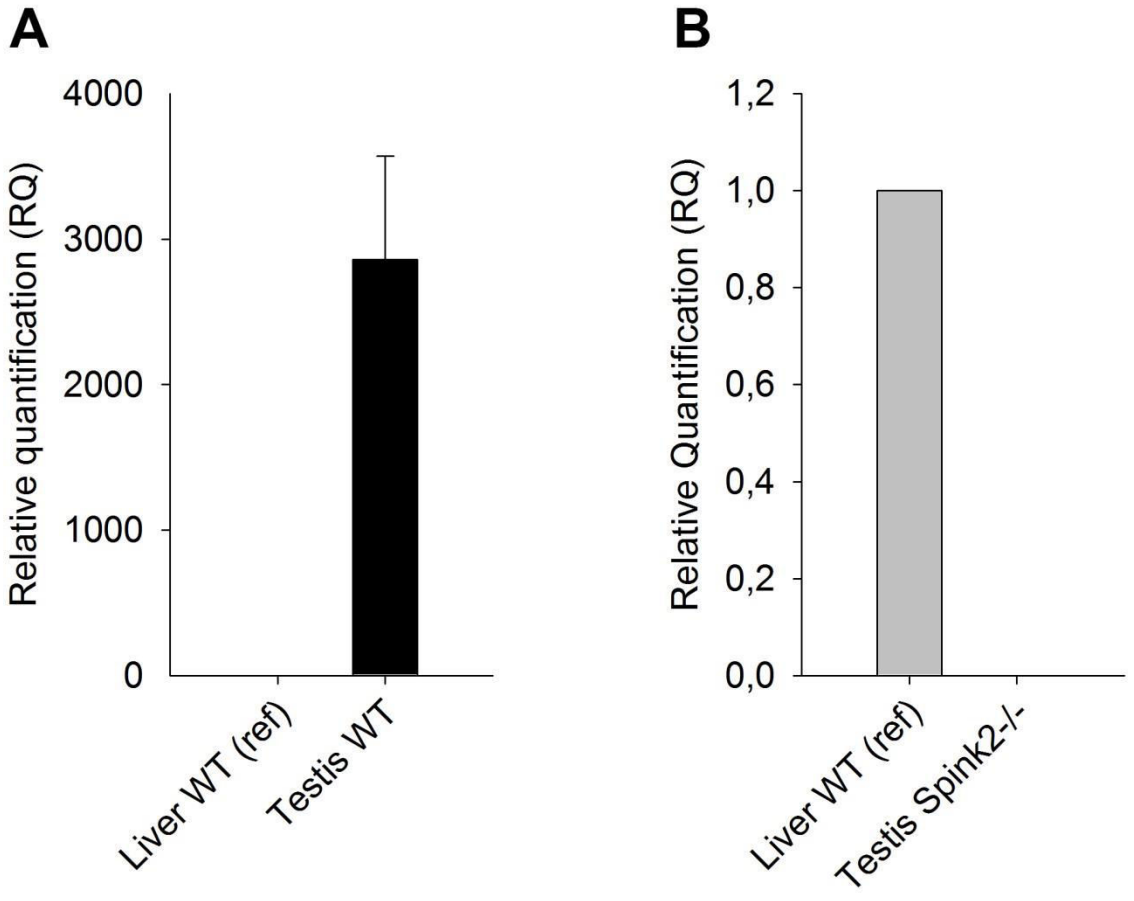


Appendix Figure S2. Impact on the protein structure of the c.206-3C>G *SPINK2* variant

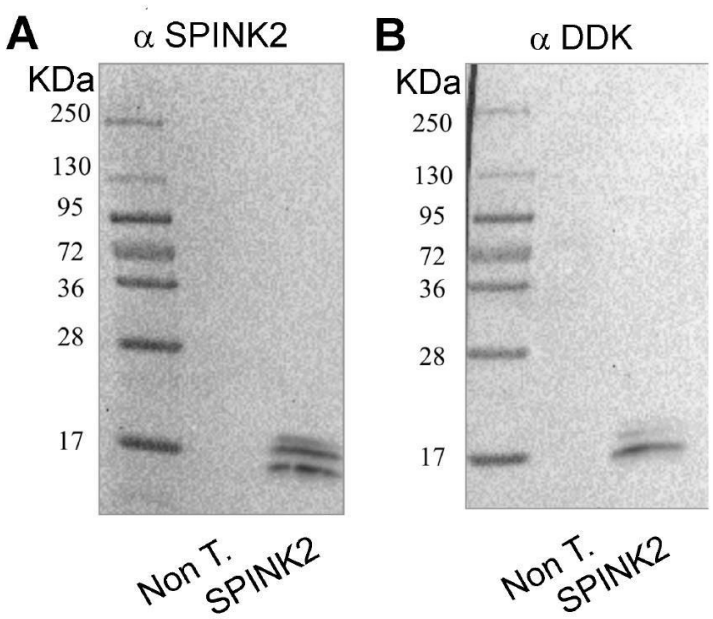
(A) The normal transcript encodes for an 84 amino acids protein predicted to contain a signal peptide at the N-terminus region (blue) and a Kazal-type serine protease inhibitor domain at the C-terminus region (orange and red). UniProtKB accession number for human SPINK2 is P20155. **(B)** The c.206-3C>G splice variant generates two abnormal transcripts, T1 and T2 (see figure 2). Translation of these transcripts is predicted to produce truncated proteins lacking the Kazal-type serine protease inhibitor domain.

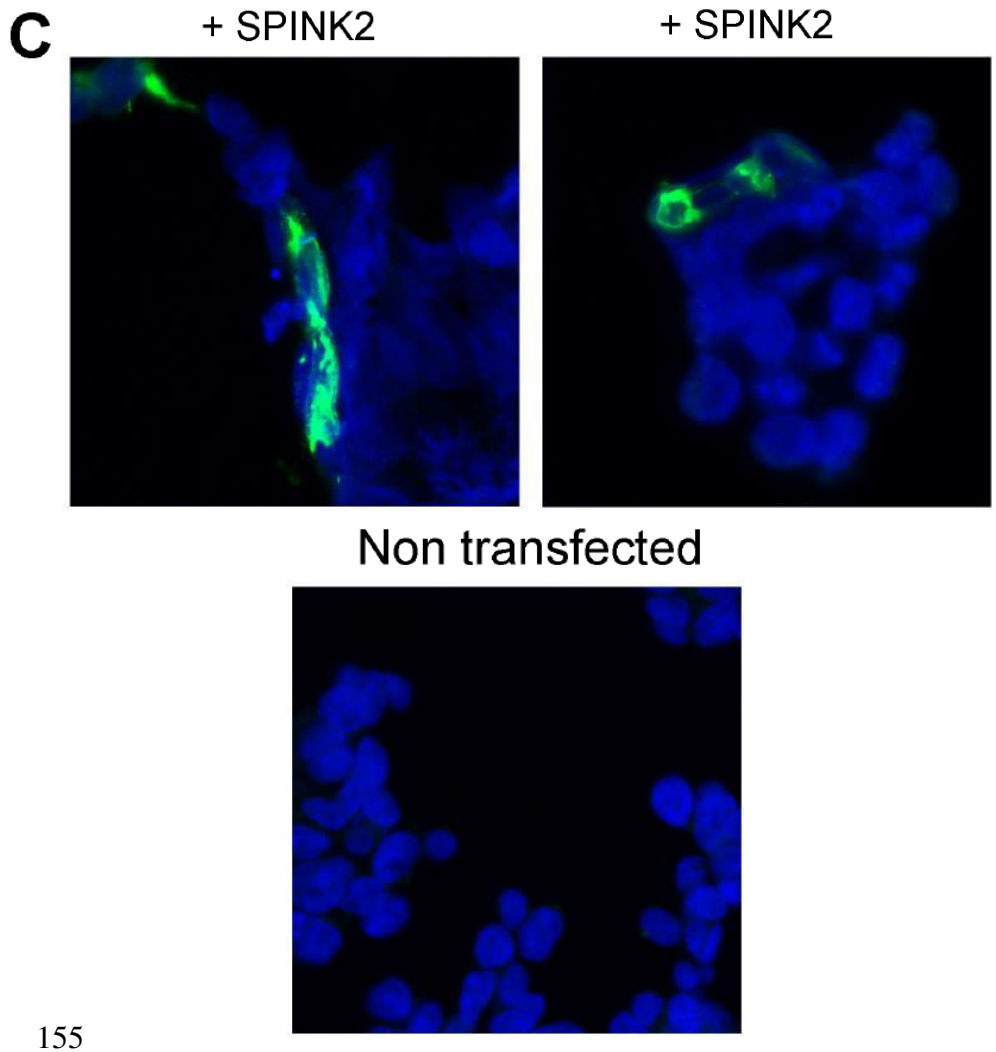


Appendix Figure S3. Absence of *Spink2* expression in testicular extracts from KO mice.
Relative quantification of *Spink2* transcript in testicular cells from **(A)** WT mice (n=3) and **(B)** *Spink2* $^{-/-}$ mice was determined by quantitative real-time RT-PCR. Values were normalized to the expression level of beta-actin via the 2-ΔΔCT method. The 2-ΔΔCT value was set at 0 in liver cells from WT mice (n=3), resulting in an arbitrary expression of 1. Data are presented as mean ± standard error mean. Statistical differences were assessed using t-test.

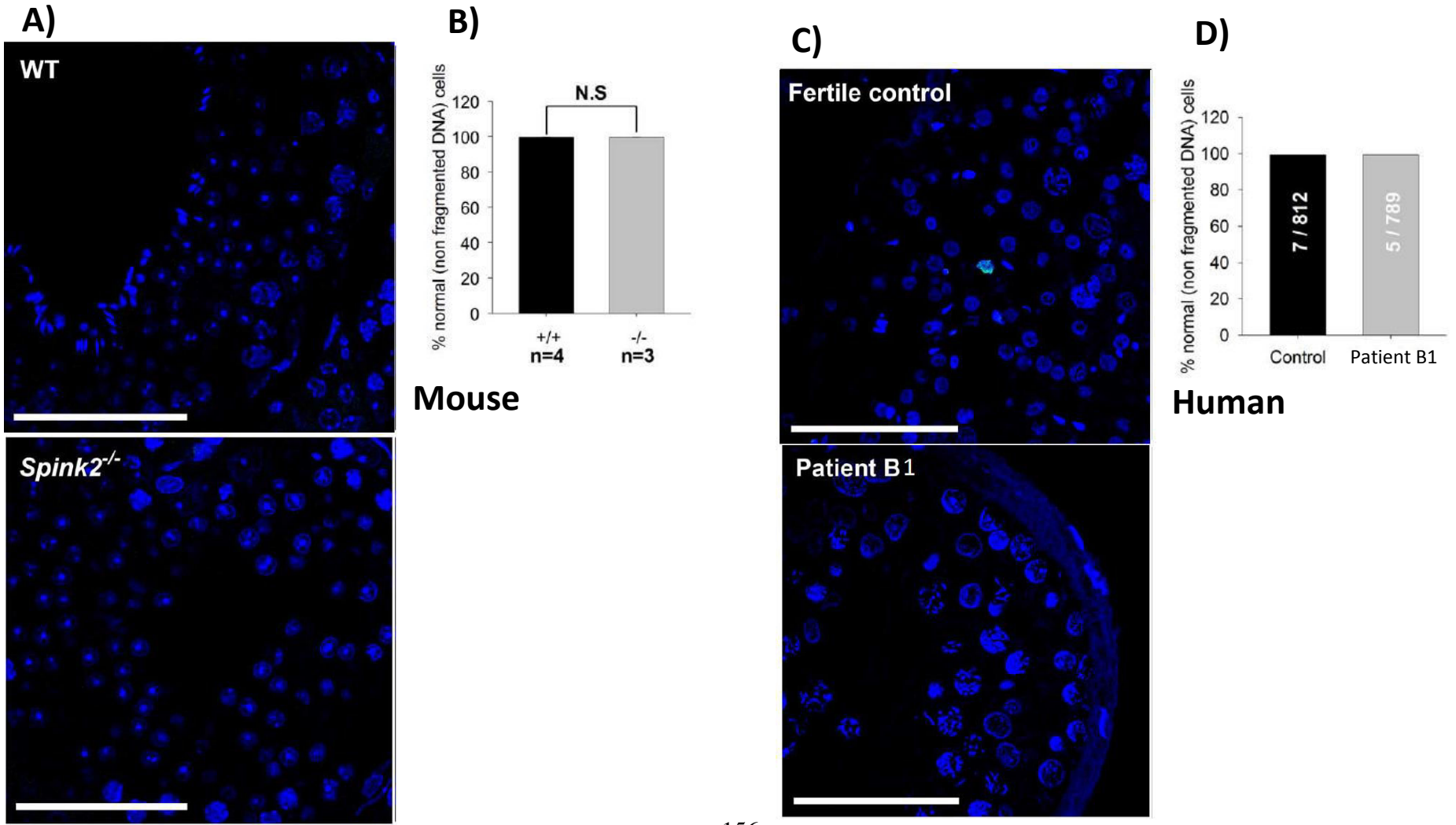


Appendix Figure S4. Specificity of antibodies targeting SPINK2.
(A) Western blot of protein extracts from HEK 293 cells heterologously transfected with or without a plasmid containing human isoform of SPINK2 tagged with DDK and revealed with an anti-SPINK2 antibody. **(B)** Similar experiment but revealed with an anti-DDK antibody. **(C)** Confocal images of HEK 293 cells transfected with human SPINK2-DDK and stained with an anti-SPINK2 antibodies (green) (top). All cells were counterstained with Hoechst (blue, staining) to mark the nuclei. Only transfected cells showed specific staining. Bottom, confoncal images of control non transfected cells showing no staining.

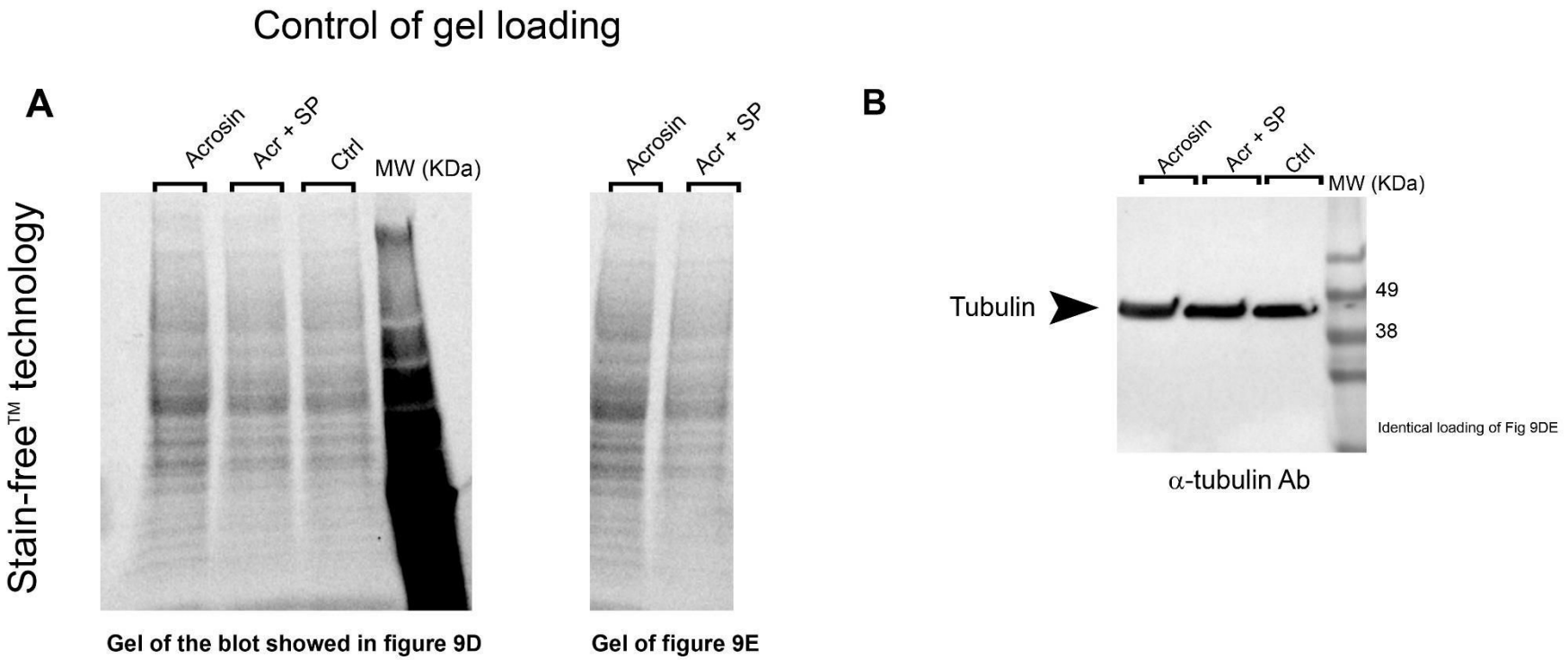




Appendix Figure S5. Absence of Spink2 does not lead to DNA fragmentation in spermatogenic cells.
(A) Representative images of TUNEL staining (green) counterstained with Hoechst (blue) of WT and *Spink2*$^{-/-}$ seminiferous tubule. Scale bar 70 µm. (**B**) Graph showing the % of cells with non fragmented DNA in WT (99.6 +/- 0.3 mean +/- SD, number of cells assessed 230-270 per biological replicate, n=3) and in *Spink2*$^{-/-}$ (99.6 +/- 0.1 mean +/- SD, number of cells assessed 300-380 per biological replicate, n=4) seminiferous tubules. NS "not statistically significant" (t-test) (**C**) Representative images of TUNEL staining (green) counterstained with Hoechst (blue) of fertile control and patient B1 seminiferous tubule. Scale bar 70 µm. (**D**) Graph showing the % of cells with non fragmented DNA in fertile control (number of cells assessed 812, n=1) and patient B1 control (number of cells assessed 789, n=1) seminiferous tubules.



Appendix Figure S6. Equal protein loading was verified by stain-free gel technology and Western blots against house keeping protein (A) Stain-free™ technology was used to control protein loading levels. Left, activated gel showing loading of Western-blot of Fig 9D and right of Fig 9E (**B**) Identical loading as tested in figure 9D and 9D were revealed by anti-tubulin antibody and confirmed equal protein loading between different lanes.



2- Discussion

Les investigations génétiques réalisées dans le cadre de l'azoospermie non-obstructive n'ont pas connu de changement depuis plusieurs décennies. Les tests génétiques réalisés en routine sont le caryotype à la recherche d'anomalies chromosomiques de structure ou de nombre et la recherche des microdélétions des régions AZFs dans le bras long du chromosome Y. Ces deux tests permettent d'établir un diagnostic génétique dans moins de 5% des cas d'infertilité, tous phénotypes confondus, et dans 20% des cas d'azoospermie (Tüttelmann et al., 2018). Grâce à l'avènement récent des technologies de séquençage haut débit comme le séquençage exomique, on espère identifier de nouvelles causes et améliorer le rendement du diagnostic génétique dans ce phénotype.

Des études antérieures de familles consanguines présentant de multiples descendants azoospermiques ont permis d'identifier plusieurs gènes candidats dans l'azoospermie autosomique récessive. En 2014, deux variants candidats ont été décrits dans les gènes *TAF4B* (*TATA-box binding protein associated factor 4b*) et *ZMYND15* (*zinc finger MYND-type containing 15*) (Ayhan et al., 2014). Les souris KO pour les orthologues de ces deux gènes sont infertiles et présentent une ANO. En 2015, trois autres variants ont été décrits chez des patients consanguins azoospermiques dans les gènes *TEX15* (*testis expressed 15*), *MCM8* (*minichromosome maintenance 8 homologous recombination repair factor*) et *SYCE1* (*synaptonemal complex central element protein 1*) (Maor-Sagie et al., 2015; Okutman et al., 2015; Tenenbaum-Rakover et al., 2015). Ces trois gènes sont fortement exprimés dans le testicule et les souris KO pour leurs orthologues présentent un blocage spermatogénique méiotique à l'instar des patients mutés (Bolcun-Filas et al., 2009; Lutzmann et al., 2012; Yang et al., 2008).

En 2015, Yatsenko et al. ont étudié la souris KO pour *Tex11* et ont observé que les mâles présentent une azoospermie associée à un blocage méiotique de la spermatogenèse (Yatsenko et al., 2015). Le séquençage ciblé de l'orthologue humain *TEX11* (*testis expressed 11*), situé sur le chromosome X, chez deux cohortes de patients azoospermiques a permis d'identifier six mutations associées à une perte de fonction de la protéine (Yang et al., 2015; Yatsenko et al., 2015). L'observation histologique des coupes de biopsies testiculaires réalisées chez trois de ces patients a montré un blocage méiotique de la spermatogenèse ce qui est concordant avec le phénotype des souris KO. La fréquence de l'ensemble de ces mutations pathogènes *TEX11* chez les patients azoospermiques a été estimée à 1%. Actuellement,

TEX11 est le seul gène auquel on a associé des mutations pathogènes chez des patients non apparentés atteints d'ANO.

Malgré ces découvertes encourageantes, la majorité des gènes candidats a été identifiée dans une seule étude et n'a pas fait l'objet d'une validation par une nouvelle étude indépendante. La difficulté de reproduire les mêmes résultats vient de l'absence de caractérisation phénotypique complète chez les patients infertiles étudiés. Dans l'azoospermie, le phénotype testiculaire est souvent inconnu au moment de l'analyse des résultats de l'exploration génétique. L'autre limite des études réalisées est leur puissance. En effet, le nombre de patients analysés est souvent faible et insuffisant pour détecter les variants génétiques rares. Un autre point à critiquer est la concentration de l'analyse sur des familles consanguines pour rechercher des causes autosomiques récessives alors que les causes autosomiques dominantes pourraient être majoritairement impliquées. D'autres efforts doivent donc être fournis pour mieux caractériser la génétique complexe de ce phénotype très sévère.

La majorité des patients qui présentent des troubles de la spermatogenèse pourraient bénéficier des techniques de procréation assistée comme la fécondation in vitro standard (FIV) ou associée à une micro-injection du spermatozoïde dans le cytoplasme de l'ovocyte (FIV-ICSI). Cependant, la majorité des patients présentant une ANO ne peuvent pas bénéficier de ces traitements. L'identification des causes génétiques responsables d'ANO ainsi que la caractérisation de leur pathogénie moléculaire pourraient constituer une base pour le développement de nouvelles stratégies thérapeutiques plus personnalisées. Dans ce cas particulier, nous avons montré que les mutations du gène *SPINK2* pouvait être associés à une ANO par blocage post-méiotique de la spermatogenèse. Le fait que le pool des spermatogonies présentes dans l'épithélium séminifère soit conservé, nous permet d'envisager des thérapies ciblant l'activité protéolytique dérégulée de l'acrosine. Des traitements similaires existent déja dans le cadre de la pancréatite chronique causée par une altération du gène *SPINK1* et une dérégulation de l'activité protéolytique de la trypsine (Kambhampati et al., 2014). Cependant, en raison de la présence d'une barrière hémato-testiculaire très sélective, il est difficile d'obtenir dans les tubules séminifères des concentrations satisfaisantes du produit administré même par voie systémique. D'autres moyens doivent être envisagés et développés pour répondre à ce challenge.

Chapitre II. Le phénotype MMAF

1- Article 3 : A homozygous ancestral SVA-insertion mediated deletion in *WDR66* induces multiple morphological abnormalities of the sperm flagellum and male infertility

Zine-Eddine Kherraf, Amir Amiri-Yekta, Denis Dacheux, Thomas Karaouzène, Charles Coutton, Marie Christou-Kent, Guillaume Martinez, Pauline Le Tanno, Selima Fourati Ben Mustapha, Lazhar Halouani, Ouafi Marrakchi, Mounir Makni, Habib Latrous, Mahmoud Kharouf, Karin Pernet Gallay, Nicolas Thierry-Mieg, Hamid Gourabi, Derrick R Robinson, Serge Crouzy, Michael Blum, Aminata Touré, Raoudha Zouari, Christophe Arnoult, Mélanie Bonhivers & Pierre F. Ray.

The American Journal of Human Genetics. Accepté.

Contexte et principaux résultats de l'étude :

La première exploration génétique du phénotype MMAF chez des cas familiaux, d'abord par cartographie d'homozygotie puis par séquençage exomique, nous a permis d'identifier *DNAH1* comme premier gène impliqué dans ce phénotype (article 6). Nous avons réalisé par la suite, le séquençage exomique chez 78 patients MMAF non apparentés et avons identifié des variants bi-alléliques délétères chez 28% des patients analysés dans les gènes *DNAH1*, *CFAP43* et *CFAP44* (article 4). Nous avons poursuivi l'analyse bio-informatique des données de séquençage exomique de ces 78 patients en ajoutant un module (logiciel ExomeDepth) à notre pipeline d'analyse bioinformatique. Ce module permet l'analyse des CNVs (copy number variants) qui sont des délétions ou des duplications de séquences qui peuvent concerner un simple exon et jusqu'à plusieurs gènes. On note que les analyses initiales permettaient de détecter uniquement les substitutions délétères ou petites insertions/délétions de quelques nucléotides. L'analyse CNV chez ces patients nous a permis d'identifier une délétion homozygote de deux exons (20/22 et 21/22) du gène *WDR66* chez 7 patients non-apparentés d'origine nord-africaine.

Grâce à une stratégie de « *PCR walking* » nous avons caractérisé l'étendue de la délétion et séquencé les points de cassure. Vingt patients supplémentaires ont été testés et cette même mutation a été retrouvée chez deux patients (résultats non-inclus dans l'article). Nous avons observé que tous les patients présentaient exactement la même délétion de 7731 pb et présentaient le même haplotype au niveau de cette région indiquant un effet fondateur. L'analyse bio-informatique des points de cassure et des séquences flanquantes montre que cet évènement est la résultante d'une recombinaison complexe entre des rétrotransposons de type SINE-VNTR-*Alus* (SVA) localisés dans *WDR66* et dans le gène *CLIP1* distant de 400Mb du locus délété. Aucune autre anomalie génétique n'a été détectée dans ce locus chez les patients de la cohorte.

Nous avons confirmé par RT-qPCR à partir d'un panel de tissus humains que *WDR66* est spécifiquement exprimé dans le testicule. Ce gène code pour une protéine appartenant à la famille des « WD-repeat proteins » qui sont impliquées dans des processus cellulaires variés. Il a été montré précédemment que *Fap251*, l'orthologue de *WDR66* chez *Chlamydomonas* code pour une protéine localisée dans l'axonème ciliaire au niveau du CSC (calmodulin and spoke-associated complex) qui est situé dans la base du radial-spoke 3 (RS3). Il a également été montré que l'inactivation de *Fap251* chez *Tetrahymena* cause une baisse

significative de la mobilité ciliaire et une déstructuration voir une perte des RS3. Grâce à l'utilisation du modèle *Trypanosoma* nous avons démontré que la partie délétée de la protéine était critique pour sa fonction. En effet, nous avons montré que l'expression de l'orthologue de *WDR66* chez ce protiste (CFap251Try) tronqué pour la région C-terminale qui est délétée chez les patients était également associée à une altération de la mobilité flagellaire. Ce résultat permet de confirmer que chez les patients délétés, c'est bien la délétion des deux exons 20/22 et 21/22 de *WDR66* qui est responsable du phénotype MMAF.

Ce travail démontre que les altérations de *WDR66* sont responsables d'environ 10% du phénotype MMAF chez les patients Nord Africains et suggère que la protéine WDR66 joue un rôle déterminant dans le maintien de l'intégrité de l'axonème et de la mobilité du flagelle spermatique.

A homozygous ancestral SVA-insertion mediated deletion in *WDR66* induces multiple morphological abnormalities of the sperm flagellum and male infertility

Zine-Eddine Kherraf[1,2], Amir Amiri-Yekta[1,2,3], Denis Dacheux[4,5], Thomas Karaouzène[1], Charles Coutton[1,6], Marie Christou-Kent[1], Guillaume Martinez[1,6], Nicolas Landrein[4], Pauline Le Tanno[1], Selima Fourati Ben Mustapha[7], Lazhar Halouani[7], Ouafi Marrakchi[7], Mounir Makni[7], Habib Latrous[7], Mahmoud Kharouf[7], Karin Pernet Gallay[8], Hamid Gourabi[3], Derrick R Robinson[4], Serge Crouzy[10], Michael Blum[9], Nicolas Thierry-Mieg[9], Aminata Touré[11,12,13], Raoudha Zouari[7], Christophe Arnoult[1], Mélanie Bonhivers[4†] and Pierre F. Ray[1,2 Σ†]

[1]Genetic Epigenetic and Therapies of Infertility, Institute for Advanced Biosciences, Inserm U1209, CNRS UMR 5309, Université Grenoble Alpes, F-38000 Grenoble, France
[2]CHU de Grenoble, UM GI-DPI, Grenoble, F-38000, France

[3] Department of Genetics, Reproductive Biomedicine Research Center, Royan Institute for Reproductive Biomedicine, ACECR, Tehran, Iran.PO Box 16635-148

[4]University Bordeaux, Microbiologie Fondamentale et Pathogénicité, CNRS UMR 5234, Bordeaux, France

[5]Institut Polytechnique de Bordeaux, Microbiologie Fondamentale et Pathogénicité, CNRS UMR 5234, Bordeaux, France

[6]CHU de Grenoble, UM de Génétique Chromosomique, Grenoble, F-38000, France
[7]Polyclinique les Jasmins, Centre d'Aide Médicale à la Procréation, Centre Urbain Nord, 1003 Tunis, Tunisia

[8]Grenoble Neuroscience Institute, INSERM 1216, Grenoble, F38000, France

[9]Univ. Grenoble Alpes / CNRS, TIMC-IMAG, F-38000 Grenoble, France

[10] Laboratoire de Chimie et Biologie des Métaux, Institut de Recherche en Technologie et Sciences pour le Vivant, CEA iRTSV/LCBM/GMCT, CNRS UMR 5249, Université Grenoble Alpes, F-38054, Grenoble Cedex 9, France.

[11]Institut National de la Santé et de la Recherche Médicale, INSERM U1016. Institut Cochin. Paris 75014, France
[12]Centre National de la Recherche Scientifique, CNRS UMR8104. Paris 75014, France
[13]Université Paris Descartes, Sorbonne Paris Cité, Faculté de Médecine. Paris 75014, France

[Σ]**Corresponding author**: Pierre F Ray, UM GI-DPI, IBP, 38043 Grenoble cedex 9, France. Tel: +33-4-76-76-55-73; E-mail: pray@chu-grenoble.fr

[†] Shared leadership

ABSTRACT

Multiple morphological abnormalities of the sperm flagellum (MMAF) is a severe form of male infertility defined by the presence of a mosaic of anomalies including short, bent, curled, thick or absent flagella caused by a severe disorganization of the axoneme and of the peri-axonemal structures. Mutations in *DNAH1*, *CFAP43* and *CFAP44*, three genes encoding axonemal related proteins, have been described to account for approximately 30% of the MMAF cases reported so far. Here, we searched for pathological copy number variants (CNVs) in whole exome sequencing (WES) data from a cohort of 78 MMAF subjects to identify new MMAF associated genes.

We identified a homozygous deletion removing the two penultimate exons of *WDR66* (also named CFAP251), a gene coding for an axonemal protein preferentially expressed in the testis and described to localize to the CSC (Calmodulin and Spoke associated Complex), at the base of radial spoke 3. Sequence analysis of the breakpoint region revealed in all deleted subjects the presence of a single chimeric SVA (SINE-VNTR-*Alu*) at the breakpoint site, suggesting that the initial deletion event was potentially mediated by an SVA insertion-recombination mechanism.

Study of *Trypanosoma WDR66*'s ortholog (*TbWDR66*) highlighted a high sequence and structural analogy with the human protein and confirmed axonemal localization of the protein. Reproduction of the human deletion in *TbWDR66* impaired flagellar movement thus confirming *WDR66* as a causative gene for MMAF phenotype and highlighted the importance of the WDR66 C-terminal region.

INTRODUCTION

Male infertility is a common reproductive disorder characterized by an extremely heterogenous pathogeny mostly due to an altered sperm production during spermatogenesis. Spermatogenesis is a complex process involving three successive steps: the mitotic proliferation of spermatogonial stem cells, the meiotic division of spermatocytes and finally the spermiogenesis, a key post-meiotic event contributing to major morphological changes of the spermatids such as chromatin compaction, acrosome biogenesis and assembly of flagellum structures [1]. Transcriptomic analyses of germ cells revealed the dynamic transcription of over 4,000 genes during human spermatogenesis [2] but the genetic basis of male infertility has so far not been extensively studied [3]. Monomorphic sperm defects such as globozoospermia, sperm macrocephaly, headless spermatozoa and multiple morphological abnormalities of the sperm flagella (MMAF) have been demonstrated to be caused by genetic defects affecting several testis-specific genes [4]. For affected men, intracytoplasmic sperm injection (ICSI), if indicated, represents the unique option for conception but significant efforts should be made to characterize the genetic causes behind their sperm defects and to better understand the nature of their molecular alterations.

In this study, we focused on subjects presenting a typical MMAF phenotype characterized by the presence of immotile sperm with short, bent, coiled, absent flagellum or flagellum with an irregular width (Figure 1). In 2014, our team first identified pathogenic mutations in the dynein axonemal heavy chain 1 (*DNAH1*) gene in 28% of the MMAF individuals analyzed [5]. Recently, whole exome sequencing (WES) of 78 MMAF individuals allowed us to identify bi-allelic mutations in *CFAP43* and *CFAP44* (previously known as *WDR96* and *WDR52* respectively) encoding tryptophan–aspartate (W-D) repeat proteins of unknown functions highly expressed in the testis [6]. Mutations in these genes have also been reported by others [7,8] and account for 20.5% of the MMAF individuals of our cohort [3]. Other rarer homozygous missense mutations have been reported in single familial cases of MMAF in *CEP135* [9] and *AK7* [10], encoding for a centriolar protein and an axonemal adenylate kinase, respectively. These results highlight the genetic heterogeneity associated with this phenotype and, importantly, they indicate that additional genes are likely to be associated with MMAF.

Copy number variations (CNVs), structural variants defined as large insertions, deletions or duplications of genomic segments of more than 1 kb [11] are important elements of human genetic diversity present in healthy people [12–14], but are also increasingly recognized as

an important etiology of many human diseases including neuro-developmental diseases, chronic inflammations, and cancers [15]. WES is a cost effective technique permitting to detect pathogenic mutations. Its efficiency at identifying single nucleotide variants (SNVs) and small InDels has been proven countless times. It also has recently been used for detecting copy number variations CNVs [16,17] but its performance and limits for detecting CNVs are still undefined. We reanalyzed the WES data from our cohort of 78 WES individuals described previously [6] focusing on CNVs detection using the ExomeDepth software [18] and identified a homozygous intragenic deletion of 8.4 kb encompassing *WDR66* exons 20 and 21 (out of twenty) in seven individuals. *WDR66*, also known as *CFAP251,* encodes a WD-repeat protein preferentially expressed in the testis. We characterized the extent of the deletion at the nucleotide level and conclude that the deletion was caused by a SVA-insertion mediated deletion mechanism [19]. Using the Trypanosoma model we confirmed the importance of the *WDR66* deleted domain. After *CFAP43 (WDR96)*, and *CFAP44 (WDR52)*, two WDR genes associated with MMAF, we demonstrated that *WDR66* is a new MMAF associated gene and confirm the importance of WD-repeat proteins in human diseases and especially in male infertility.

MATERIAL AND METHODS

Study Participants

We included 78 subjects consulting for isolated primary infertility between 2008 and 2016 and presenting asthenozoospermia due to a combination of morphological defects of the sperm flagella described as follow: absent, short, bent, coiled flagella and/or flagella of irregular width. Forty-six individuals were of North African origin and consulted for primary infertility at the Clinique des Jasmin in Tunis. Ten individuals originated from the Middle East (Iranians) and were treated in Tehran at the Royan Institute (Reproductive Biomedicine Research Center) for primary infertility and twenty-two individuals were recruited in France: 21 at the Cochin Hospital and one in Lille's hospital. Saliva and/or peripheral blood were obtained for all participants. During their medical consultation, all subjects answered a health questionnaire focused on primary ciliary dyskinesia (PCD) manifestations. All had an isolated MMAF phenotype with no mention of any other PCD symptoms. Ciliary function could however not be examined (in nasal or lung cilia) and we cannot exclude the possibility of a sub-clinical alteration of the ciliary function. Informed consent was obtained from all the subjects participating in the study according to the local protocols and the principles of the Declaration of Helsinki. Controls from fertile individuals with normal spermograms were obtained from CRB Germetheque. Consent for CRB storage was approved by the CPCP Sud-Ouest of Toulouse (coordination of the multi-sites CRB Germetheque). The detailed description of the cohort can be found in [6].

Study of sperm morphology, spermograms and transmission electron microscopy

Sperm morphology was first analyzed on slides under a microscope after modified Papanicolaou staining. Sperm analysis was carried out in the source laboratories during the course of the routine biological examination of the patient, according to World Health Organization (WHO) guidelines [20]. Then, transmission electron microscopy (TEM) assessment was carried out. Sperm cells from an affected individual deleted for *WDR66* were fixed in PBS with 4% paraformaldehyde for 60 minutes. They were pelleted (1,000 x g 10 min), resuspended in fixation buffer (2.5 % glutaraldehyde, 2% PFA, 100 mM Phosphate buffer pH 7.4, 50 mM sucrose for 2 h). Fixed cells were washed in buffer for 10 min, post-fixed in 1% OsO_4 for 1 h, washed 3 times in water, then samples were stained in 2% uranyl acetate in water at 4°C overnight. Samples were next washed in water, dehydrated in ethanol, embedded in Epoxy resin and polymerized during 60 hours at 60°C. Sections were stained with 2% uranyl and lead

citrate, and visualized in a transmission electron microscope (JEOL 1200EX). Trypanosomes were prepared for TEM as above with the exception that they were embedded in Spur resin.

Whole-Exome Sequencing and data processing

Genomic DNA was isolated from saliva using Oragen DNA extraction kit (DNAgenotech®, Ottawa, Canada). Coding regions and intron/exon boundaries were enriched using the "all Exon V5 kit" (Agilent Technologies, Wokingham, UK). WES and initial data analyses were performed as described in [6]. CNVs, which had not been analyzed previously, were analyzed with the ExomeDepth software package [18]. We retained only overlapping CNVs found in at least two subjects and not described as benign in Cathagenia. All data are based on GRCh37/hg19.

RT-PCR and qRT-PCR

qRT-PCR was performed on a cDNA Panel from testis, lung, kidney and heart tissues (Clontech), using a StepOnePlus™ Real-Time PCR System (Life Technologies®) with Power SYBR® Green PCR Master Mix (Life Technologies®) according to the manufacturer's protocol. Quantification of the fold change in gene expression was determined by the relative quantification (RQ) method ($2^{-\Delta\Delta CT}$) using the *GAPDH* gene as a reference. Data are shown as the mean of RQ ± standard deviation. Primers are described in Table S2.

Immunofluorescence and confocal microscopy on human sperm cells

Immunostaining was carried out on samples from two patients P7 and P9. Sperm was fixed in PBS, 4% paraformaldehyde for 7 min. After washing in PBS, the sperm suspension was spotted onto 0.1% poly L-lysine pre-coated slides (Thermo Scientific). After attachment, sperm was permeabilized with 0.1% (v/v) Triton X-100-DPBS (Triton X-100; Sigma-Aldrich) for 15 min at room temperature. Slides were then blocked in 5% corresponding normal serum-DPBS (normal goat or donkey serum; GIBCO, Invitrogen) and incubated overnight at 4°C with primary antibodies. Polyclonal rabbit anti-WDR66 antibody was purchased from Sigma-Aldrich, France and used at 1:50. Monoclonal mouse anti-acetylated α-tubulin antibody was purchased from Sigma-Aldrich and used at 1:2,500 dilution. The slides were then washed and incubated with secondary antibody diluted at 1:800 (Dylight 549-conjugated goat anti-mouse IgG and Dylight 488-conjugated goat anti-rabbit IgG, Jackson ImmunoResearch) and Hoechst

33342 (Sigma-Aldrich) for 1h at RT, rinsed and mounted with Dako mounting medium (Life Technology). Images were acquired on a confocal microscopy Zeiss LSM 710 confocal microscope and processed using Zen 2009 software. The list of the antibodies used in the study is provided in Table S3.

Breakpoints investigation

A walking PCR approach was conducted to get closer to the deleted fragment. A long-range PCR was performed to amplify the CNV breakpoint junctions and the product was analyzed by Sanger sequencing to refine the breakpoints. To investigate the potential mechanism causing the deletion, the sequences around the breakpoints were aligned with each other using the Basic Local Alignment Search Tool (BLAST) and in addition, were analyzed for different repetitive elements using repeatmasker software (http://www.repeatmasker.org). PCR primers and protocols used for each individual are listed in Table S1. Sequencing reactions were carried out with BigDye Terminator v3.1 (Applied Biosystems). Sequence analysis was carried out on ABI 3130XL (Applied Biosystems). Sequences were analyzed using SeqScape software (Applied Biosystems).

Trypanosoma brucei cell lines, culture, transfection, and western blotting

The trypanosome cell lines used in this study derived from the parental (WT) procyclic form *T. brucei* 427 29-13 (present in the insect host, PCF) and bloodstream form *T. brucei* 427 90-13 strains (present in the mammalian host, BSF), both co-expressing the T7 RNA polymerase and tetracycline repressor [21]. Cells were cultured, transfected and cloned as described in [22,23] in medium supplemented with hygromycin (25 µg/mL for PCF, 5 µg/mL for BSF), neomycin (10 µg/mL for PCF, 2.5 µg/mL for BSF), and blasticidin (20 µg/mL for PCF and BSF) for constitutive expression of 10TY1-tagged proteins, and with phleomycin (5 µg/mL for PCF, 2.5 µg/mL for BSF) for RNAi inducible cells. RNAi interference was induced with tetracycline (10 µg/mL).

To produce transgenes at their endogenous loci encoding proteins tagged at the C-terminus with a 10TY1 tag (TbWDR66$_{TY1}$, TbWDR66-ΔCter$_{TY1}$), WT cells were transfected with tagging cassettes that were obtained by PCR using 1) a forward primer containing 80 nucleotides from the 3′-end of *TbWDR66* or *TbWDR66-ΔCter* ORF and a sequence specific of the pPOTv7-10TY1-neo vector [24] to amplify the TY1 tag, and 2) a reverse primer containing

80 nucleotides from the 3'-UTR of *TbWDR66*, and a sequence specific of the vector to amplify a selection gene [24]. For RNA interference (RNAi), a fragment of *TbWDR66* ORF (bp 2719 to 3070, corresponding to the aa 903 - 1023 C-terminal domain) was cloned into p2T7tiB [25] and transfected into the TbWDR66$_{TY1}$ and TbWDR66-ΔCter$_{TY1}$ background cell lines. The primer sequences are presented in Table S4 and a schematic of the constructs in Figure S1. RNA interference was induced by tetracycline (10 µg/mL). For western blotting, samples were prepared and treated as in [6].

Immunofluorescence on Trypanosoma cells

Cells were collected, washed, and processed for immuno-labelling as in [3]. Briefly, cells were pelleted and washed, loaded on poly-L-lysine coated slide, and air-dried. After rehydration, cells were fixed and permeabilized in methanol at -20°C, rehydrated in PBS, then incubated with the primary antibodies in two steps: 1) anti-TY1 (BB2) followed by anti-mouse IgG1 Alexa594-conjugated (Thermofischer A21125), 2) anti-PFR2 (rabbit polyclonal) and anti-tubulin TAT1 mouse IgG2a primary antibodies followed by anti-rabbit ATTO-N647-conjugated (Sigma 40839) and anti-mouse IgG2a Alexa-conjugated (Thermofischer A21131) secondary antibodies. Nuclei and kinetoplasts were stained with DAPI (10 µg/mL) in epifluorescence experiments. Epifluorescence images were acquired on a Zeiss Imager Z1 microscope, using a Photometrics Coolsnap HQ2 camera, with a Zeiss 100x objective (NA 1.4) using Metamorph software (Molecular Devices), and analysed with ImageJ. STED images were acquired on a Leica DMI6000 TCS SP8 X-STED microscope with an HC PL APO CS2 100x/1.40 oil objective and de-convolved with Huygens Pro 16.10. A list of all the antibodies is supplied in Table S5.

Video-microscopy on Trypanosoma cells

Video-microscopy was carried out as described in [26]. Briefly, BSF cells (WT and RNAi induced cells for 48 h) were washed and cell mobility was recorded by phase contrast on a Zeiss AxioImager with a 40X objective (NA 1.3). Fifty seconds of digital video from separate regions were captured and analysed using Metamorphoh software (Molecular Devices) and video edited using ImageJ (setting frame rate were10 fps).

Electron microscopy on Trypanosoma cells

Cells were fixed in culture medium, by the addition of glutaraldehyde to a final concentration of 2.5%, for 60 min. They were pelleted (1,000 g, 10 min), resuspended in fixation buffer (2.5 % glutaraldehyde, 2% PFA, 100 mM Phosphate buffer pH 7.4, 50 mM sucrose for 2 H). Fixed cells were washed in water for 10 min, post-fixed in 1% OsO4 for 1 H, washed 3 times in water, then en bloc stained in 2% uranyl acetate in water at 4°C. Samples were next washed in water 3x for 15 min, dehydrated in ethanol, embedded in Spurr resin and polymerised overnight at 60°C. Sections (50-60 nm) were stained with UranyLess 1 minute, washed in water followed by lead citrate. Sections were visualized on a TECNAI 12 TEM.

RESULTS

Identification of the identified CNV

Using the ExomeDepth software package [18] we re-analyzed the WES data obtained from the 78 MMAF individuals described in [6]. Only one homozygous deleterious deletion was identified in several subjects and affected a testis-expressed gene. It was found in seven apparently unrelated North-African individuals and removed exons 20 and 21 (out of 22) of the *WDR66* gene. The deletion was found in 7 out of the 46 Tunisian MMAF subjects analyzed here, indicating that in this population, the *WDR66* deletion is responsible for 15% of the cases. The presence of the deletion was visualized in all deleted individuals using the Integrative Genomics Viewer (IGV) software (Figure 2A). The presence of the deletion in patients was also confirmed by CGH array (Figure S2). Consistent with a negative selection of homozygotes throughout evolution, the identified variant was absent from public databases, including DECIPHER (https://decipher.sanger.ac.uk/) [27], Database of Genomic Variants (DGV; http://dgv.tcag.ca/) [28], and Copy Number Variation in Disease (CNVD; http://bioinfo.hrbmu.edu.cn/CNVD) [29] and only found in one heterozygous Non-Finnish European (NFE) from the 60,706 unrelated individuals aggregated in the ExAC database (http://exac.broadinstitute.org). We note however that all deleted subjects were North African, a population which is not well represented in these databases.

WDR66 gene description, transcripts expression and protein localization.

WDR66, also known as *CFAP251* (cilia- and flagella-associated protein 251) is located on chromosome 12 (ENSG00000158023, chr12: 122,355,768-122,441,833), approximately 11 Mb from the telomere of 12q and contains 22 exons. Transcription of *WDR66* leads to two isoforms T1 (NM_144668) encoding 1,149 amino-acids (aa) (Q8TBY9-1) and T2 (NM_001178003) encoding 941 aa (Q8TBY9-2). As the deleted exon 20 and 21 are only present in T1 it is important to measure the respective expression of both transcripts in tissues of interest to confirm that T1 is expressed in the testis. We therefore measured the respective expression levels of the transcripts by performing quantitative RT-PCR (qRT-PCR) in different human tissues using specific primers for each isoform T1 and T2. Results showed that T1 is highly and predominantly expressed in the testis and moderately expressed in other ciliated tissues (especially the lung) whereas T2 has a negligible expression in all tested tissues (Figure 2B). These results thus indicate that T1, the transcript affected by the identified deletion, is the

main physiological transcript and is highly expressed in the testis (Figure 2C). In this manuscript we will therefore refer to this transcript (ENST00000288912.8) and to the protein it encodes. Consequences of the deletion on the transcript are c.[3007_3337del].

Using the InterPro domain annotation program (http://www.ebi.ac.uk/interpro/), we found that nine WD40 repeat domains are present in both isoforms T1 and T2, and one calcium-binding EF-hand domain is present only in isoform T1 in the C-terminus region (Figure 2D and Table S7) indicating that the activity of this isoform could be modulated by calcium. In NP_653269 (the protein encoded by T1) the first amino acid (aa) affected by the identified deletion is predicted to be aa1003. Deletion of exons 20 and 21 induce a frameshift and translation of exon 22 is predicted to be incorrect for the first 25 aa before inducing a stop codon 26 nucleotides after the first abnormally coded aa: p.[Ile1003Lysfs26].

We then wanted to determine if the deleted transcript is present in sperm from individuals carrying the genomic deletion or subjected to mRNA decay. Unfortunately, we could not obtain enough biological material to perform a reliable RT-PCR. mRNA decay has however been shown to be elicited by abnormally long 3'UTRs and to be more efficient when a premature stop codon is introduced near the 5' end of the transcript [30]. The deletion identified here removes the two penultimate exons and creates a new stop codon in the last exon near the transcript's 3'end. We, therefore, do not expect this mechanism to be active here. Out of the 4 commercially available antibodies we observed that only one provided a specific staining (Table S3). This antibody has been raised against a 101 amino acid epitope overlapping with the patient's deletion. We observed that, although the staining was weaker than in control sperm, the patient's short and abnormal flagella were immunodecorated along the flagella suggesting that the truncated protein is present in the deleted subjects and located to the flagella (Figure 2E). The intensity of the signal is patient's sperm is however very weak and in the absence of a negative control we cannot exclude the possibility that it could be due to non-specific staining and that the protein is completely absent in sperm from deleted subjects.

Breakpoint characteristics and CNV mutational mechanisms

Then, to unequivocally confirm the exact size and position of the deletion, we performed a walking PCR approach. We validated a PCR protocol that amplified specifically a 950 bp fragment in deleted individuals using the primers WDR66-F located 5209 bp upstream of exon 20 and WDR66-R positioned 1597 bp downstream of exon 21 (Figure 2A). As could be expected the 8541 bp interval in controls individuals could not be amplified and yielded no signal whereas a band of the expected size was obtained from all deleted individuals (Figure 2B). Sanger analysis of the PCR sequence covering the breakpoint region in the deleted individuals allowed the characterization of the breakpoint at the single-nucleotide resolution (Figure 2C). We observed that the deletion spans 7591 bp with exons 20-21 of the WDR66 gene completely deleted (chr12:122,432,883-122,440,629). All deleted individuals carried the exact same deletion with the same nucleic acid sequence at the breakpoint locus suggesting that they had all inherited the same deleted allele from a common ancestor. The loss of exons 20 and 21 is predicted to cause a frameshift leading to a premature stop codon at the beginning of the last exon 22. Therefore, the 147 amino acids encompassing the C-terminus region of the protein is predicted to be lost (=12.8% of the full-length protein).

To estimate the allelic frequency of this variant and to confirm that it is not a frequent CNV in the North African population we carried out a multiplex PCR allowing the co-amplification of the deleted allele (950 bp of product size) with the normal allele (220 bp) thus including a positive control in all reactions. A total of 100 North-African individuals were analyzed (Figure S3). None of the tested controls amplified the deletion indicating that the variant is also rare within the North-African population with an allelic frequency < 1/200. We then re-analyzed the whole data from WES analysis (SNPs and indels) of the 78 individuals but we did not identify any other *WDR66* deleterious variants.

To investigate the origin of this complex rearrangement, we carried out a series of bio-informatic analyses focusing on the genomic region surrounding the deletion. Clustering the deleted sequence using repeatmasker software (http://www.repeatmasker.org) indicated that the deleted region is rich in transposable elements and prone to genomic rearrangements. In addition, the centromeric breakpoint of the identified deletion was localized at the exact ending of an SVA_F retrotransposable element. Intriguingly, detailed sequence analysis revealed that in the deleted subjects, the telomeric fragment of this element (122, 432,706-122,432-883) did not perfectly match the expected SVA-F sequence (Figure S4). Alignment of this sequence

using BLAT showed 100% similarity to a distal sequence of an SVA-D element (Figure S4) located in intron 5 of *CLIP1*, a gene located 400 kb telomeric from *WDR66* (Figure 3A). In non-deleted control individuals, the abnormal telomeric end of *WDR66* SVA_F sequence perfectly matched the sequence of reference (GRCh37/hg19) indicating that the short SVA_D sequence observed only in deleted subjects was specific to the deletion event and came from *CLIP1* SVA_D retrotransposable element. We conclude that SVA-D was likely retrotransposed in intron 21 of *WDR66* and initiated a non-allelic homologous recombination (NAHR) with the SVA-F element in intron 19, inducing the formation of a single chimeric SVA-F/SVA-D sequence, which preceded the *WDR66* deletion (Figure 3B-D). This indicates that *WDR66*-SVA-F recombined at the following localization: 122,432,691-122,432,706. Such a mechanism has already been observed and has been described as a SVA-insertion mediated deletion [19].

Dating of the identified mutation

To compute the age of the deletion, we use the size of the shared ancestral haplotypes surrounding the *WDR66* deletion locus (Table S6) that is equal to 2.60 cM when assuming a recombination rate of 1.30 cM/Mb [31]. The probability to observe an ancestral haplotype of size c (measured in Morgans) or larger around the *WDR66* deletion is equal to $(1-c)^g$ where g is the age of the deletion measured in generations [32]. The probability that the 2n haplotypes carried by the 7 individual is larger than c is equal to $(1-c)^{2ng}$, which can be approximated by . Therefore, the length of the shared ancestral haplotype can be approximated by e^{-2ngc} an exponential distribution rate of $2ng$. The parameter g can be estimated by $\frac{1}{2nL}$, which provides an estimate of $g=27$ generations. However, Labuda et al. have shown that dating arising from genetic clock methods is biased downwards because they do not account for population expansion [33]. To account for an exponential expansion of rate r per generation, the mathematical expression: $\frac{1}{r}\ln\left(\frac{c.e^r}{e^r-1}\right)$ should be added to the estimates provided by genetic clock methods. Assuming that an upper bound for the expansion rate is given by the expansion $r = 0.4$, as was calculated for the Ashkenazy Jewish populations [34], we found that the age of the mutation is in the range of 27–43 generations. Taking 25 years for the generation time, the mutation would have then arisen 675–1075 years ago. Heterozygous individuals have apparently not been negatively selected and the deleted allele now seems relatively frequent in Tunisia as 7 out of the 46 Tunisian MMAF subjects analyzed here (15%) had a homozygous *WDR66* deletion.

Phenotypic analysis of MMAF individuals carrying genomic deletion in WDR66

The analysis of semen samples from the seven individuals carrying the same mutation showed normal semen volume and concentration although with large inter-individual variations (Table 1). A total absence of sperm motility was observed for all individuals carrying the *WDR66* deletion (Table 1). Sperm morphology appeared seriously altered under light microscopy examination with 100% of abnormal spermatozoa as illustrated in Figure 1. The most common morphological defects observed in the semen of all individuals concerned the flagellum dimensions (reduced length and irregular caliber). A high rate of acrosomal abnormalities was also observed in many deleted subjects affecting 56 to 95 % of spermatozoa. We compared the sperm parameters of individuals with different genotypes (mutated for *DNAH1*, *CFAP43, CFAP44* and *WDR66*), however, no significant difference was observed among the different groups (Table S8).

Due to an insufficient number of sperm cells collected from mutated individuals, only sperm cells from one individual carrying the WDR66 deletion could be analyzed by transmission electron microscopy (TEM) (Figure 1). Observation of longitudinal sections showed severe axonemal and peri-axonemal defects affecting the outer dense fibers (ODF), the fibrous sheath (FS) and the mitochondrial sheath (MS), which appeared completely disorganized resulting in aborted flagella or their replacement by a cytoplasmic mass englobing unassembled axonemal component (not shown). Due to the oligo-teratozoospermia observed in this individual, only a few cross-sections (<10) presented a sufficient quality to be analyzed. Among these sections, 100% were abnormal and the main defect constantly observed concerned the axonemal and the peri-axonemal structures such as unassembled FS (Figure 1F).

Functional study of the *Trypanosoma* WDR66 ortholog

To better characterize WDR66 localization and function we decided to use a tractable model organism, which could be used for forward and reverse genetics. We choose *Trypanosoma brucei* (*T. brucei*) because this flagellated, pathogenic, protozoan is a common model to study the function of axonemal proteins and has contributed to elucidating the molecular pathogeny of human ciliopathies [35] and more recently of MMAF syndrome [6].

BLASTp analysis on *T. brucei* genome database [36] using the human WDR66 sequence identified the *T. brucei* ortholog Tb927.3.1670 (named TbWDR66 in this study). TbWDR66 is a 1027 amino acids long protein. The sequence identity calculated with Clustal omega [37] between the 2 proteins is 30.55 %. Furthermore, alignment of the secondary structures of the human and the *T. brucei* protein shows a near perfect match (Figure S5 and Table S7) suggesting that the different protein domains (including the C-terminal domain containing an EF-hand motif) have preserved their functionality throughout evolution. Moreover, previous functional genomics and proteomic studies identified TbWDR66 as a flagellar protein [38,39]. In addition, TbWDR66 is the *Tetrahymena* and *Chlamydomonas* CFAP251 ortholog that were shown to play a role in building the radial spoke 3 in cilia [40].

Because the localization and function of WDR66 in the trypanosome flagellum is currently unknown, we localized TbWDR66 in bloodstream form (BSF, present in the mammalian host) and procyclic form (PCF, present in the insect host) *T. brucei*, using 10xTY1-tagged protein by generating *T. brucei* cell lines expressing endogenous levels of C-terminal tagged TbWDR66 ($TbWDR66_{TY1}$) [24,41]. $TbWDR66_{TY1}$ was found in the axoneme in BSF as substantiated by double labelling with antibodies against the paraflagellar rod structure (PFR) and TY1 (Figure 5A). We also used stimulated emission depletion microscopy (STED) and triple PFR-tubulin-TY1 labelling, which confirmed the axoneme localization in the PCF (Figure S6). $TbWDR66_{TY1}$ signal extended throughout the whole flagellum length, as demonstrated by the TY1 labelling preceding the PFR at the proximal end of the flagellum and along the flagellum up to its distal end (Figure 5A, white boxes). We also generated BSF and PCF cell lines expressing TbWDR66 deleted of its C-terminal domain (deleted aa 907-1027, TbWDR66-$\Delta Cter_{TY1}$), corresponding to the deletion of exons 20-22 in the human protein, and showed that the absence of the C-ter domain did not affect protein localization (Figure 5A and Figure S6) similar to our observations in the sperm of patients with the WDR66 C-terminal deletion.

To evaluate the impact of TbWDR66 knockdown on flagellum structure and function, we generated a BSF cell line expressing $TbWDR66_{TY1}$ and inducible for *TbWDR66* by RNA interference (RNAi) targeting the sequence encoding the C-terminal domain (Figure S1). Although RNAi induction did not fully deplete $WDR66_{TY1}$ levels as shown by immunofluorescence and western-blot detection of the tagged protein (Figure 5AB) it did lead to a decrease in cell proliferation compared to WT or to non-induced cultures (Figure 5C). More importantly, RNAi induction strongly affected flagellum motility as shown by video-

microscopy (Video S1 for WT to be compared to Video S2 for cells expressing *Tb*WDR66$_{TY1}$ and RNAi induced for 48 h). Anecdotally but notably, we have observed by electron microscopy some flagella sections (2%, n>149) showing a disorganization of the axoneme with displaced DMTs, a phenotype not observed in the WT cell control (Figure 5D). Together, these data indicate that TbWDR66 plays an important role in flagellar motility and is also likely to be involved in axoneme structure, two functions that are in agreement with what has been described for CFAP251 in *Tetrahymena* [40] .

To assess the role of the TbWDR66 C-terminal domain in protein function, we generated a cell line expressing TbWDR66-ΔCter$_{TY1}$ and inducible for the RNAi targeting only the endogenous *TbWDR66* (Figure S1B). Cell proliferation was not affected by the expression of TbWDR66-ΔCter$_{TY1}$ in RNAi non-induced cells (Figure 5C). As expected, expression and localization of TbWDR66-ΔCter$_{TY1}$ was not affected by the RNAi as shown by immunofluorescence and western-blot analyses (Figure 5AB). However, flagellum motility was impaired similarly to the phenotypes induced by RNAi in the WDR66$_{TY1}$ background cell line (Video S3) demonstrating that expression of TbWDR66 deleted of its C-terminal domain cannot complement for the down-regulation of the full-length TbWDR66 protein. Further, cell proliferation decrease was slightly stronger than in the WDR66$_{TY1}$ background cell line suggesting that, in the context of a protein complex, presence of a non-functional protein might be more deleterious than absence of the protein (Figure 5C). Taken together, these data demonstrate that TbWDR66 is involved in flagellum motility and that its C-terminal domain is required for protein function.

DISCUSSION

WDR66 is a new gene involved human male infertility accounting for 9 % of MMAF cases.

In the present study, we analyzed 78 MMAF individuals combining SNPs/indels with CNV analysis on WES data and identified a homozygous intragenic deletion encompassing exons 20 and 21 of the *WDR66* gene, which encodes a WD-repeat protein. Interestingly, we showed previously that two other WDR proteins (CFAP43 and CFAP44) are responsible for MMAF phenotype and account for 20.5% of the same cohort [6]. We report in this study the implication of a new WD-repeat protein in 15% of MMAF subjects from Tunisia indicating that *WDR66* is another important gene for this phenotype. This suggests that WDR proteins are enriched in axonemal and periaxonemal structures. Interestingly, we observed the exact same deletion in all seven individuals, all were of North African origin indicating a founder effect. Study of the SNPs present around the deleted regions confirmed this and showed that all deleted subjects shared an ancestral haplotype of 2.60 cM suggesting that their common ancestor lived 675–1075 years ago.

Interest of WES to detect pathological CNVs and importance of SVA retrotransposons elements.

Copy number variants (CNVs) are a common cause of disease but they had not been explored from whole exome sequencing (WES) data in the field of male infertility and particularly in MMAF phenotype. In the current study, we analyzed WES data using the ExomeDepth software, which is based on the assessment of sequencing read depth. We analyzed the breakpoints to elucidate the mechanisms underlying the deletion. Results showed a high density of *Alu* retrotransposons in the region flanking the deletion breakpoints suggesting that this region exhibits a high degree of genomic instability. Interestingly, we observed a single chimeric SVA element located precisely adjacent to the centromeric breakpoint strongly suggesting that the creation of this abnormal SVA and the deletion event are the consequence of the same non-allelic homologous recombination (NAHR) between SVA_F in intron 19 of *WDR66* and the SVA_D retrotransposed from intron 5 of *CLIP1* to intron *21* of *WDR66.* This event resulted in the deletion of a portion of each SVA element and of the intervening DNA sequences (Figure 3). This mechanism was previously described as SVA insertion-mediated deletion (SIMD) [19]. SVAs, originally named SINE-R, are composite transposons which make

up 0.2% of the human genome and represent the youngest and currently the most active retrotransposable elements (RE) of the human genome [42]. SVAs are mobilized by L1 enzymatic machinery throughout the genome using a "copy-and-paste" mechanism [43]. It has been shown that retrotransposition of SVA elements is a source of insertions, deletions and rearrangements within the target site leading to various pathologies [44]. The spectrum of diseases that have been reported related to SVA transposition is increasing progressively including Marfan syndrome, hemophilia B, neurofibromatosis type 1 and 2, leukemia and congenital muscular dystrophy [45–50]. In this study, we report for the first time, a homozygous CNV generated by an SVA transposition event causing multiple morphological abnormalities of the sperm flagellum and leading to male infertility.

WDR66 function and effect of the identified deletion

The axoneme, the internal cytoskeleton of motile cilia and flagella, is a highly evolutionary conserved structure which consists of nine doublet microtubules (DMTs) circularly arranged around the central pair complex of microtubules (CPC) ('9+ 2' structure) [51]. The beating of cilia and flagella is orchestrated by multiprotein-ATPase complexes that are located on the peripheral doublets, which provide the sliding force for sperm motility. In addition, the sperm flagellum harbors specific peri-axonemal structures, which are not found in other motile cilia, a helical mitochondrial sheath (MS) in the midpiece, the fibrous sheath (FS) in the principal piece (PP) and outer dense fibers (ODF) in the midpiece and the proximal part of the PP [52].

Previous reports showed that *WDR66* ortholog (*FAP251*) in *Chlamydomonas* localizes to the CSC (Calmoduline and Spoke associated Complex) at the base of the radial spoke 3 (RS3) [53]. The CSC was recently identified as an axonemal complex that associates the bases of RS2 and 3 and the N-DRC (Nexin-Dynein Regulatory Complex) and is involved in the modulation of the dynein activity and therefore in the regulation of flagellar motility[53]. Recent work in *Tetrahymena* showed that, in the absence of CFAP251, RS3 was either missing or incomplete and the protozoa had an altered swimming ability with abnormal ciliary waveform [40]. Here we studied the localization of *WDR66* ortholog in *Trypanosoma brucei* and used this model to confirm the deleterious effect of the identified deletion. Immunofluorescence and STED microscopy confirmed that TbWDR66 is an axonemal protein and confirmed that in *Trypanosoma,* knockdown by RNAi affected flagellar structure and motility. We then generated

cell lines with a deletion of TbWDR66 C-terminal domain corresponding to the deletion identified in MMAF individuals. We observed that the absence of the C-ter domain did not affect the protein localization (Figure 5A and Figure S6) but severely affected the cells mobility, as was observed on the sperm of deleted individuals. This work confirmed that the WDR66 Cter domain, which contains a calcium-regulating EF-hand domain is critical to maintaining flagella function by probably regulating flagellar beating but we also observed that WDR66 also plays a critical structural role, perhaps by consolidating the links between the RS3s and the microtubule doublets.

Impact of the MMAF genes on flagellar structure and ultrastructure in mammals and Trypanosoma

The 7 subjects with a partial deletion of *WDR66* described here present a classical MMAF syndrome with the typical hallmarks of the phenotype: a near absence of motility (3%), a high proportion of short (39%), absent (27%), coiled (21%) flagella and of flagella of irregular caliber (43%) as described in table 1 and highlighted in Figure 1 A-D). These parameters are similar to those found previously in patients with mutations in other MMAF genes such as *DNAH1*, *CFAP43* and *CFAP44* [6], confirming the relative homogeneity of the phenotype and underlining its genetic heterogeneity. Previous ultrastructural analyses of the sperm flagella of subjects mutated for the three initial genes highlighted a severe disorganization of the axoneme affecting in particular the central pair microtubules but also the other doublets and even the extra-axonemal structures. For instance, we have shown that the absence of DNAH1, a protein located in the inner dynein arm facing RS3 (and thus a putative partner of WDR66) induces a strong disorganization of the axonemal structure. Similarly, the absence of CFAP43 and 44, two proteins playing a role in the interaction between the axoneme and periaxonemal structures leads to a strong disorganization of the axonemal structure in human and mouse. Here we showed that the absence of a functional WDR66 protein also induces a spectacular disorganization of the flagella ultrastructure (Figure 1 E,F). Interestingly, we observed that the level of flagellar disorganization is much more pronounced in man and mouse sperm than in flagellate protozoan depleted of the MMAF orthologues (as illustrated by the study of CAFAP43, CFAP44 and here of WDR66 orthologues in *Trypanosoma*). In the studied mammals, the absence of these axonemal proteins seems to have a profound destabilizing effect during flagellar growth inducing the different morphological anomalies described above which

even include malformations of the fibrous sheath and other peri-axonemal structures. The absence of structural defects in *Trypanosoma* may indicate that axonemal growth and intra-flagellar transport (IFT) is more resilient in the parasite, suggesting that sperm and parasite's flagellum IFT are different. Moreover, in Tb the flagella is attached to the body of the parasite and we can speculate that the axoneme is not subjected to the same constraints during flagellar formation thus perhaps also explaining the absence of marked ultrastructural alterations.

Overall we showed that the deletion of the WDR66 C-terminal domain is a common genetic cause of male infertility in Tunisian men due to MMAF. This finding highlights the importance of this domain and of the WDR66 protein for flagellar structure, assembly and beating. We confirmed that WES is efficient to identify small CNV events (here, a deletion of two exons) and could at term replace or at least complement conventional microarray analysis. Importantly, we showed, as previously observed for CFAP43 and CFAP44, the high level of conservation of flagellar proteins between humans and *Trypanosoma brucei*, confirming that this organism is an excellent model to investigate the molecular function and pathogeny of genes associated with flagellar dysfunction in humans.

ACKNOWLEDGMENTS

We thank our patients and control individuals for their participation. This work was mainly supported by the following grants: The "MAS-Flagella" project financed by French ANR and the DGOS for the program PRTS 2014 (to PFR, MB, AT, CA, NTM), the "Whole genome sequencing of individuals with Flagellar Growth Defects (FGD)" financed by the fondation maladies rares (FMR) for the program Séquençage à haut débit 2012 (to PFR) and the ParaFrap Grant financed by the ANR LabEx (ANR-11-LABX-0024) to DRR (http://www.labex-parafrap.fr/fr). The STED microscopy was done in the Bordeaux Imaging Centre, a service unit of the CNRS-INSERM and Bordeaux University, member of the national infrastructure France BioImaging supported by the French National Research Agency (ANR-10-INBS-04). The help of Patrice Mascalchi is acknowledged. Some of the electron microscopy was done at the Bordeaux Imaging Centre. We thank Nicolas Biteau (Bordeaux University) for the rabbit anti-PFR2 antibody, Sam Dean (Oxford University) for the pPOTv7 plasmid, Keith Gull (Oxford University) for the anti-tubulin (TAT1) antibody, Philippe Bastin (Institut Pasteur, Paris) for the anti-TY1 antibody. We thank also the GIN electron microscopy platform and Anne Bertrand, and the IAB microscopy platform and Alexei Grichine and Jacques Mazzega for their technical help.

AUTHORS' ROLES

Z-EK, MB, CA and PFR analyzed the data and wrote the manuscript; Z-EK, AA-Y and CC performed molecular work; TK, NT-M, PFR analyzed the genetic data; Z-EK, MC-K, CC, CC, GM, PLT performed IF experiments; Z-EK and SC performed protein *in silico* analysis, Z-EK, GM, CC and KP-G performed EM experiments, DD, DRR, NL and MB performed *Trypanosoma* work, SFBM, LH, OM, MM, HL, MK, HG, RZ provided clinical samples and data; MB and PFR designed the study, supervised all molecular laboratory work, had full access to all of the data in the study and take responsibility for the integrity of the data and its accuracy. All authors contributed to the report.

CONFLICT OF INTEREST STATEMENT

The authors have declared that no conflict of interest exists.

REFERENCES

1. Neto, F.T.L., Bach, P.V., Najari, B.B., Li, P.S., and Goldstein, M. (2016). Spermatogenesis in humans and its affecting factors. Semin. Cell Dev. Biol. *59*, 10–26.

2. Jan, S.Z., Vormer, T.L., Jongejan, A., Röling, M., Silber, S.J., de Rooij, D.G., Hamer, G., Repping, S., and van Pelt, A.M.M. (2017). Unraveling transcriptome dynamics in human spermatogenesis. Dev. Camb. Engl.

3. Coutton, C., Escoffier, J., Martinez, G., Arnoult, C., and Ray, P.F. (2015). Teratozoospermia: spotlight on the main genetic actors in the human. Hum. Reprod. Update *21*, 455–485.

4. Ray, P.F., Toure, A., Metzler-Guillemain, C., Mitchell, M.J., Arnoult, C., and Coutton, C. (2017). Genetic abnormalities leading to qualitative defects of sperm morphology or function. Clin. Genet. *91*, 217–232.

5. 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.

6. Coutton, C., Vargas, A.S., Amiri-Yekta, A., Kherraf, Z.-E., Ben Mustapha, S.F., Le Tanno, P., Wambergue-Legrand, C., Karaouzène, T., Martinez, G., Crouzy, S., et al. (2018). Mutations in CFAP43

and CFAP44 cause male infertility and flagellum defects in Trypanosoma and human. Nat. Commun. *9*,.

7. Tang, S., Wang, X., Li, W., Yang, X., Li, Z., Liu, W., Li, C., Zhu, Z., Wang, L., Wang, J., et al. (2017). Biallelic Mutations in CFAP43 and CFAP44 Cause Male Infertility with Multiple Morphological Abnormalities of the Sperm Flagella. Am. J. Hum. Genet. *100*, 854–864.

8. Yan-Wei Sha, Xiong Wang, Xiaohui Xu, Zhi-Ying Su, Yuanqing Cui, Li-Bin Mei, Xian-Jing Huang, Jie Chen, Xue-Mei He, Zhi-Yong Ji, et al. (2017). Novel Mutations in CFAP44 and CFAP43 Cause Multiple Morphological Abnormalities of the Sperm Flagella (MMAF). Reprod. Sci. 1933719117749756.

9. Sha, Y.-W., Xu, X., Mei, L.-B., Li, P., Su, Z.-Y., He, X.-Q., and Li, L. (2017). A homozygous CEP135 mutation is associated with multiple morphological abnormalities of the sperm flagella (MMAF). Gene *633*, 48–53.

10. Lorès, P., Coutton, C., El Khouri, E., Stouvenel, L., Givelet, M., Thomas, L., Rode, B., Schmitt, A., Louis, B., Sakheli, Z., et al. (2018). Homozygous missense mutation L673P in adenylate kinase 7 (AK7) leads to primary male infertility and multiple morphological anomalies of the flagella but not to primary ciliary dyskinesia. Hum. Mol. Genet.

11. Feuk, L., Carson, A.R., and Scherer, S.W. (2006). Structural variation in the human genome. Nat. Rev. Genet. *7*, 85–97.

12. Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W., and Lee, C. (2004). Detection of large-scale variation in the human genome. Nat. Genet. *36*, 949–951.

13. Mills, R.E., Walter, K., Stewart, C., Handsaker, R.E., Chen, K., Alkan, C., Abyzov, A., Yoon, S.C., Ye, K., Cheetham, R.K., et al. (2011). Mapping copy number variation by population-scale genome sequencing. Nature *470*, 59–65.

14. Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Månér, S., Massa, H., Walker, M., Chi, M., et al. (2004). Large-scale copy number polymorphism in the human genome. Science *305*, 525–528.

15. Almal, S.H., and Padh, H. (2012). Implications of gene copy-number variation in health and diseases. J. Hum. Genet. *57*, 6.

16. Ma, D., Yang, J., Wang, Y., Huang, X., Du, G., and Zhou, L. (2017). Whole exome sequencing identified genetic variations in Chinese hemangioblastoma patients. Am. J. Med. Genet. A. *173*, 2605–2613.

17. Tsuchida, N., Nakashima, M., Kato, M., Heyman, E., Inui, T., Haginoya, K., Watanabe, S., Chiyonobu, T., Morimoto, M., Ohta, M., et al. (2017). Detection of copy number variations in epilepsy using exome data. Clin. Genet.

18. Plagnol, V., Curtis, J., Epstein, M., Mok, K.Y., Stebbings, E., Grigoriadou, S., Wood, N.W., Hambleton, S., Burns, S.O., Thrasher, A.J., et al. (2012). A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. Bioinforma. Oxf. Engl. *28*, 2747–2754.

19. Lee, J., Ha, J., Son, S.-Y., and Han, K. (2012). Human Genomic Deletions Generated by SVA-Associated Events. Comp. Funct. Genomics *2012*, 807270.

20. Jungwirth, A., Giwercman, A., Tournaye, H., Diemer, T., Kopa, Z., Dohle, G., Krausz, C., and European Association of Urology Working Group on Male Infertility (2012). European Association of Urology guidelines on Male Infertility: the 2012 update. Eur. Urol. *62*, 324–332.

21. Wirtz, E., Leal, S., Ochatt, C., and Cross, G.A. (1999). A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in Trypanosoma brucei. Mol. Biochem. Parasitol. *99*, 89–101.

22. Brun, R., and Schönenberger, null (1979). Cultivation and in vitro cloning or procyclic culture forms of Trypanosoma brucei in a semi-defined medium. Short communication. Acta Trop. *36*, 289–292.

23. Schumann Burkard, G., Jutzi, P., and Roditi, I. (2011). Genome-wide RNAi screens in bloodstream form trypanosomes identify drug transporters. Mol. Biochem. Parasitol. *175*, 91–94.

24. Dean, S., Sunter, J., Wheeler, R.J., Hodkinson, I., Gluenz, E., and Gull, K. (2015). A toolkit enabling efficient, scalable and reproducible gene tagging in trypanosomatids. Open Biol. *5*, 140197.

25. LaCount, D.J., Barrett, B., and Donelson, J.E. (2002). Trypanosoma brucei FLA1 is required for flagellum attachment and cytokinesis. J. Biol. Chem. *277*, 17580–17588.

26. Oberholzer, M., Lopez, M.A., Ralston, K.S., and Hill, K.L. (2009). Approaches for functional analysis of flagellar proteins in African trypanosomes. Methods Cell Biol. *93*, 21–57.

27. Bragin, E., Chatzimichali, E.A., Wright, C.F., Hurles, M.E., Firth, H.V., Bevan, A.P., and Swaminathan, G.J. (2014). DECIPHER: database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. Nucleic Acids Res. *42*, D993.

28. MacDonald, J.R., Ziman, R., Yuen, R.K.C., Feuk, L., and Scherer, S.W. (2014). The Database of Genomic Variants: a curated collection of structural variation in the human genome. Nucleic Acids Res. *42*, D986-992.

29. Qiu, F., Xu, Y., Li, K., Li, Z., Liu, Y., DuanMu, H., Zhang, S., Li, Z., Chang, Z., Zhou, Y., et al. (2012). CNVD: text mining-based copy number variation in disease database. Hum. Mutat. *33*, E2375-2381.

30. Schweingruber, C., Rufener, S.C., Zünd, D., Yamashita, A., and Mühlemann, O. (2013). Nonsense-mediated mRNA decay — Mechanisms of substrate mRNA recognition and degradation in mammalian cells. Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1829*, 612–623.

31. Jensen-Seaman, M.I., Furey, T.S., Payseur, B.A., Lu, Y., Roskin, K.M., Chen, C.-F., Thomas, M.A., Haussler, D., and Jacob, H.J. (2004). Comparative recombination rates in the rat, mouse, and human genomes. Genome Res. *14*, 528–538.

32. Slatkin, M., and Rannala, B. (2000). Estimating allele age. Annu. Rev. Genomics Hum. Genet. *1*, 225–249.

33. Labuda, M., Labuda, D., Korab-Laskowska, M., Cole, D.E., Zietkiewicz, E., Weissenbach, J., Popowska, E., Pronicka, E., Root, A.W., and Glorieux, F.H. (1996). Linkage disequilibrium analysis in young populations: pseudo-vitamin D-deficiency rickets and the founder effect in French Canadians. Am. J. Hum. Genet. *59*, 633–643.

34. Risch, N., de Leon, D., Ozelius, L., Kramer, P., Almasy, L., Singer, B., Fahn, S., Breakefield, X., and Bressman, S. (1995). Genetic analysis of idiopathic torsion dystonia in Ashkenazi Jews and their recent descent from a small founder population. Nat. Genet. *9*, 152–159.

35. Vincensini, L., Blisnick, T., and Bastin, P. (2011). 1001 model organisms to study cilia and flagella. Biol. Cell *103*, 109–130.

36. Aslett, M., Aurrecoechea, C., Berriman, M., Brestelli, J., Brunk, B.P., Carrington, M., Depledge, D.P., Fischer, S., Gajria, B., Gao, X., et al. (2010). TriTrypDB: a functional genomic resource for the Trypanosomatidae. Nucleic Acids Res. *38*, D457-462.

37. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539.

38. Broadhead, R., Dawe, H.R., Farr, H., Griffiths, S., Hart, S.R., Portman, N., Shaw, M.K., Ginger, M.L., Gaskell, S.J., McKean, P.G., et al. (2006). Flagellar motility is required for the viability of the bloodstream trypanosome. Nature *440*, 224–227.

39. Subota, I., Julkowska, D., Vincensini, L., Reeg, N., Buisson, J., Blisnick, T., Huet, D., Perrot, S., Santi-Rocca, J., Duchateau, M., et al. (2014). Proteomic analysis of intact flagella of procyclic Trypanosoma brucei cells identifies novel flagellar proteins with unique sub-localization and dynamics. Mol. Cell. Proteomics MCP *13*, 1769–1786.

40. Urbanska, P., Song, K., Joachimiak, E., Krzemien-Ojak, L., Koprowski, P., Hennessey, T., Jerka-Dziadosz, M., Fabczak, H., Gaertig, J., Nicastro, D., et al. (2015). The CSC proteins FAP61 and FAP251 build the basal substructures of radial spoke 3 in cilia. Mol. Biol. Cell *26*, 1463–1475.

41. Bastin, P., Bagherzadeh, Z., Matthews, K.R., and Gull, K. (1996). A novel epitope tag system to study protein targeting and organelle biogenesis in Trypanosoma brucei. Mol. Biochem. Parasitol. *77*, 235–239.

42. Cordaux, R., and Batzer, M.A. (2009). The impact of retrotransposons on human genome evolution. Nat. Rev. Genet. *10*, 691–703.

43. Raiz, J., Damert, A., Chira, S., Held, U., Klawitter, S., Hamdorf, M., Löwer, J., Strätling, W.H., Löwer, R., and Schumann, G.G. (2012). The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. Nucleic Acids Res. *40*, 1666–1683.

44. Hancks, D.C., and Kazazian, H.H. (2016). Roles for retrotransposon insertions in human disease. Mob. DNA *7*, 9.

45. Brett, M., Korovesis, G., Lai, A.H.M., Lim, E.C.P., and Tan, E.-C. (2017). Intragenic multi-exon deletion in the FBN1 gene in a child with mildly dilated aortic sinus: a retrotransposal event. J. Hum. Genet. *62*, 711–715.

46. Legoix, P., Sarkissian, H.D., Cazes, L., Giraud, S., Sor, F., Rouleau, G.A., Lenoir, G., Thomas, G., and Zucman-Rossi, J. (2000). Molecular characterization of germline NF2 gene rearrangements. Genomics *65*, 62–66.

47. Nakamura, Y., Murata, M., Takagi, Y., Kozuka, T., Nakata, Y., Hasebe, R., Takagi, A., Kitazawa, J., Shima, M., and Kojima, T. (2015). SVA retrotransposition in exon 6 of the coagulation factor IX gene causing severe hemophilia B. Int. J. Hematol. *102*, 134–139.

48. Takasu, M., Hayashi, R., Maruya, E., Ota, M., Imura, K., Kougo, K., Kobayashi, C., Saji, H., Ishikawa, Y., Asai, T., et al. (2007). Deletion of entire HLA-A gene accompanied by an insertion of a retrotransposon. Tissue Antigens *70*, 144–150.

49. Vogt, J., Bengesser, K., Claes, K.B.M., Wimmer, K., Mautner, V.-F., van Minkelen, R., Legius, E., Brems, H., Upadhyaya, M., Högel, J., et al. (2014). SVA retrotransposon insertion-associated deletion represents a novel mutational mechanism underlying large genomic copy number changes with non-recurrent breakpoints. Genome Biol. *15*, R80.

50. Xiong, H., Wang, S., Kobayashi, K., Jiang, Y., Wang, J., Chang, X., Yuan, Y., Liu, J., Toda, T., Fukuyama, Y., et al. (2009). Fukutin gene retrotransposal insertion in a non-Japanese Fukuyama congenital muscular dystrophy (FCMD) patient. Am. J. Med. Genet. A. *149A*, 2403–2408.

51. Ishikawa, T. (2017). Axoneme Structure from Motile Cilia. Cold Spring Harb. Perspect. Biol. *9*, a028076.

52. Lindemann, C.B., and Lesich, K.A. (2016). Functional anatomy of the mammalian sperm flagellum. Cytoskelet. Hoboken NJ *73*, 652–669.

53. Heuser, T., Dymek, E.E., Lin, J., Smith, E.F., and Nicastro, D. (2012). The CSC connects three major axonemal complexes involved in dynein regulation. Mol. Biol. Cell *23*, 3143–3155.

Table 1. Semen parameters of the 7 deleted MMAF subjects.

Patient Identification	P1	P2	P3	P4	P5	P6	P7	Mean ± SD	Lower Reference Limits (WHO 2010)
Geographical origin	Tunisia	Tunisia	Tunisia	Tunisia	Tunisia	Tunisia	Tunisia	-	-
Consanguinity	Yes	Yes	Yes	Yes	No	Yes	Yes	-	-
Age	45	54	51	45	31	28	59	44.7 ± 10.6 (n'=7)	-
Sperm volume (mL)	4	1.8	1.7	2.2	3.4	3	5	3.0 ± 1.1 (n'=7)	1.5
Sperm concentration (million/mL)	11.3	6	56	56	60	7,8	15.5	30.3 ± 23.5 (n'=7)	15
Total motility 1 h (%)	2	1	5	0	5	5	4	3.1 ± 1.9 (n'=7)	40
Vitality (%)	53	80	73	79	62	NC	47	65.6 ± 12.6 (n'=6)	58
Normal spermatozoa (%)	0	0	0	0	0	0	0	0 (n'=7)	4
Abnormal acrosome	66	56	56	78	80	NC	95	71.8 ± 15.3 (n'=6)	-
Abnormal head morphology	42	10	20	58	8	NC	7	24.1 ± 25.1 (n'=6)	-
Abnormal base	36	2	32	14	2	NC	3	14.8 ± 15.6 (n'=6)	-
Absent flagella (%)	31	6	22	50	24	NC	30.5	27.2 ± 13.1 (n'=6)	-
Short flagella (%)	34	32	59	32	50	NC	25	38.6 ± 11.8 (n'=6)	-
Coiled flagella (%)	20	17	27	34	8	NC	27	21.1 ± 8.3 (n'=6)	-
Flagella of irregular caliber (%)	4	68	37	50	68	NC	33	43.3 ± 22.1 (n'=6)	-
Multiple anomalies index	2.5	2.3	2.7	3.4	2.5	NC	2.56	2.6 ± 0.4 (n'=6)	-

LEGENDS TO FIGURES

Figure 1: Defective sperm flagellar formation in *WDR66* mutated individuals.
(A-D) Light microscopy analysis of control **(A)** and individual's sperm **(B-D)** reveals multiple morphological abnormalities of the sperm flagella in individuals. **(E-F)** TEM micrographs of cross-sections through sperm cell midpieces. **(E)** Control sperm shows a homogeneous ring-like mitochondrial sheath (MS) surrounding 9 outer dense fibers (ODF, arrowheads) and the central axoneme characterized by the typical presence of 9 microtubule doublets (MD, arrows) and the two central microtubules singlets (MS). **(F)** Representative sperm from an affected men showing a partitioned MS abnormally positioned and assembled ODF and MD and the absence of the MS. Scale bars, 10 µm **(A–D)** and 200 nm **(E–F)**.

Figure 2: Confirmation of the presence of the WDR66 deletion in patients and validation of its effect on mRNA and protein.
(A) Deleted individuals (P1 and P2) show a complete absence of sequence coverage of exons 20 and 21 as illustrated by IGV as opposed to controls (C1)
(B) In human, *WDR66* encodes for two transcripts: T1 (ENST00000288912.8, NM_144668), with 22 exons coding for a 1149 amino acid protein and T2 (ENST00000397454.2, NM_001178003) with 18 exons coding for a 941 amino acid protein. Only T1 is affected by the identified deletion (black double arrow).
(C) Relative expression ($2^{-\Delta\Delta Ct}$) of both transcripts (T1 and T2) performed on human panel of cDNAs from two tissues expressing 9+2 secondary motile axonemes (testis and lung) and two tissues expressing 9+0 primary immotile axonemes (kidney and lung). *GAPDH* was used as reference gene and T2 expression level in brain as reference sample. Data are expressed as mean ± SD.
(D) Representation of WDR66 protein encoded by T1. We note the presence of nine WD40 repeat domains and one calcium-binding EF-hand (α) in the C-terminus extremity. The first deleted amino acid (aa) is aa1003. Deletion of exons 20 and 21 induce a frameshift. Translation of exon 22 is then incorrect for the first 25 aa (orange line) before inducing a stop codon 26 nucleotides after the first abnormal aa. The protein encoded by the deleted T1 is therefore expected to stop at aa1028.
(E) Double immunofluorescence labelling of WDR66 and α-tubulin in human sperm cells: In control sperm, the yellow stain resulted from merged signals indicates the co-localization of both proteins within the flagellar axoneme. The images shown are representative of many cells examined in two independent experiments. Scale Bar: 10 µm.

Figure 3: Fine characterization of the deletion and theoretical description of the initial deletion event

(A) Representation in control subjects of part of the *WDR66* gene (between exon 19 and 22) and localization of the deletion-specific primers F1 and R1.

(B) Representation of the same region in deleted subjects. Detailed analysis of the deleted region within *WDR66* highlights the presence of a type F SINE-VNTR-Alu (SVA) retrotransposon (SVA-F) positioned at the centromeric border of the identified *WDR66* deletion (orange box).
(C) PCR of the deleted region. In controls (C1-C3), F1 and F2 primers are too distant (8541 bp) and do not yield any amplification. PCR amplification with F1 and F2 of patients (P1-P7) permits the amplification of a fragment of 950nt.
(D) Sanger sequencing through the deletion breakpoints permits to localize the exact position of the breakpoints, indicating that a total of 7591nt were deleted. It also highlights the presence of a 165nt inserted sequence (yellow) at the junction point.
(E) Blast analysis of the 165nt inserted sequence shows that an identical sequence (100% identity) is present within an SVA-D sequence located 400kb telomeric of SVA-F between exons 6 and 5 of the *CLIP1* gene (yellow box).
(F) SVA-D was likely retrotransposed between *WDR66* exons 21 and 22
(G) An intra-chromatidien non-allelic homologous recombination (NAHR) was then initiated
(H) Inducing the formation of a hybrid SVA-F/SVA-D sequence preceding the *WDR66* exons 20 and 21 deletions.

Figure 5: TbWDR66 is an axoneme protein and its C-terminal domain is required for function in BSF trypanosome.
(A) Control BSF cells (WT) or expressing TbWDR66$_{TY1}$ or TbWDR66-ΔCter$_{TY1}$ were immunolabelled with anti-PFR (white) and anti-TY1 (red) showing that anti-TY1 decorates the tagged proteins throughout the full-length of the flagellum. Induction of RNAi for 48 h (+RNAi) showed a decrease of fluorescence intensity of TbWDR66$_{TY1}$ labelling but not of TbWDR66-ΔCter$_{TY1}$ labelling. Nuclei and kinetoplasts (mitochondrial genomes) were labelled with DAPI (blue). Scale bar represents 5 μm. (B) Western-blot analysis of the expression of TbWDR66$_{TY1}$ and TbWDR66-ΔCter$_{TY1}$ without (NI) and with 24 h, 48 h, 72 h of RNAi induction shows a decrease of TbWDR66$_{TY1}$ (indicated by the % of labelling relative to the non-induced condition) expression, whilst expression of TbWDR66-ΔCter$_{TY1}$ is not affected. Tubulin labeling was used as loading control. (C) *TbWDR66* RNAi affects cell proliferation. Growth curves for WT cells, and RNAi non-induced or induced cells expressing TbWDR66$_{TY1}$ or TbWDR66-ΔCter$_{TY1}$. Cells were counted every 24 h. The graph represents the cumulative number of cells per ml. Error bars represent the standard error from 3 independent experiments. (D) Electron microscopy images of thin sections of a WT cell (a), and a cell expressing TbWDR66$_{TY1}$ (b) or TbWDR66-ΔCter$_{TY1}$ (c) and RNAi-induced for 48 h show that the decrease TbWDR66 expression induces a severe disorganization of the axoneme and of the paraxonemal structure, which is not rescued by the expression of TbWDR66-ΔCter$_{TY1}$. Scale bars represent 50 nm.

LEGENDS OF SUPPLEMENTARY MATERIAL

Figure S1. Schematic representation of the RNAi and endogenous tagging strategies.

A. One allele out of two is replaced by a PCR product allowing the insertion of the TY1 tag and a selection gene. The RNAi will target both alleles, and upon induction, both WT and TY1-tagged protein will be knocked down. **B.** One allele out of two is replaced by a PCR product allowing the shortening of *TbWDR66* ORF, the insertion of the C-terminal TY1 tag and a selection gene. The RNAi targets only the WT allele, and as a consequence, only the WT protein will be knocked down, whilst TbWDR66-$\Delta Cter_{TY1}$ expression is not affected.

Figure S2: Array-CGH profile of chromosome 12 showing a large homozygous deletion of *WDR66* gene in the patient P1.
(A) Each spot represents an oligonucleotide on the array. Spots are graphically represented according to their corresponding position along the chromosome. Values indicated on the top are expressed as log2 ratios of fluorescence intensity. We observe that three spots are deviated and have ratios between -2.358 and -4.621, indicating that the patient has a homozygous deletion in the *WDR66* gene.

Figure S3: gel electrophoresis illustrating the duplex PCR used to screen 100 North-African individuals for WDR66 deletion
The primer used permit to amplify a 220 nucleotide sequence located within the *WDR66* deletion hence present only in non-deleted alleles and a 950 nucleotide sequence encompassing the deletion breakpoint hence only present in the deleted alleles. In lane 1 (-/-) the amplified DNA was extracted from a non-deleted control, in lane 2 (+/+) from the homozygous deleted patient 1. Lane 3 (+/-) was obtained from mixing deleted and non-deleted DNAs. S1-S4 corresponds to the first 4 tested North African subjects. MW= molecular weight.

Figures S4: Alignment of the SVA-F/D sequence encompassing the centromeric breakpoint in patients with the SVA-F and SVA-D sequences in WDR66 (intron 19) and CLIP1 (intron 5) respectively.
Alignment was performed using CLC sequence viewer. D indicates the variants identical between the patients and the *CLIP1* SVA-D sequence, whereas F indicates the variants identical between the patient and the WDR66 SVA-F sequence. Results show that in patients, the telomeric part of the recombined SVA is homologous to the *CLIP1* SVA-D (yellow line) whereas the centromeric SVA sequence corresponds to the original*WDR66* SVA-F intron 19 (grey line). The 15 nucleotides indicated between the D and the F sequence (double black arrow) corresponds to the minimal region localizing the centromeric breakpoint. The telomeric breakpoint is indicated by a black vertical arrow. In deleted men, the sequence following the telomeric breakpoint corresponds to WDR66 intron 21.

Figure S5. Secondary structure prediction of human WDR66 and Trypanosoma TbWDR66.
Elements of the secondary structures of WDR66 and tbWDR66 were predicted by Porter 4.0. C, E and H correspond to coils, extended conformations (β-sheets) and helices (α-helices), respectively.

Figure S6. $TbWDR66_{TY1}$ and TbWDR66-$\Delta Cter_{TY1}$ are axoneme proteins in PCF *Trypanosoma*. A. Epifluorescence analysis of immunolabeled PCF cells expressing TbWDR66TY1 (a-d) and TbWDR66-$\Delta Cter_{TY1}$ (e-h) using anti-PFR2 (grey) that labels the paraflagellar rod and anti-TY1 (red) antibodies. Scale bar represents 5 μm. **B.** Deconvolved images of STED confocal microscopy experiments on PCF cells expressing $TbWDR66_{TY1}$ (a-d) or TbWDR66-$\Delta Cter_{TY1}$ (e-h) and immuno-labeled with anti-PFR2 (grey), anti-TY1 (red), and anti-tubulin (green) that labels the axonemal microtubules as well as the subpellicular microtubules (delimitated by the dotted white lines). Scale bar represents 0.5 μm.

Video S1. Illustration of mobility of WT cells showing active flagellar beating.

Video S2. Illustration of mobility of cells expressing TbWDR66$_{TY1}$ and RNAi-induced for 48 h. Flagellar beating is slower compared to WT cells, and the cells remain primarily in one location demonstrating a flagellum motility defect.

Video S3. Illustration of mobility of cells expressing TbWDR66-ΔCter$_{TY1}$ and RNAi-induced for 48 h.
Flagellar beating is slower compared to WT cells, and the cells remain primarily in one location demonstrating a flagellum motility defect demonstrating that TbWDR66-ΔCter$_{TY1}$ cannot complement for the down-regulation of the full-length TbWDR66 protein.

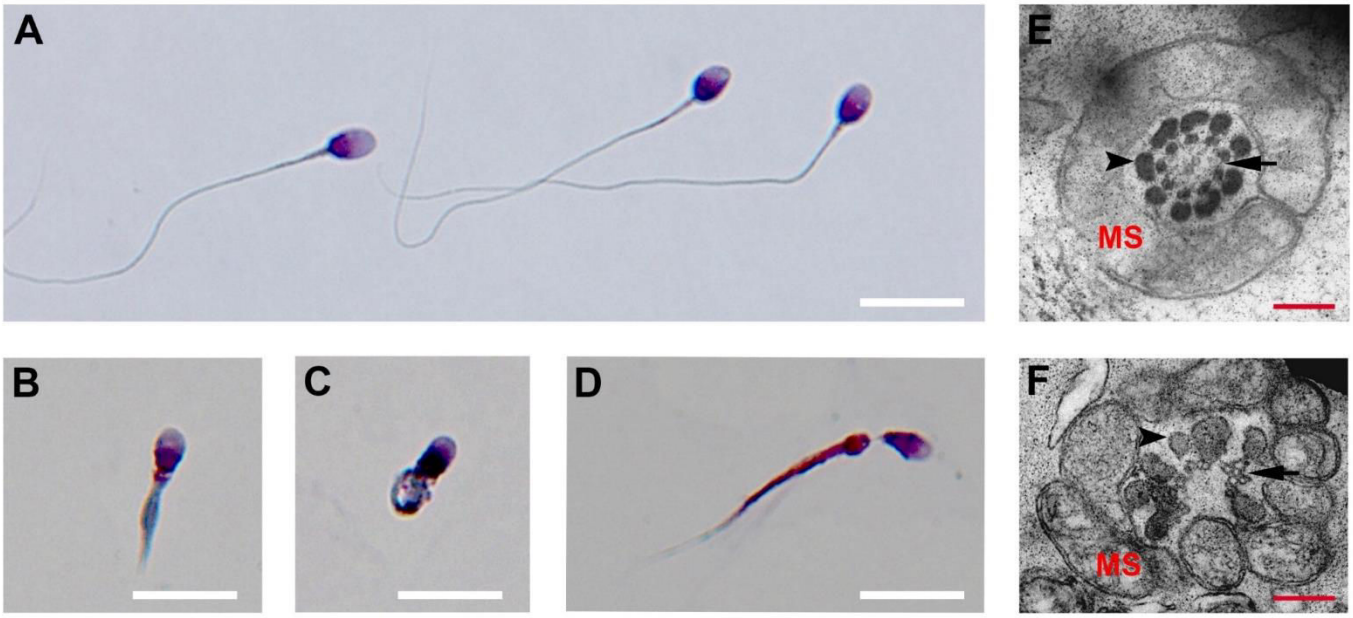


Figure 1

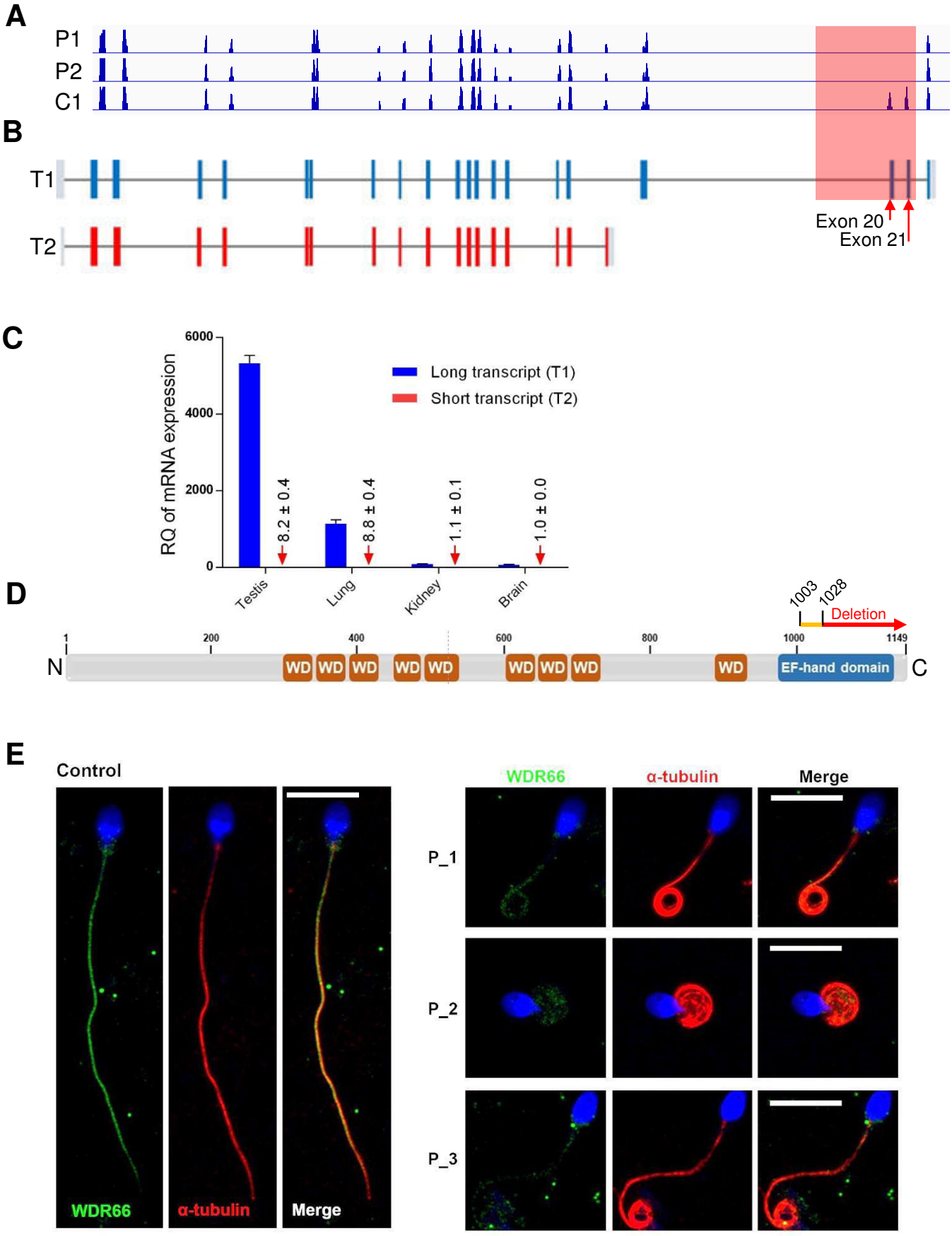

A
P1
P2
C1
B
T1
T2
Exon 20
Exon 21
C
RQ of mRNA expression
6000
4000
2000
0
Long transcript (T1)
Short transcript (T2)
8.2 ± 0.4
8.8 ± 0.4
1.1 ± 0.1
1.0 ± 0.0
Testis
Lung
Kidney
Brain
D
1003
1028
Deletion
1
200
400
600
800
1000
1149
N
WD
EF-hand domain
C
E
Control
WDR66
α-tubulin
Merge
P_1
P_2
P_3

Figure 2

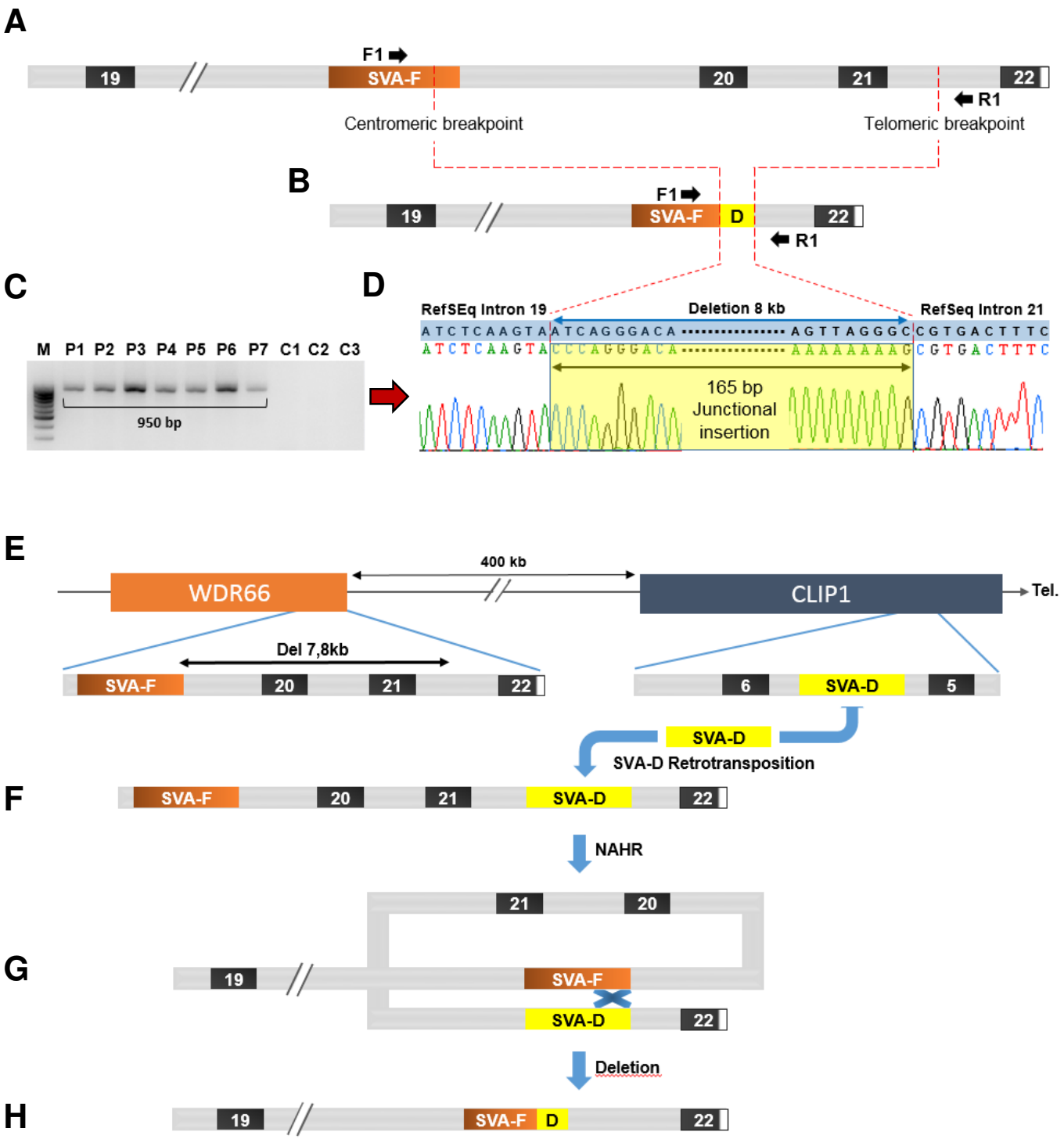
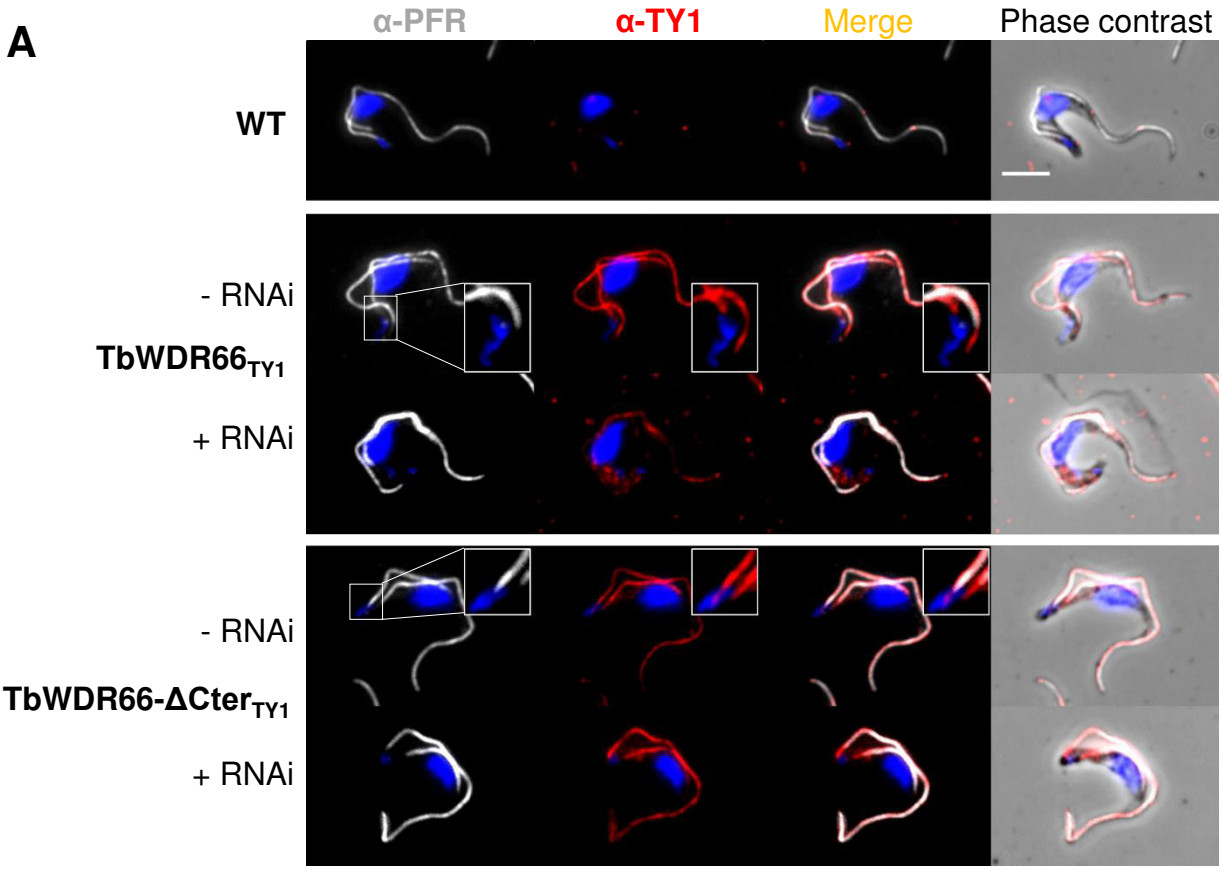
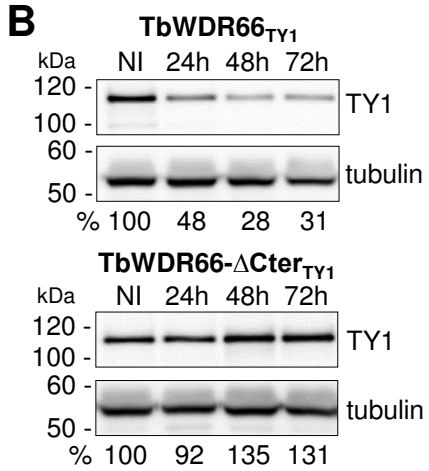
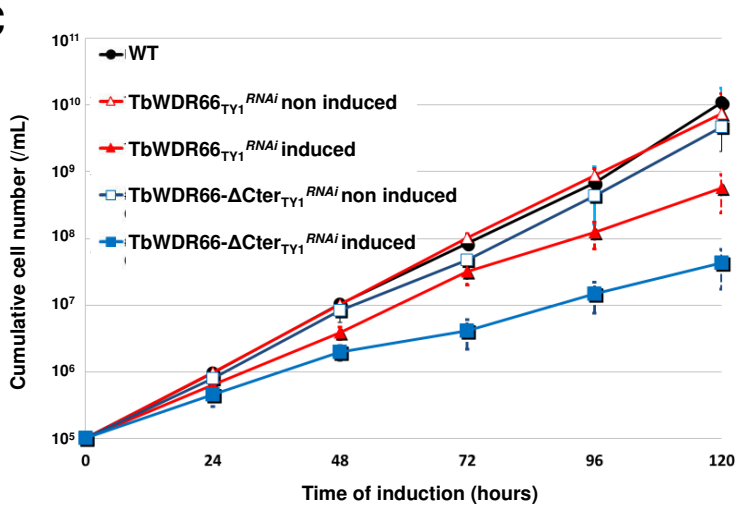

A
F1
19
SVA-F
20
21
22
R1
Centromeric breakpoint
Telomeric breakpoint
B
F1
19
SVA-F
D
22
R1
C
M P1 P2 P3 P4 P5 P6 P7 C1 C2 C3
950 bp
D
RefSEq Intron 19
Deletion 8 kb
RefSeq Intron 21
165 bp
Junctional
insertion
E
400 kb
WDR66
CLIP1
Tel.
Del 7,8kb
SVA-F
20
21
22
6
SVA-D
5
SVA-D
SVA-D Retrotransposition
F
SVA-F
20
21
SVA-D
22
NAHR
G
21
20
19
SVA-F
SVA-D
22
Deletion
H
19
SVA-F
D
22

Figure 3


A
α-PFR
α-TY1
Merge
Phase contrast
WT
- RNAi
TbWDR66TY1
+ RNAi
- RNAi
TbWDR66-ΔCterTY1
+ RNAi


B
TbWDR66TY1
kDa
NI 24h 48h 72h
120 -
100 -
60 -
50 -
TY1
tubulin
% 100 48 28 31
TbWDR66-ΔCterTY1
kDa
NI 24h 48h 72h
120 -
100 -
60 -
50 -
TY1
tubulin
% 100 92 135 131


C
WT
TbWDR66TY1RNAi non induced
TbWDR66TY1RNAi induced
TbWDR66-ΔCterTY1RNAi non induced
TbWDR66-ΔCterTY1RNAi induced
Cumulative cell number (/mL)
10^11
10^10
10^9
10^8
10^7
10^6
10^5
0
24
48
72
96
120
Time of induction (hours)

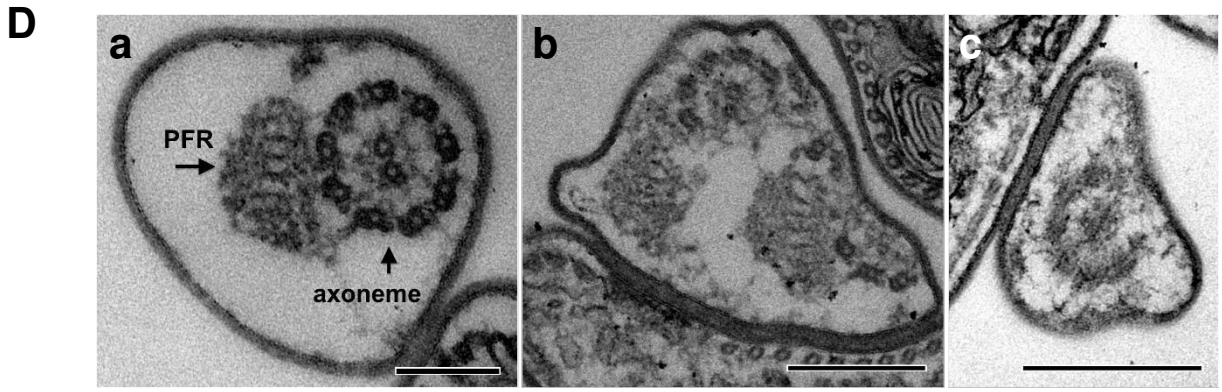

D
a
PFR
axoneme
b
c

Figure 4

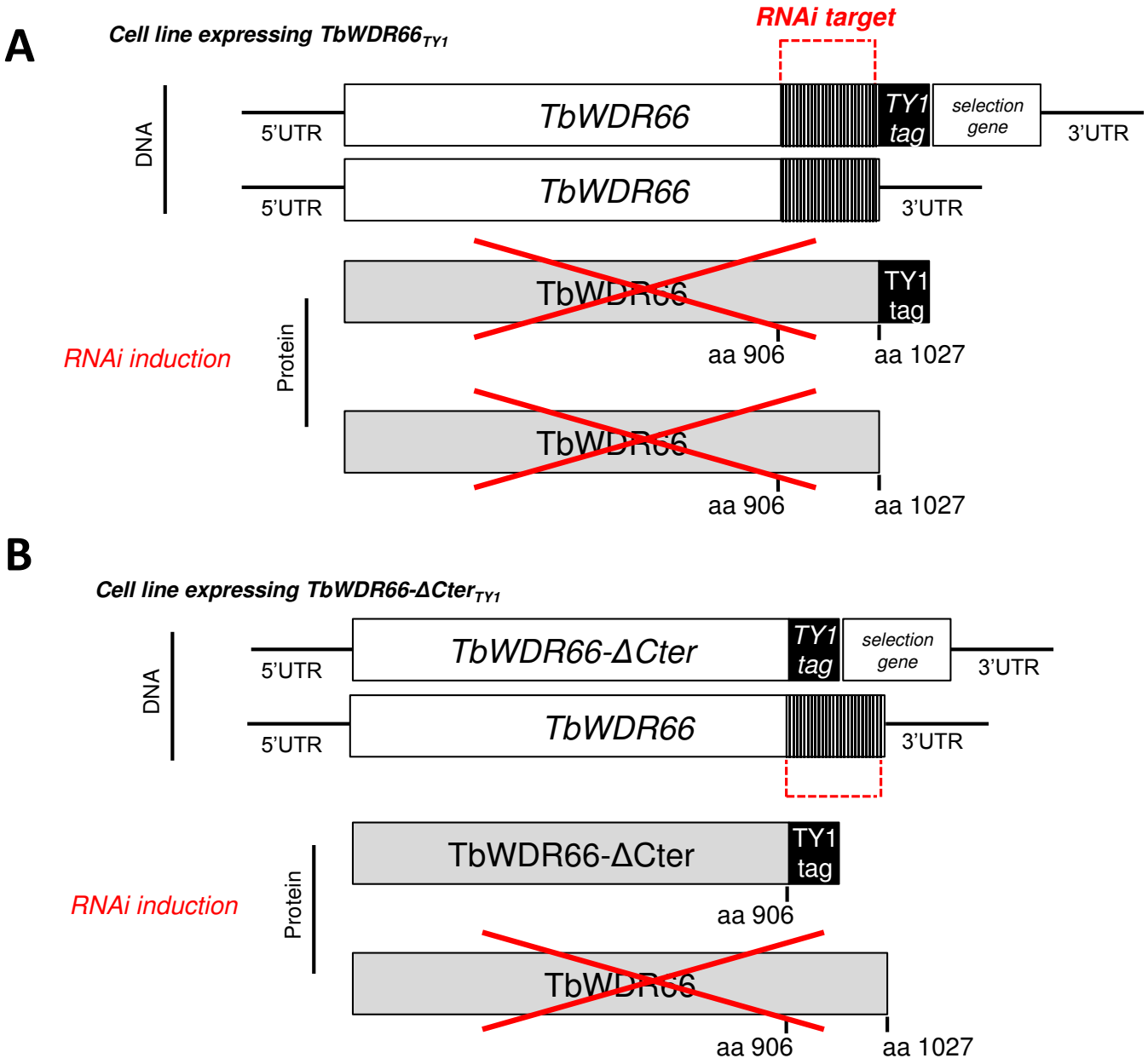

A
Cell line expressing TbWDR66TY1
RNAi target
DNA
5'UTR
TbWDR66
TY1 tag
selection gene
3'UTR
5'UTR
TbWDR66
3'UTR
RNAi induction
Protein
TbWDR66
TY1 tag
aa 906
aa 1027
TbWDR66
aa 906
aa 1027
B
Cell line expressing TbWDR66-ΔCterTY1
DNA
5'UTR
TbWDR66-ΔCter
TY1 tag
selection gene
3'UTR
5'UTR
TbWDR66
3'UTR
RNAi induction
Protein
TbWDR66-ΔCter
TY1 tag
aa 906
TbWDR66
aa 906
aa 1027

Figure S1

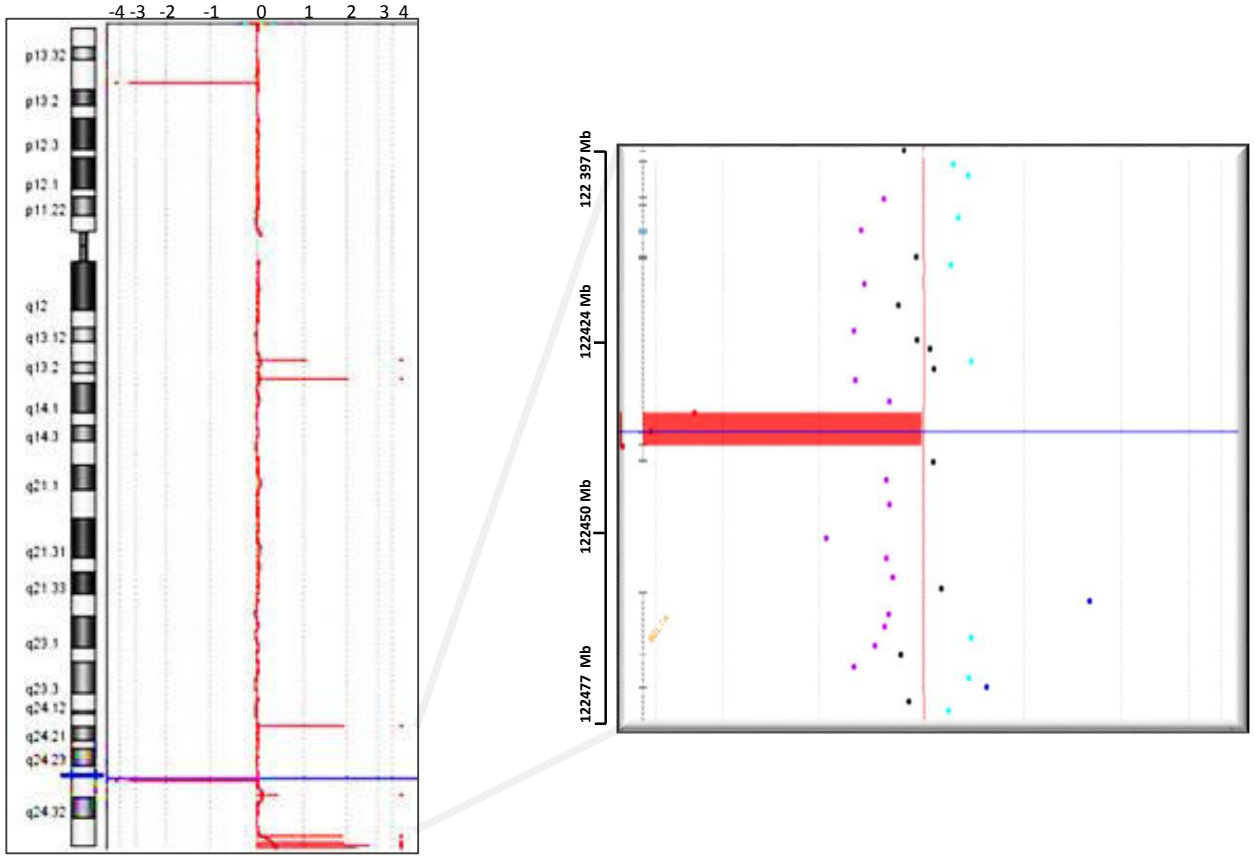

-4 -3 -2 -1 0 1 2 3 4
p13.32
p13.2
p12.3
p12.1
p11.22
q12
q13.12
q13.2
q14.1
q14.3
q21.1
q21.31
q21.33
q23.1
q23.3
q24.12
q24.21
q24.23
q24.32
122 397 Mb
122424 Mb
122450 Mb
122477 Mb

Figure S2

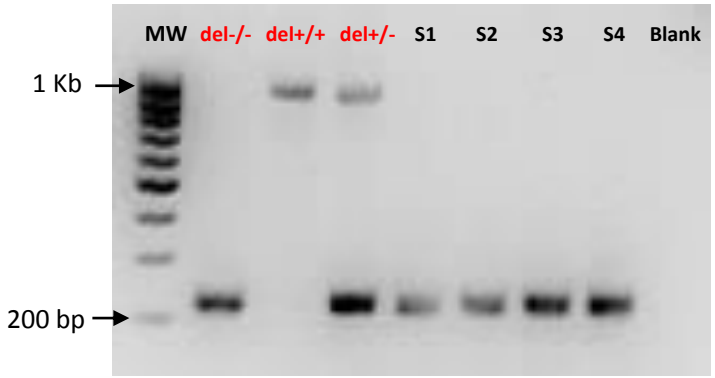


Figure S3

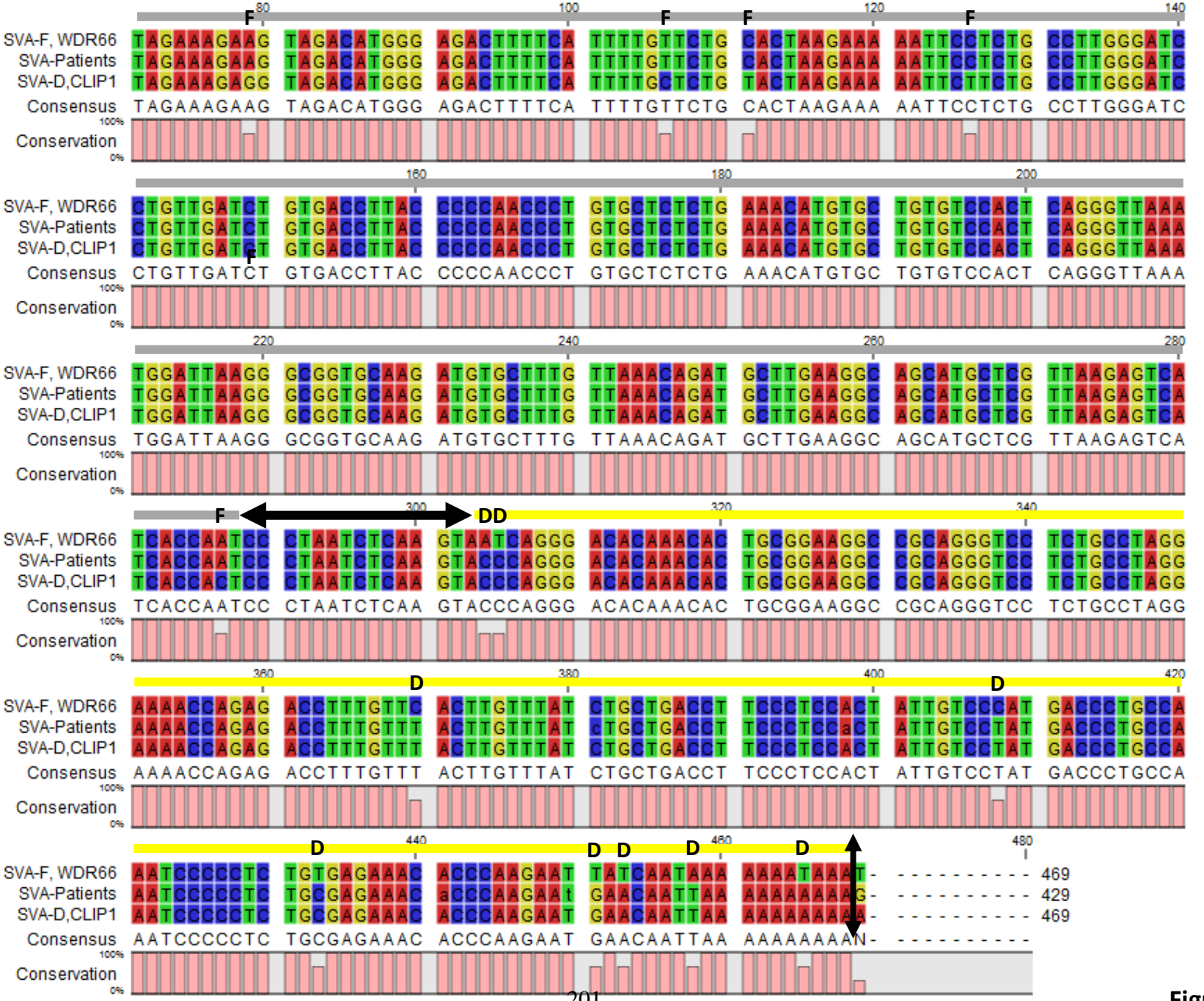

SVA-F, WDR66
SVA-Patients
SVA-D,CLIP1
Consensus
Conservation
100%
0%
F
D
DD
TAGAAAGAAG TAGACATGGG AGACTTTTCA TTTTGTTCTG CACTAAGAAA AATTCCTCTG CCTTGGGATC
CTGTTGATCT GTGACCTTAC CCCCAACCCT GTGCTCTCTG AAACATGTGC TGTGTCCACT CAGGGTTAAA
TGGATTAAGG GCGGTGCAAG ATGTGCTTTG TTAAACAGAT GCTTGAAGGC AGCATGCTCG TTAAGAGTCA
TCACCAATCC CTAATCTCAA GTACCCAGGG ACACAAACAC TGCGGAAGGC CGCAGGGTCC TCTGCCTAGG
AAAACCAGAG ACCTTTGTTT ACTTGTTTAT CTGCTGACCT TCCCTCCACT ATTGTCCTAT GACCCTGCCA
AATCCCCCTC TGCGAGAAAC ACCCAAGAAT GAACAATTAA AAAAAAAAAN
AATCCCCCTC TGTGAGAAAC ACCCAAGAAT TATCAATAAA AAAAATAAAT
AATCCCCCTC TGCGAGAAAC aCCCAAGAAt GAACAATTAA AAAAAAAAAG
AATCCCCCTC TGCGAGAAAC ACCCAAGAAT GAACAATTAA AAAAAAAAAA
469
429
469

Figure S4

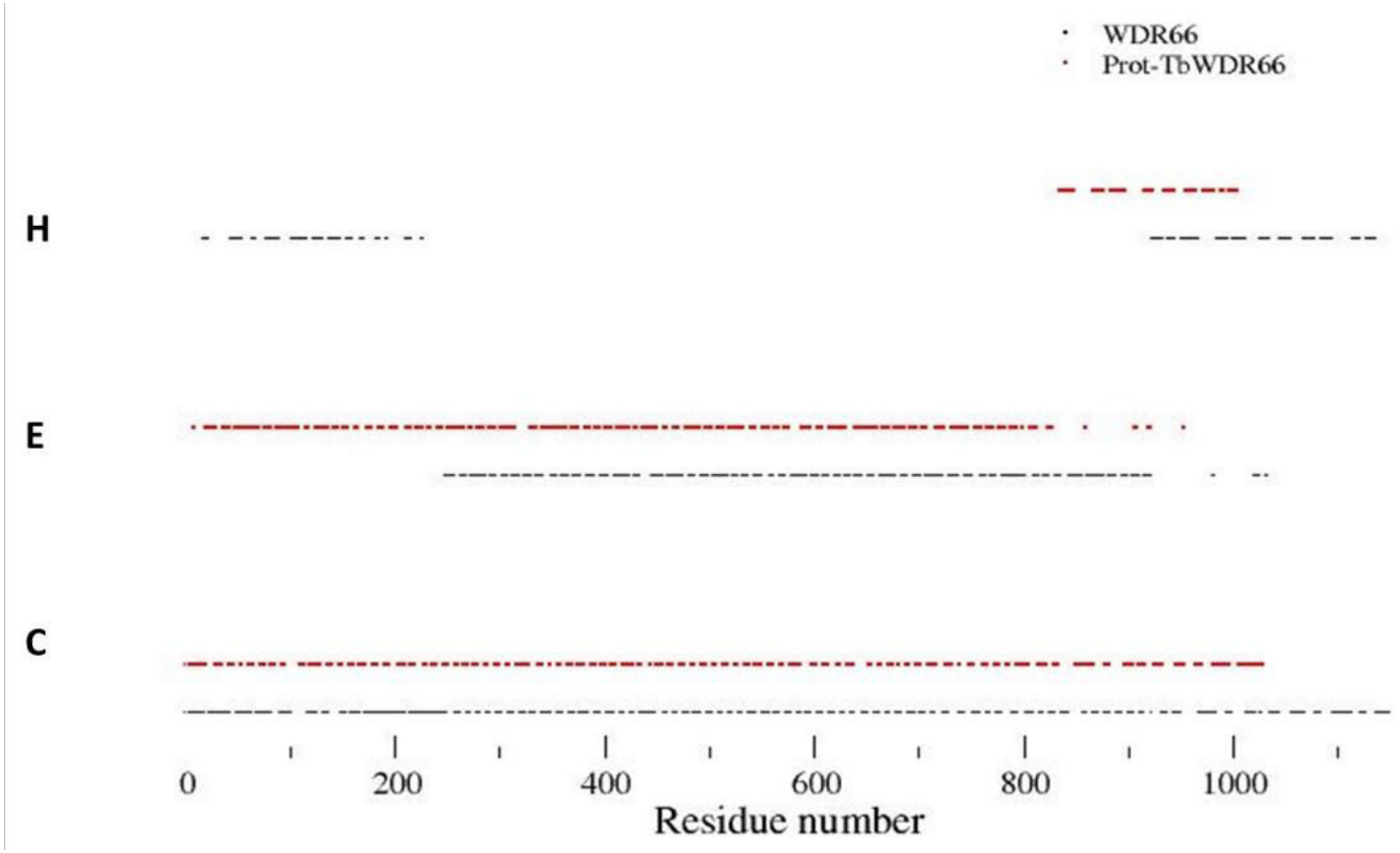


Figure S5

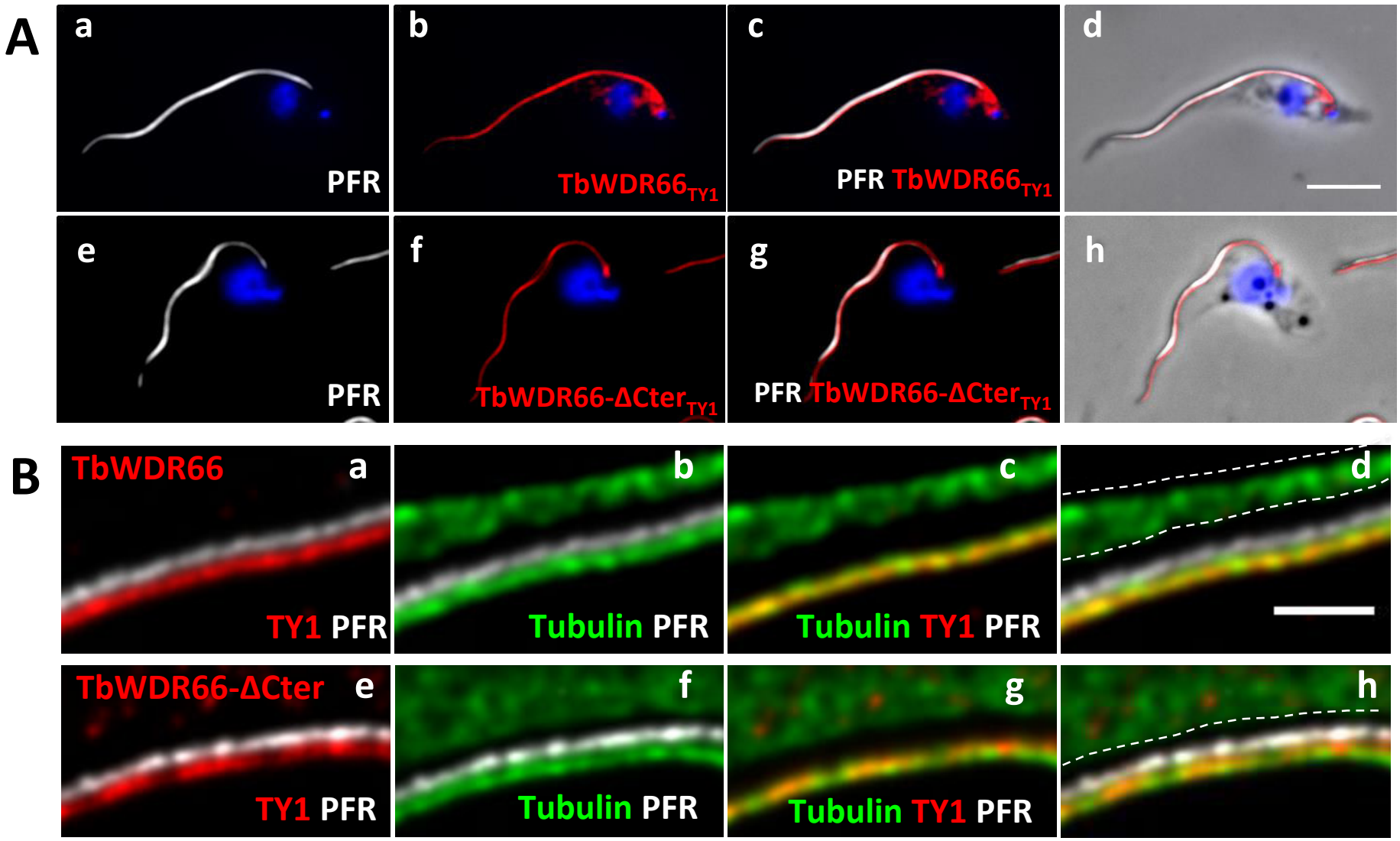

A
a
PFR
b
TbWDR66TY1
c
PFR TbWDR66TY1
d
e
PFR
f
TbWDR66-ΔCterTY1
g
PFR TbWDR66-ΔCterTY1
h
B
TbWDR66
a
TY1 PFR
b
Tubulin PFR
c
Tubulin TY1 PFR
d
TbWDR66-ΔCter
e
TY1 PFR
f
Tubulin PFR
g
Tubulin TY1 PFR
h

Figure S6

Table S1. List of primers used for multiplex PCR to amplify differentially wild-type and mutated allele and for deleted allele sequencing.

Primer name	Primer sequence (5'-3')	Tm (°C)	Product length (bp)
WDR66-Ex20F	AGCTTTGAGGTGCTCGGTTA	59	220 (in controls)
WDR66-Int20R	CCTCATGGCTCAAATCTGGT	58	
WDR66-Int19F (F1)	GAATAGAAAGGCGGGAAAGG	56	950 (in mutated patients)
WDR66-Int22R (R1)	AACTCTGGGTGGAAATTGCTTA	58	

Table S2. List of primers used for qRT-PCR experiment on human sperm extracts.
Transcripts' names are listed in Ensembl database (assembly GRCh37/hg19). Their respective transcript ID is ENST00000288912.4 (T1) and ENST00000397454.2 (T2).

Primer name	Primer sequence (5'-3')	Tm (°C)	Product length (bp)
GAPDH-F	ACCCACTCCTCCACCTTTGA	60	101
GAPDH-R	CTGTTGCTGTAGCCAAATTCGT	60	
WDR66-T1F	AGCTTTGAGGTGCTCGGTTA	60	132
WDR66-T1R	CAGTGAAGCAAAGCAATCCA	60	
WDR66-T2F	GGTGCAGTGGAAAATCACCT	60	139
WDR66-T2R	CTGTGAATCCCAGCTCCACT	60	

Table S3. Antibodies used for immunofluorescence experiments on human sperm cells.

Primary antibodies	Source	Type	Dilution	Reference
WDR66	Rabbit	Polyclonal	1/20	Sigma-Aldrich- HPA040005
α-tubulin	Mouse	Monoclonal	1/200	Sigma-Aldrich-T7451

Table S4. Endogenous tagging and RNAi primer sequences for *T. brucei* experiments.

***TbWDR66* RNAi**	**1036**	5'-GCCGGCCGCTCTAGAtgtgatgtgagcagcattccg-3'
	1037	5'-TAAGCTTGCTCTAGAtttcctcttccccaccgtca-3'
***TbWDR66*$_{TY1}$ endogenous tagging**	**1038**	5'-CGGAAAATCTTCTTGGATTTGAAGATTATGATGCTATGGCCCAGGGTGATGACGGTGGGGAAGAGGAAATGAGTTTGCAGggttctggtagtggttcc-3'
	1039	5'-GAGAGAAAACGAGCATTCCACTGATGGTTTGGTACCTTTAGTCTCCAAATTCACCTGCCCATAACTCTCTGTTCCCTCACccaatttgagagacctgtgc-3'
***TbWDR66-ΔCter*$_{TY1}$ endogenous tagging**	**1039**	5'-GAGAGAAAACGAGCATTCCACTGATGGTTTGGTACCTTTAGTCTCCAAATTCACCTGCCCATAACTCTCTGTTCCCTCACccaatttgagagacctgtgc-3'
	1040	5'-CGGAAAATCTTCTTGGATTTGAAGATTATGATGCTATGGCCCAGGGTGATGACGGTGGGGAAGAGGAAATGAGTTTGCAGggttctggtagtggttcc-3'

Table S5. List of primary and secondary antibodies used in immunofluorescence experiments in *Trypanosoma*.

Antibodies	References	Species	Protein localization	Dilution
TAT1 (tubulin)	Described in (Woods et al., 1989)	Mouse IgG2a	Cytoskeleton/flagellum	1/50
PFR2	Gift from Dr N. Biteau*	Rabbit	PFR	1/400
BB2 (TY1)	Described in (Bastin et al., 1996)	Mouse IgG1	TY1-tagged protein	1/500
Anti-mouse IgG2a Alexa-488	Invitrogen A21131	Goat		1/100
Anti-mouse IgG1 Alexa-594	Invitrogen A21125	Goat		1/100
Anti-rabbit ATTO-N647	Sigma 40839	Goat		1/100

Bastin, P., Bagherzadeh, Z., Matthews, K.R., and Gull, K. (1996). A novel epitope tag system to study protein targeting and organelle biogenesis in Trypanosoma brucei. Mol. Biochem. Parasitol. *77*, 235–239.
Woods, A., Sherwin, T., Sasse, R., MacRae, T.H., Baines, A.J., and Gull, K. (1989). Definition of individual components within the cytoskeleton of Trypanosoma brucei by a library of monoclonal antibodies. J. Cell Sci. *93*, 491–500.
* MFP, University of Bordeaux, CNRS UMR 5234

Table S6. List of all chromosome 12 SNVs obtained by WES of all deleted subjects

The smallest common region of homozygosity was establish by taking into account the SNPs common for at least 6 of the 7 analysed patients (P1-P7) that are different from the control individual (C). The region taken into consideration is surrounded by a black box. The extend of the WDR66 gene is highlighted in yellow.

See the excel file **"Table S6"**

Table S7. Protein folds found in the 2 target sequences
WD40 folds and other β-domains are found in the two proteins by program Superfamily with very good confidence (e-value corresponds to the probability of finding such folds by chance in the sequence). Both proteins also share α-EF-Hand domains in their C-Terminal region.

Protein	Family	Domains (aa number)	E-value
WDR66	Quinoprotein alcohol dehydrogenase-like (β)	249-407, 460-486, 601-774	$1.26e^{-33}$
	WD40 repeat-like (β)	691-922	$1.83e^{-16}$
	EF-hand (α)	980-1099	$3.33e^{-7}$
TbWDR66	WD40 repeat-like (β)	218-372, 458-630	$2.75e^{-40}$
	YVTN repeat-like/Quinoprotein amine dehydrogenase (β)	20-94, 148-319	$1.65e^{-8}$
	WD40 repeat-like (β)	591-631, 699-796	$6.59e^{-6}$
	EF-hand (α)	863-981	$2.24e^{-2}$

Table S8. Average semen parameters in the different genotype groups for the 78 MMAF subjects.

Semen parameters	MMAF *WDR66* patients n=7	MMAF *CFAP43* patients n=10	MMAF *CFAP44* patients n=6	MMAF *DNAH1* patients n=6	MMAF all mutations n=29	MMAF with unknown causes n=49	Overall MMAF patients n=78
Mean age (years)	44.7 ± 10.6 (n'=7)	37.7 ± 9.6 (n'=9)	41.3 ± 4.3 (n'=6)	41.5 ± 4.4 (n'=6)	41.0 ± 8.4 (n'=28)	41.9 ± 7.3 (n'=49)	41.6 ± 7.7 (n'=77)
Sperm volume (ml)	3.0 ± 1.1 (n'=7)	3.5 ± 1.4 (n'=8)	3.2 ± 0.87 (n'=6)	3.5 ± 1.2 (n'=6)	3.3 ± 1.1 (n'=28)	3.6 ± 1.5 (n'=48)	3.5 ± 1.4 (n'=75)
Sperm concentration (10^6/ml)	30.3 ± 23.5 (n'=7)	27.2 ± 23.4 (n'=8)	7.9 ± 8.4 (n'=6)	22.9 ± 15.2 (n'=6)	22.3 ± 20.4 (n'= 28)	27.2 ± 37.2 (n'=48)	25.6 ± 32.1 (n'=75)
Motility (a+b+c) 1 h	3.1 ± 1.9 (n'=7)	0 ± 0 (n'=9)	0 ± 0 (n'=6)	2.6 ± 4.2 (n'=6)	1.3 ± 2.5 (n=29)	5.3 ± 6.4 (n'=48)	3.9 ± 5.6 (n'=76)
Vitality	65.6 ± 12.6 (n'=6)	55.5 ± 24 (n'=8)	43.3 ± 22.6 (n'=6)	51.2 ± 23.1 (n'=5)	54.5 ± 21.2 (n'=26)	51.9 ± 20.2 (n'=47)	52.7 ± 20 (n'=72)
Normal spermatozoa	0 (n'=7)	1.25 ± 3.5 (n'=8)	0 ± 0 (n'=6)	0 ± 0 (n'=6)	0.4 ± 1.9 (n'=27)	2.3 ± 4.3 (n'=48)	1.6 ± 2.7 (n'=61)
Abnormal acrosome	71.8 ± 15.3 (n'=6)	41.1 ± 29.8 (n'=6)	85.2 ± 7.9 (n'=5)	47.0 ± 31.3 (n'=5)	60.8 ± 28.3 (n'=22)	63.6 ± 25.8 (n'=44)	62.6 ± 26.5 (n'=66)
Abnormal head morphology	24.1 ± 25.1 (n'=6)	8.0 ± 3.5 (n'=5)	54.6 ± 27.8 (n'=5)	45.0 ± 55.6 (n'=4)	31.9 ± 33.7 (n'=20)	32.3 ± 21.7 (n'=43)	32.2 ± 25.8 (n'=63)
Abnormal base	14.8 ± 15.6 (n'=6)	22.2 ± 14.5 (n'=5)	54.0 ± 29.7 (n'=5)	26.5 ± 22.3 (n'=4)	28.8 ± 24.8 (n'=20)	32.4 ± 19.2 (n'=42)	31.2 ± 21.0 (n'=62)
Absent flagella	27.2 ± 13.1 (n'=6)	21.8 ± 17.6 (n'=5)	36.8 ± 4.1 (n'=5)	25.6 ± 15.9 (n'=5)	26.9 ± 14.3 (n'=22)	17.3 ± 15.5 (n'=45)	20.7 ± 15.7 (n'=66)
Short Flagella	38.6 ± 11.8 (n'=6)	65.3 ± 31.7 (n'=8)	52.2 ± 27 (n'=6)	49.8 ± 24.3 (n'=5)	53.3 ± 25.2 (n'=26)	38.6 ± 27.0 (n'=47)	43.7 ± 27.3 (n'=72)
Coiled Flagella	21.1 ± 8.3 (n'=6)	8.2 ± 6 (n'=6)	14.4 ± 7 (n'=5)	9 ± 6.3 (n'=5)	12.8 ± 9.0 (n'=23)	13.0 ± 9.7 (n'=47)	12.8 ± 9.4 (n'=69)
Flagella of irregular caliber	43.3 ± 22.1 (n'=6)	20.2 ± 19.3 (n'=5)	28.4 ± 16.9 (n'=5)	35 ± 22.7 (n'=5)	32.0 ± 21.8 (n'=22)	31.3 ± 26.6 (n'=46)	31.7 ± 25.1 (n'=67)
Multiple anomalies index	2.6 ± 0.4 (n'=6)	2.3 ± 0.2 (n'=4)	3.4 ± 0.4 (n'=5)	2.3 ± 1.4 (n'=5)	2.6 ± 0.8 (n'=20)	2.7 ± 0.6 (n'=41)	2.7 ± 0.7 (n'=61)

Values are expressed in percent, unless specified otherwise.Values are mean ± SD; n= total number of patients in each group; n'= number of patients used to calculate the average based on available data. We compared statistical differences between MAAF due to *WDR66*, *CFAP43*, *CFAP44* and *DNAH1* mutations versus MAAF due to uncharacterized genetic cause. No significant statistical difference was observed between the two groups.

2- Article 4 : Mutations in *CFAP43* and *CFAP44* cause male infertility and flagellum defects in Trypanosoma and human

Charles Coutton, Alexandra S. Vargas 1, Amir Amiri-Yekta , **Zine-Eddine Kherraf**, Selima Fourati Ben Mustapha, Pauline Le Tanno, Clémentine Wambergue-Legrand, Thomas Karaouzène, Guillaume Martinez, Serge Crouzy, Abbas Daneshipour, Seyedeh Hanieh Hosseini, Valérie Mitchell, Lazhar Halouani, Ouafi Marrakchi, Mounir Makni, Habib Latrous, Mahmoud Kharouf, Jean-François Deleuze, Anne Boland, Sylviane Hennebicq, Véronique Satre, Pierre-Simon Jouk, Nicolas Thierry-Mieg, Beatrice Conne, Denis Dacheux, Nicolas Landrein, Alain Schmitt, Laurence Stouvenel, Patrick Lorès, Elma El Khouri, Serge P. Bottari, Julien Fauré, Jean-Philippe Wolf, Karin Pernet-Gallay, Jessica Escoffier, Hamid Gourabi, Derrick R. Robinson, Serge Nef, Emmanuel Dulioust, Raoudha Zouari, Mélanie Bonhivers, Aminata Touré, Christophe Arnoult & Pierre F. Ray.

Nature Communications. Février 2018.

Contexte et principaux résultats de l'étude :

La première exploration génétique du phénotype MMAF chez des cas familiaux, d'abord par cartographie d'homozygotie puis par séquençage exomique, nous a permis d'identifier *DNAH1* comme premier gène impliqué dans ce phénotype. Cependant, les mutations identifiées étaient variées et la majorité des individus testés par séquençage exomique est restée sans diagnostic génétique. Ceci suggère fortement que d'autres gènes sont impliqués dans ce phénotype. Nous avons donc réalisé une étude par séquençage exomique chez une cohorte de 78 sujets MMAF non apparentés et ayant consulté pour infertilité primaire isolée. Ces patients ont été recrutés à partir de centres de procréation médicalement assistée (PMA) en France, Iran et Tunisie. A cette date, seul le gène *DNAH1* était décrit somme étant associé au phénotype MMAF. Du fait de la taille importante de *DNAH1*, les 78 exons de ce gène n'ont pas été séquencés par la technique de Sanger avant la réalisation du séquençage exomique pour cette nouvelle cohorte. L'analyse des données de séquençage exomique nous a permis d'identifier des mutations homozygotes du gène *DNAH1* chez six patients (8%). D'autres mutations homozygotes délétères ont été identifiées chez 20% des sujets analysés dans deux gènes codant pour des protéines associées aux cils et aux flagelles : *CFAP43* (n=10) et *CFAP44* (n=6). Ces protéines font partie de la famille des protéines WDR (WD40-repeats) qui sont impliquées dans de nombreuses fonctions cellulaires mais dont le rôle dans l'assemblage et la fonction du flagelle spermatique reste inconnu.

Nous avons pu produire des souris knock-out pour ces deux gènes en utilisant la nouvelle technologie d'édition du génome, CRISPR/Cas9. Nous avons montré que les souris mâles adultes sont infertiles et présentent une asthénozoospermie sévère associée à des anomalies morphologiques flagellaires multiples. Nous avons également pu réaliser des études fonctionnelles chez le protiste *Trypanosoma brucei* dont le flagelle présente une forte homologie structurale avec le flagelle spermatique chez l'homme et nous avons observé après atténuation de l'activité des gènes orthologues, une altération de la mobilité flagellaire. Nous avons également déterminé grâce au STED (Immuno-electron and stimulated-emission-depletin) la localisation flagellaire des protéines codées par ces gènes chez *Trypanosoma* entre les microtubules périphériques de l'axonème et le PFR (paraflagellar rod).

Ce travail a donc permis d'identifier deux nouveaux gènes associés au phénotype MMAF. Par ailleurs, l'étude de ces deux modèles animaux nous a permis de valider ces résultats et de suggérer que les protéines CFAP43 et CFAP44 ont probablement une fonction

structurale inédite consistant à relier l'axonème aux structures péri-axonémales. Des études plus approfondies sont nécessaires chez le mammifère afin d'identifier la localisation exacte de ces protéines et les structures axonémales et péri-axonémales avec lesquelles elles interagissent.

nature COMMUNICATIONS

ARTICLE

DOI: 10.1038/s41467-017-02792-7 OPEN

Mutations in CFAP43 and CFAP44 cause male infertility and flagellum defects in *Trypanosoma* and human

Charles Coutton *et al.*[#]

Spermatogenesis defects concern millions of men worldwide, yet the vast majority remains undiagnosed. Here we report men with primary infertility due to multiple morphological abnormalities of the sperm flagella with severe disorganization of the sperm axoneme, a microtubule-based structure highly conserved throughout evolution. Whole-exome sequencing was performed on 78 patients allowing the identification of 22 men with bi-allelic mutations in *DNAH1* ($n = 6$), *CFAP43* ($n = 10$), and *CFAP44* ($n = 6$). CRISPR/Cas9 created homozygous *CFAP43/44* male mice that were infertile and presented severe flagellar defects confirming the human genetic results. Immunoelectron and stimulated-emission-depletion microscopy performed on CFAP43 and CFAP44 orthologs in *Trypanosoma brucei* evidenced that both proteins are located between the doublet microtubules 5 and 6 and the paraflagellar rod. Overall, we demonstrate that CFAP43 and CFAP44 have a similar structure with a unique axonemal localization and are necessary to produce functional flagella in species ranging from *Trypanosoma* to human.

Correspondence and requests for materials should be addressed to P.F.R. (email: pray@chu-grenoble.fr)
[#]A full list of authors and their affliations appears at the end of the paper.

Medical treatment of infertility has rapidly evolved over the past four decades, but much remains to be accomplished[1]. Despite recent success in identifying infertility genes[2–4], most genetic causes of male infertility are currently uncharacterized and additional efforts should be pursued to better characterize male infertility. We demonstrated previously that mutations in the *DNAH1* gene are responsible for multiple morphological abnormalities of the flagella (MMAF), an infertility phenotype characterized by severe asthenozoospermia due to a combination of flagellar defects including short, curled, abnormal width, rolled, or absent flagella[5,6]. *DNAH1* encodes an axonemal inner dynein arm heavy chain, the lack of which leads to a strong disorganization of the axoneme[5]. *DNAH1* mutations were identified in approximately one-third of the studied patients, indicating that MMAF is genetically heterogeneous and that other genes are likely to be involved in this syndrome[2].

In the present study, we analyzed 78 MMAF patients using whole-exome sequencing (WES) and showed that in addition to mutations in *DNAH1*, mutations in *CFAP43* and *CFAP44*, two genes encoding for WD repeat domains (WDR) containing proteins, are responsible for MMAF syndrome and account for 20.5% of this cohort. Most importantly, we investigated the role of these novel genes by performing gene invalidation and silencing in two evolutionary distant models, *Trypanosoma* and mouse, yet sharing an extremely conserved flagellar structure. Using this original approach, we demonstrate the importance of WDR proteins for axonemal structure of flagella and male fertility in humans.

Results

Identification of *CFAP43* and *CFAP44* mutations in MMAF patients. In the present study, we analyzed a cohort of 78 individuals presenting with a MMAF phenotype defined by the presence in the ejaculate of immotile spermatozoa with several abnormalities of the sperm flagellum including short, coiled, absent, and flagella of irregular caliber[5] (Fig. 1a). A majority of patients originated from North Africa, 46 were recruited in Tunisia, 10 in Iran, and 22 in France. The average semen parameters of all 78 MMAF patients included in the cohort are

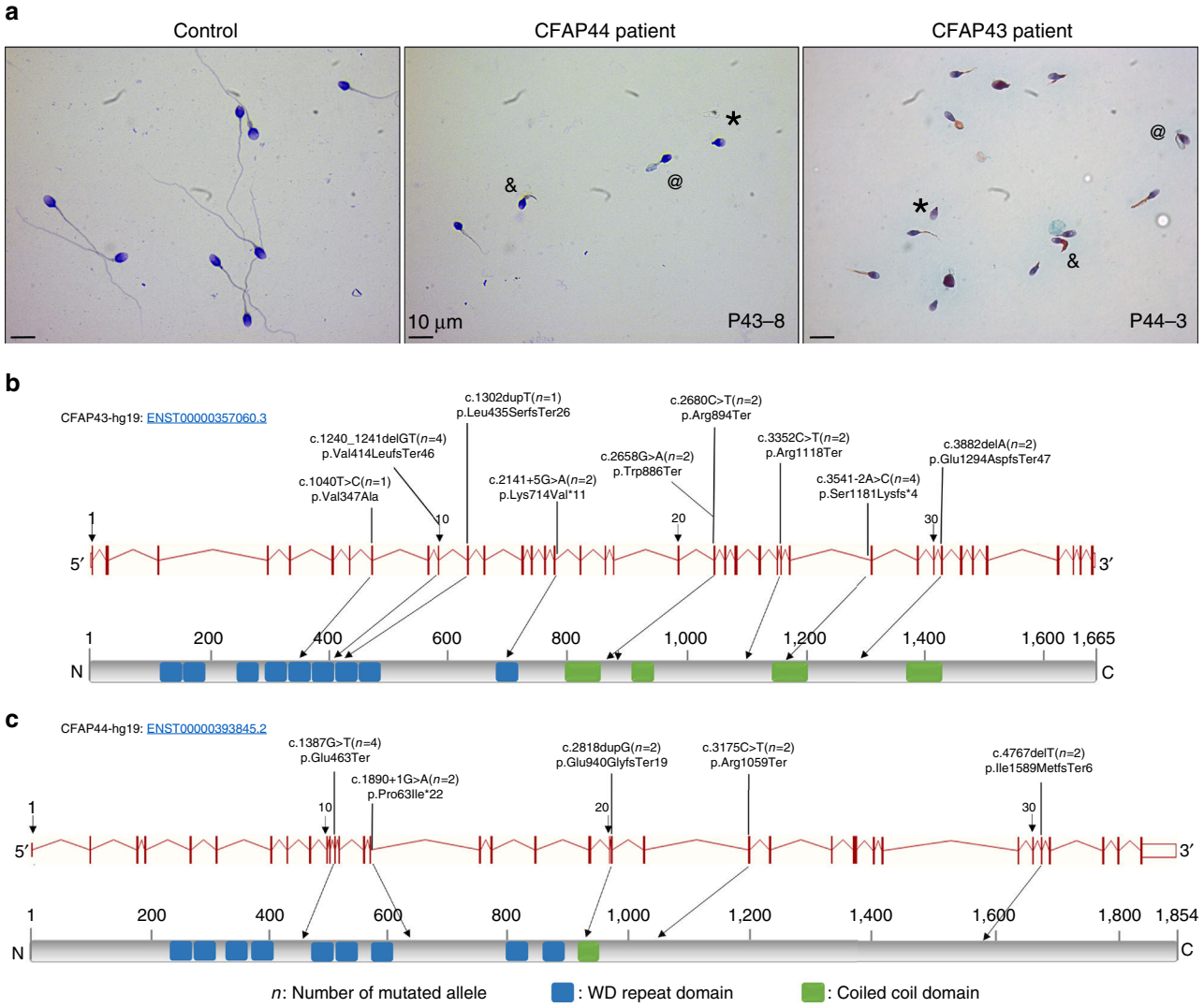


Fig. 1 Morphology of spermatozoa from the patients P_{43}-8 and P_{44}-3 and position of *CFAP43* and *CFAP44* mutations in the intron–exon structure and in the protein. **a** All spermatozoa have a much shorter flagellum than those of controls. Additional features of MMAF spermatozoa are absent (*), thick (&), and rolled flagella (@). **b** Mutations identified in the *CFAP43* gene. **c** Mutations identified in the *CFAP44* gene. Blue squares stand for WD repeat domains and green squares for coiled-coiled domains as described by Uniprot server. Mutations are annotated in accordance to the HGVS's recommendations

Table 1 Average semen parameters in the different genotype groups for the 78 MMAF subjects

Semen parameters	MMAF *CFAP43* patients, $n = 10$	MMAF *CFAP44* patients, $n = 6$	MMAF *DNAH1* patients, $n = 6$	MMAF all mutations, $n = 22$	MMAF with unknown causes, $n = 56$	Overall MMAF, $n = 78$
Mean age (years)	37.7 ± 9.6 ($n' = 9$)	41.3 ± 4.3 ($n' = 6$)	41.5 ± 4.4 ($n' = 6$)	39.8 ± 7 ($n' = 21$)	42.3 ± 7.9 ($n' = 56$)	41.6 ± 7.7 ($n' = 77$)
Sperm volume (ml)	3.5 ± 1.4 ($n' = 8$)	3.2 ± 0.87 ($n' = 6$)	3.5 ± 1.2 ($n' = 6$)	3.4 ± 1.2 ($n' = 20$)	3.5 ± 1.5 ($n' = 55$)	3.5 ± 1.4 ($n' = 75$)
Sperm concentration (10^6/ml)	27.2 ± 23.4 ($n' = 8$)	7.9 ± 8.4 ($n' = 6$)	22.9 ± 15.2 ($n' = 6$)	20.1 ± 18.8 ($n' = 20$)	27.6 ± 35.7 ($n' = 55$)	25.6 ± 32.1 ($n' = 75$)
Motility (a + b + c) 1 h	0 ± 0 ($n' = 9$)	0 ± 0 ($n' = 6$)	2.6 ± 4.2 ($n' = 6$)	0.7 ± 2.4* ($n' = 21$)	5 ± 6.1 ($n' = 55$)	3.9 ± 5.6 ($n' = 76$)
Vitality	55.5 ± 24 ($n' = 8$)	43.3 ± 22.6 ($n' = 6$)	51.2 ± 23.1 ($n' = 5$)	50.5 ± 22.7 ($n' = 19$)	53.4 ± 20 ($n' = 53$)	52.7 ± 20 ($n' = 72$)
Normal spermatozoa	1.25 ± 3.5 ($n' = 8$)	0 ± 0 ($n' = 6$)	0 ± 0 ($n' = 6$)	0.5 ± 2.3 ($n' = 20$)	2.1 ± 4.1 ($n' = 54$)	1.6 ± 2.7 ($n' = 61$)
Absent flagella	21.8 ± 17.6 ($n' = 5$)	36.8 ± 4.1 ($n' = 5$)	25.6 ± 15.9 ($n' = 5$)	28.1 ± 14.4* ($n' = 15$)	18.5 ± 15.5 ($n' = 51$)	20.7 ± 15.7 ($n' = 66$)
Short flagella	65.3 ± 31.7 ($n' = 8$)	52.2 ± 27 ($n' = 6$)	49.8 ± 24.3 ($n' = 5$)	57.1 ± 27.9* ($n' = 19$)	38.9 ± 25.7 ($n' = 53$)	43.7 ± 27.3 ($n' = 72$)
Coiled flagella	8.2 ± 6 ($n' = 6$)	14.4 ± 7 ($n' = 5$)	9 ± 6.3 ($n' = 5$)	10.4 ± 6.6 ($n' = 16$)	13.5 ± 10 ($n' = 53$)	12.8 ± 9.4 ($n' = 69$)
Bent flagella	10.3 ± 6 ($n' = 3$)	9 ($n' = 1$)	6 ± 8.5 ($n' = 2$)	8.7 ± 5.8 ($n' = 6$)	12 ± 9 ($n' = 20$)	4.2 ± 8.4 ($n' = 26$)
Flagella of irregular caliber	20.2 ± 19.3 ($n' = 5$)	28.4 ± 16.9 ($n' = 5$)	35 ± 22.7 ($n' = 5$)	27.9 ± 19.4 ($n' = 15$)	32.8 ± 26.7 ($n' = 52$)	31.7 ± 25.1 ($n' = 67$)
Multiple anomalies index	2.3 ± 0.2 ($n' = 4$)	3.4 ± 0.4 ($n' = 5$)	2.3 ± 1.3 ($n' = 5$)	2.7 ± 1 ($n' = 14$)	2.7 ± 0.6 ($n' = 47$)	2.7 ± 0.7 ($n' = 61$)

Values are expressed in percent, unless specified otherwise
Values are mean ± SD; n = total number of patients in each group; n' = number of patients used to calculate the average based on available data
We compared statistical differences between MMAF due to *CFAP43*, *CFAP44*, and *DNAH1* mutations versus MMAF due to uncharacterized genetic cause
CFAP cilia and flagella associated protein; MMAF, multiple morphological abnormalities of the flagella
*A significant difference $P < 0.05$

Table 2 *CFAP43* (*WDR96*), *CFAP44* (*WDR52*), and *DNAH1* variants identified by WES for all the analyzed subjects ($n = 78$)

Gene	Variant coordinates	Transcript	cDNA variation	Amino-acid variation	Patients	Nationality	Hom./Het.
CFAP43	Chr10:105,912,486	ENST00000357060	c.3541−2A>C	p.Ser1181Lysfs*4	P_{43}-1, P_{43}-2	Tunisia	Homozygous
CFAP43	Chr10:105,956,662	ENST00000357060	c.1240_1241delGT	p.Val414LeufsTer46	P_{43}-3, P_{43}-4	Afghanistan, Iran	Homozygous
CFAP43	Chr10:105,928,535	ENST00000357060	c.2658G>A	p.Trp886Ter	P_{43}-5	Algeria	Homozygous
CFAP43	Chr10:105,928,513	ENST00000357060	c.2680C>T	p.Arg894Ter	P_{43}-6	Algeria	Homozygous
CFAP43	Chr10:105,905,296	ENST00000357060	c.3882delA	p.Glu1294AspfsTer47	P_{43}-7	Tunisia	Homozygous
CFAP43	Chr10:105,921,781	ENST00000357060	c.3352C>T	p.Arg1118Ter	P_{43}-8	Tunisia	Homozygous
CFAP43	Chr10:105,953,765	ENST00000357060	c.1302dupT	p.Leu435SerfsTer26	P_{43}-9	France	Heterozygous
CFAP43	Chr10:105,963,485	ENST00000357060	c.1040T>C	p.Val347Ala	P_{43}-9	France	Heterozygous
CFAP43	Chr10:105,944,769	ENST00000357060	c.2141+5G>A	p.Lys714Val*11	P_{43}-10	Turkey	Homozygous
CFAP44	Chr3:113,114,596	ENST00000393845	c.1890+1G>A	p.Pro631Ile*22	P_{44}-1	Tunisia	Homozygous
CFAP44	Chr3:113,063,450	ENST00000393845	c.3175C>T	p.Arg1059Ter	P_{44}-2	Tunisia	Homozygous
CFAP44	Chr3:113,082,107_11,3082,108	ENST00000393845	c.2818dupG	p.Glu940GlyfsTer19	P_{44}-3	Morocco	Homozygous
CFAP44	Chr3:113,119,479	ENST00000393845	c.1387G>T	p.Glu463Ter	P_{44}-4, P_{44}-5	Algeria	Homozygous
CFAP44	Chr3:113,023,990	ENST00000393845	c.4767delT	p.Ile1589MetfsTer6	P_{44}-6	Algeria	Homozygous
DNAH1	Chr3:52,414,073	ENST00000420323	c.7531delC	p.Gln2511SerfsTer27	P_{DNA}-1	Tunisia	Homozygous
DNAH1	Chr3:52,382,924	ENST00000420323	c.2127dupC	p.Ile710HisfsTer4	P_{DNA}-2	Tunisia	Homozygous
DNAH1	Chr3:52,395,227	ENST00000420323	c.4744_4752 delCCAGCTGGC	p.Pro1582_Gly1584del	P_{DNA}-3	Tunisia	Homozygous
DNAH1	Chr3:52,394,055	ENST00000420323	c.4531G>A	p.Val1511Met	P_{DNA}-4	Tunisia	Heterozygous
DNAH1	Chr3:52,394,397	ENST00000420323	c.4642C>G	p.Leu1548Val	P_{DNA}-4	Iran	Heterozygous
DNAH1	Chr3:52,409,423	ENST00000420323	c.7153T>A	p.Trp2385Arg	P_{DNA}-5	Iran	Heterozygous
DNAH1	Chr3:52,423,486	ENST00000420323	c.9505C>G	p.Arg3169Gly	P_{DNA}-5	France	Heterozygous

cDNA, complementary DNA; CFAP cilia and flagella associated protein; WES, whole-exome sequencing

described in Table 1. Nearly no spermatozoa with normal morphology could be observed in the ejaculate of MMAF individuals (1.6%); an average of 20.7 and 43.7% of spermatozoa had no flagella and short flagella, respectively, and 31.7% of the spermatozoa had flagella with an irregular caliber. As a result, total sperm motility was dramatically reduced to 3.9% (normal value >40%), which prevented natural conception for all individuals. Given the notion of consanguinity for most individuals from the cohort, we postulated that infertility was likely transmitted through recessive inheritance and probably often resulted from homozygous mutations. After exclusion of frequent variants and applying stringent filters, a limited list of homozygous variants was identified for each proband. First, we identified six patients (7.7%) with mutations in the *DNAH1* gene (Table 2), previously identified as the main cause of MMAF phenotype[5,6]. We subsequently identified 10 subjects with variants in *CFAP43* (12.8%), eight of which had a homozygous loss-of-function variant and two with two likely deleterious variants (Table 2). In addition, six subjects (7.7%) had a homozygous loss-of-function variant in *CFAP44* (Table 2). These two *CFAP* genes (for cilia and flagella associated protein) were reported in public expression databases to be strongly expressed in the testis and to be connected with cilia and flagella structure and/or functions[7]. Quantitative real-time reverse transcription PCR (RT-qPCR) experiments performed in human and mouse tissue panels confirmed that *CFAP43* and *CFAP44* mRNA in testis was predominant and very significantly higher than in the other tested tissues (Supplementary Fig. 1). Taking into account the high number of mutated patients and the specific expression pattern of the two genes, we focused on these two genes, which appeared as

the best candidates to explain the primary infertility observed for these individuals. All individuals with *CFAP43* or *CFAP44* mutations were unrelated; to our knowledge, none carried rare variants in genes previously reported to be associated with male infertility. *CFAP43* and *CFAP44* encoded proteins that belong to the WDR protein family and are both composed of nine WD repeats[8].

CFAP43 (also known as *WDR96*, NM_025145) is localized on chromosome 10 and contains 38 exons encoding a predicted 1,665-amino-acid protein (Q8NDM7). We identified 9 different pathogenic variants in *CFAP43* in 10 unrelated individuals from the cohort. The splicing variant c.3541−2A>C was identified in two unrelated individuals and affects a consensus splice acceptor site of *CFAP43* intron 27 (Table 2, Fig. 1b). This variant was absent from the 60,706 unrelated individuals aggregated in the ExAC database (http://exac.broadinstitute.org), consistent with a deleterious effect on protein function. Three other variants correspond to stop-gain mutations identified in three subjects: c.2658G>A and c.2680C>T are located in exon 21 and c.3352C>T in the exon 26 (Fig. 1b). These three variants were found in the ExAC database with very low allele frequency ranging from $8.24e^{-06}$ to $9.89e^{-05}$. Two others mutations were small frameshift indels not listed in ExAC: c.1240_1241delGT (found in two unrelated patients) and c.3882delA (Table 2, Fig. 1b). All these mutations generate premature stop codons and are predicted to produce no protein or truncated non-functional proteins. Apart from these obvious harmful mutations found in eight patients, we also identified two patients harboring likely deleterious mutations in *CFAP43*. The first patient harbors a homozygous splicing mutation c.2141+5G>A, not listed in ExAC which, according to the splice site prediction algorithm Human Splicing Finder (http://www.umd.be/HSF3), likely alters the consensus splice donor site of *CFAP43* exon 16. Unfortunately, we could not obtain any additional biological material from this patient and could not verify the effect of this variant on mRNA expression. The second patient is a compound heterozygous carrying one single-nucleotide duplication, c.1320dupT, located in the exon 11 and an additional missense mutation, Val347Ala, located in exon 8 of the *CFAP43* gene. Although prediction software classified this last variant as likely "benign," this mutation affects a conserved residue located in the N-terminal part of the protein within a WDR known to be important for protein/protein interactions[9]. Moreover, this missense variant is found at a very low prevalence in the general population estimated to 7.45e−5 (9/12078). Altogether the above-identified mutations clearly indicate that in humans, loss of function of *CFAP43* is associated with asthenozoospermia and sperm flagellum defects.

CFAP44 (also known as WDR52, NM_001164496) is localized on chromosome 3 and contains 35 exons encoding a predicted 1,854-amino-acid protein (Q96MT7). Five different homozygous variants were identified in six unrelated patients (Table 2, Fig. 1c). Stop-gain mutations c.1387G>T and c.3175C>T are located in exons 12 and 23, respectively. The frameshift mutations, c.4767delT, with a single-nucleotide deletion, are located in exon 31, and the variant c.2818dupG, with a single-nucleotide duplication, is located in exon 21. The last mutation, c.1890 +1C>T, affects exon 15 consensus splice donor site. All these mutations are predicted to generate a premature stop codon. None of these variants were listed in the ExAC database.

The presence of all variants was verified by bidirectional Sanger sequencing (Supplementary Fig. 2). The detailed sperm parameters of the patients carrying *DNAH1*, *CFAP43*, or *CFAP44* mutations were compared with each other. There was no significant difference between the three groups (Table 1) as illustrated by the similar morphologies of sperm from patient P_{43}-8 and P_{44}-3, mutated for *CFAP43* and *CFAP44*, respectively (Fig. 1a). Data from all these groups were therefore pooled and compared with data from the patients with no identified mutation. Patients with a mutation had a significantly higher rate of spermatozoa with short or absent flagella and presented a significantly lower motility rate compared to patients with no identified mutations. There was no difference in the other sperm parameters (Table 1).

***CFAP43* and *CFAP44* mutations induce severe axonemal disorganization**. The internal skeleton of motile cilia and flagella, named the axoneme, is a highly evolutionary conserved structure, which consists of nine doublets of microtubules (DMTs) circularly arranged around a central pair complex (CPC) of microtubules ('9+2' structure). All DMTs are not identical and are organized in a highly reproducible fashion with substructures specific to certain doublets. The convention adopted in the numbering of the DMTs was based on the relative position of each DMT with regard to the plane of the CPC[10,11]. The beating of cilia and flagella is orchestrated by multiprotein-ATPase complexes, located on the peripheral doublets, which provide the sliding force for sperm motility. In addition, the sperm flagellum harbors specific peri-axonemal structures, which are not found in other motile cilia, a helical mitochondrial sheath (MS) in the midpiece, the fibrous sheath (FS) in the principal piece (PP), and outer dense fibers (ODFs) in the midpiece and the proximal part of the PP.

We studied the ultrastructure of sperm cells from mutated patients in *CFAP43* and *CFAP44* by transmission electron microscopy (TEM) (Fig. 2). Due to insufficient amount of sperm cells collected from most *CFAP44* and *CFAP43* patients, only one patient for each gene was studied by TEM. For each patient, we could observe longitudinal sections and >20 cross-sections presenting a sufficient quality to observe the ultrastructure of the axonemal components. Longitudinal sections showed severe axonemal and peri-axonemal defects affecting the ODF, the FS, and the MS, which appeared completely disorganized resulting in aborted flagella or their replacement by a cytoplasmic mass englobing unassembled axonemal components (Fig. 2). In *CFAP43*-mutated and *CFAP44*-mutated patients, approximately 95% of the cross-sections were abnormal and the main defect observed was the absence of the CPC (9+0 conformation) observed in 81.8, and 66.7% of the cases, respectively, compared to 0% in control fertile subject (Supplementary Table 1). In the residual fraction (~5%) of axoneme with normal (9+2) conformation, peri-axonemal structures abnormalities were constantly observed (Supplementary Table 1). The lack of central pair defects was associated with peripheral doublets disorganization in 13.6 and 19% for *CFAP43* and *CFA44*, respectively (Supplementary Table 1). Cross-sections with a single central microtubule (9+1 conformation) were observed in about 10% of cases, only for the CFAP44 patient. Interestingly, in CFAP43 patients, the CPC, when present, was misoriented compared to control sections in which the CPC is normally parallel to the axis of the two longitudinal columns of the FS (Fig. 2).

To define the ultrastructural defects evidenced by TEM, we performed immunofluorescence (IF) experiments using antibodies targeting different axonemal proteins. In *CFAP43*-mutated and *CFAP44*-mutated patients, staining of SPAG6, a protein located in the CPC, was abnormal and atypical compared to control staining (Fig. 3a–c). SPAG6 staining was absent in CFAP43 patient sperm cells, whereas in *CFAP44* patients' sperm cells, SPAG6 immunostaining was present but displayed an abnormal and diffuse pattern concentrated in the midpiece of the spermatozoa, quickly diminishing along the flagellum (Fig. 3d–i).

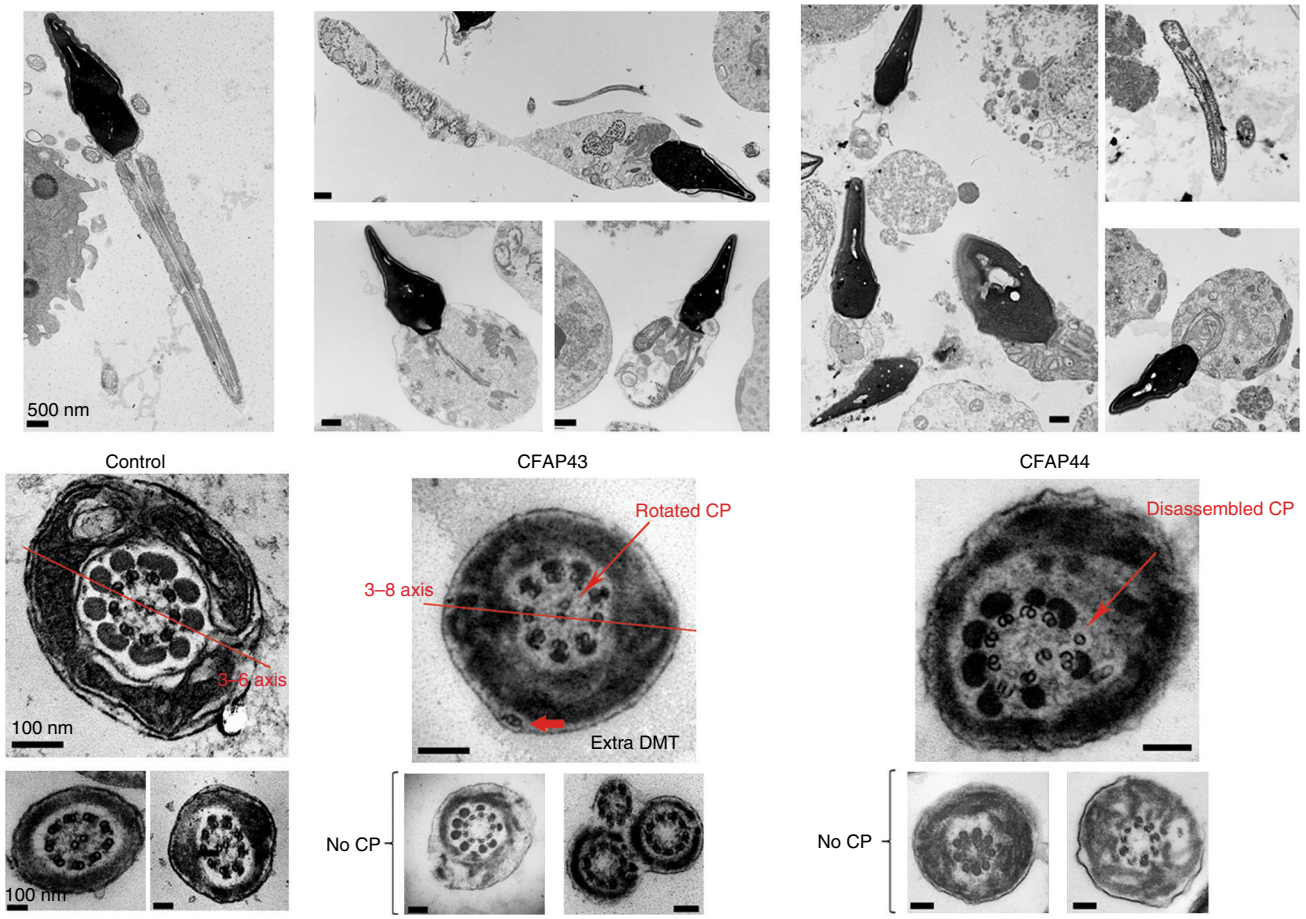


Fig. 2 Transmission electron microscopy analysis of sperm cells from *CFAP43* and *CFAP44* patients shows a severe axonemal disorganization. (Left panel) Longitudinal sections (scale bar = 500 nm) and cross-sections (scale bars, 100 nm) of sperm flagellum from control. (Central panel) Longitudinal sections (scale bars, 500 nm) and cross-sections (scale bars, 100 nm) of sperm flagellum from *CFAP43*-mutated patient (P_{43}-8). We can notice a short short tail corresponding to a cytoplasmic mass containing the different components of the flagellum, all unorganized. In CFAP43 upper cross-section, the CPC is not aligned with DMTs 3 and 8 (red line) and is rotated by 90°. We can observe the absence of central pair of microtubules in other cross-sections. (Right panel) Ultrastructure of *CFAP44*-mutated sperm (P_{44}-3) longitudinal sections (scale bars = 500 nm) show similar ultrastructure of short tail (cytoplasmic mass). In CFAP44 upper cross-section, the central pair is disassembled and displaced (red arrow). We can observe the absence of central pair of microtubules in other cross-sections. Scales bars for cross-sections = 100 nm

Additionally, staining of the radial spoke head protein RSPH1 presented significantly different patterns compared to controls. In *CFAP43*-mutated patients, the RSPH1 staining was completely abnormal with a marked diffuse staining, concentrated in the midpiece, whereas the tubulin staining remained restricted to the axoneme (Fig. 3m–o). In *CFAP44*-mutated patients' sperm cells, RSPH1 staining was significantly reduced (Fig. 3p–r). For *CFAP43*-mutated and *CFAP44*-mutated patients, immunostaining for AKAP4, DNALI1, DNAI2, and GAS8 were all comparable with controls (Fig. 3j–l), suggesting that FS, outer dynein arms, inner dynein arms, and the nexin-dynein regulatory complex, respectively, were not directly affected by mutations in *CFAP43* or *CFAP44* (Supplementary Table 2—Human). Unfortunately, we could not obtain any specific antibodies against CFAP43 and CFAP44that would allow the localization of these proteins in mouse and human flagella.

Analysis of *Cfap43* and *Cfap44* knockout mice. We assessed the impact of *Cfap43* and *Cfap44* absence on mouse spermatogenesis by generating knockout (KO) animals using the CRISPR-Cas9 technology. We obtained four independent mutational events for *Cfap43* and three for *Cfap44*; all were insertions/deletions of a few nucleotides inducing a translational frameshift expected to lead to the complete absence of the protein or the production of a truncated protein. Because all KO strains for *Cfap43* presented the same reproductive phenotype (identical sperm morphology associated with complete infertility), we randomly chose one and restricted our study to a strain with a 4 bp deletion in exon 21 (delAAGG). Similarly, the *Cfap44* lines presented the same reproductive phenotype, and we focused on a strain with a 7 bp insertion in exon 3 (InsTCAGATA). RT-PCR was performed on testis RNA from $Cfap43^{-/-}$ and $Cfap44^{-/-}$ mice, which confirmed the production of abnormal transcripts in both mutants leading to a premature stop codon (Supplementary Fig. 3).

The reproductive phenotype was studied for both KO mice models. Homozygous KO females were fully fertile and gave litters of normal size (7.8 ± 1.8 and 7.3 ± 3.5 versus 6.7 ± 0.5 pups/litter (mean ± SD, $n = 7$) for $Cfap43^{-/-}$, $Cfap44^{-/-}$, and wild-type (WT), respectively), contrary to homozygous KO males, which exhibited complete infertility when mated with WT females (Fig. 4a). Epididymal sperm concentrations were 11.4 ± 0.9, 11.2 ± 0.4, and $14.4 \pm 3.1 \times 10^6$ sperm/ml (mean ± SD, $n = 3$), for $Cfap43^{-/-}$, $Cfap44^{-/-}$, and WT, respectively, and fall

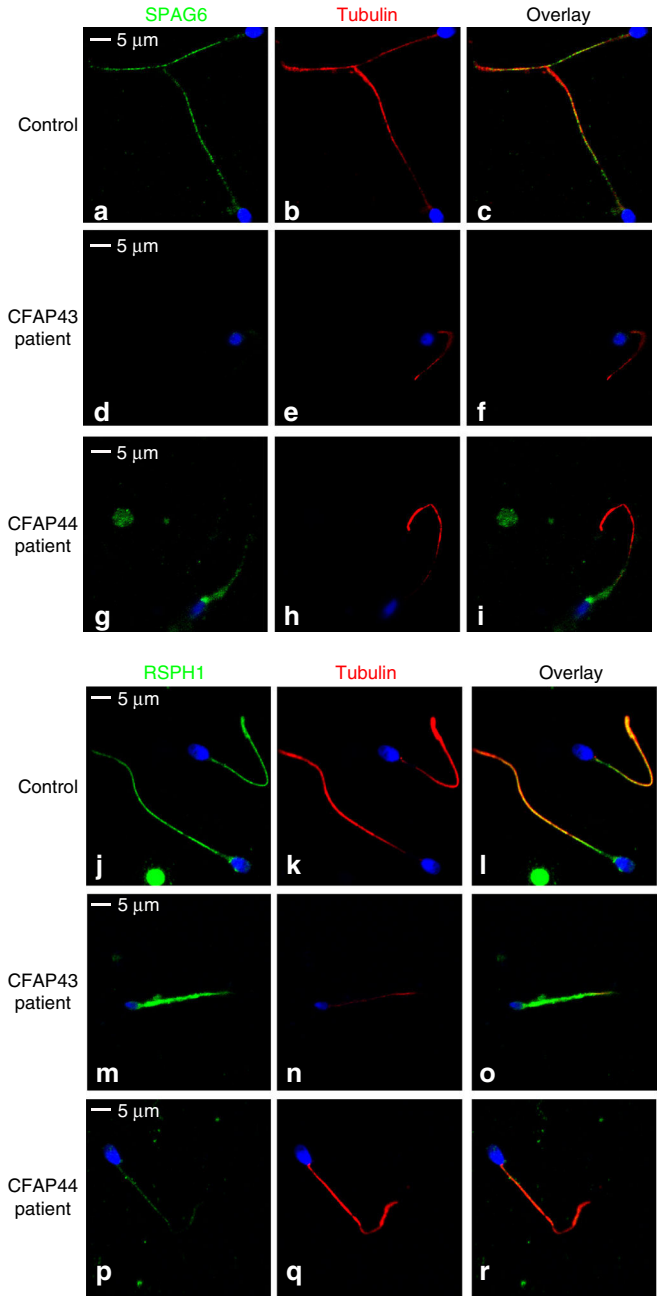


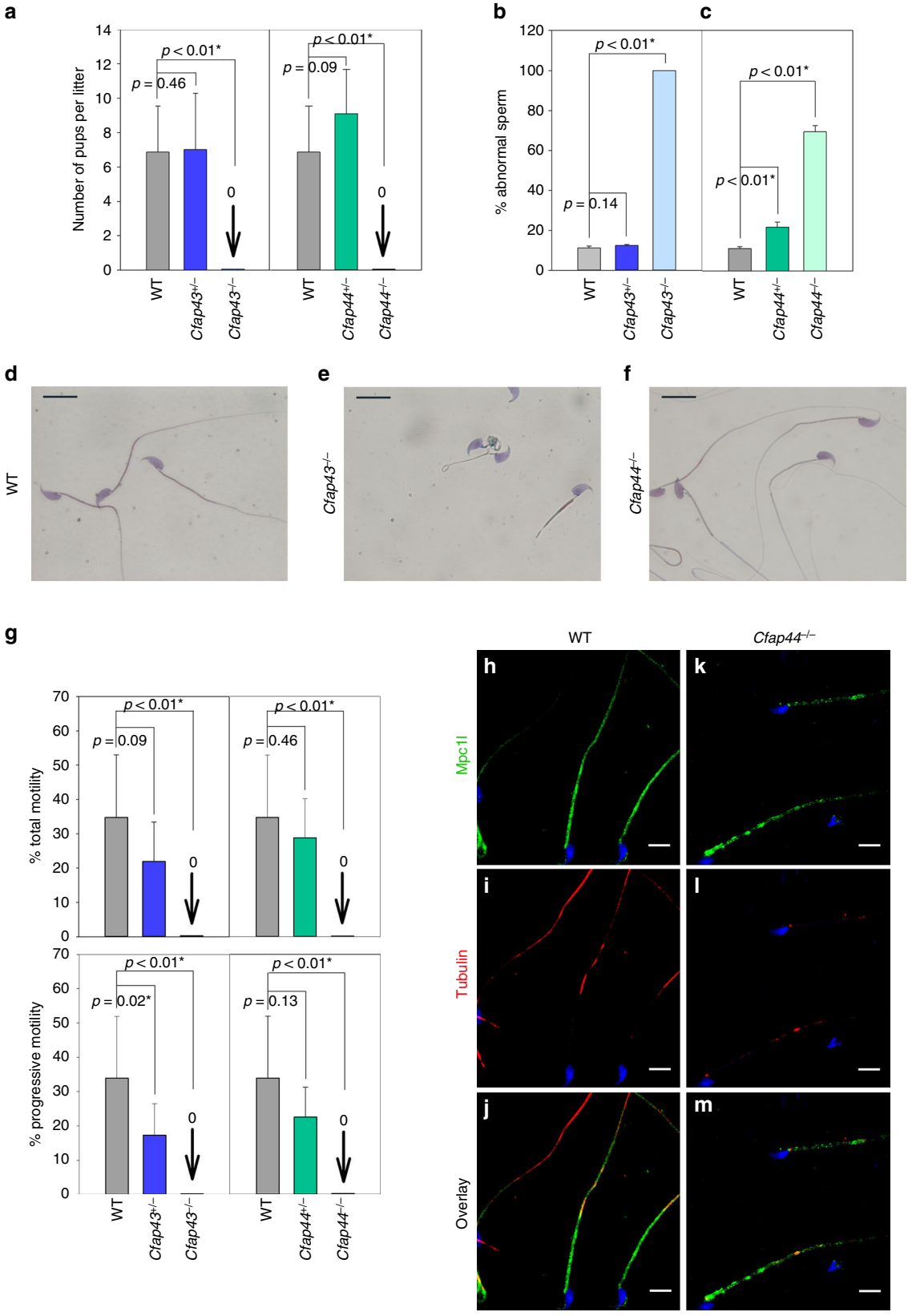
Fig. 3 Immunofluorescence staining in *CFAP43* and *CFAP44* patients reveals an abnormal axonemal organization. **a**–**c** sperm cells from a fertile control stained with anti SPAG6 (green), a protein located in the CPC, and anti-acetylated tubulin (red) antibodies. DNA was counterstained with Hoechst 33342. **c** Corresponds to **a** and **b** overlay and shows that in control sperm, SPAG6 and tubulin staining superimpose. Scale bars = 5 µm. **d**–**f** SPAG6 staining is absent in sperm from the patient P_{43}-5 homozygous for the c.2658C>T variant in *CFAP43*. **d**–**i** Similar IF experiments performed with sperm cells from the patient P_{44}-2 homozygous for the c.3175C>T variant in *CFAP44*. Scale bar = 5 µm. Contrary to the control, the SPAG6 immunostaining (green) is abnormal with a diffuse pattern concentrated in the midpiece of the spermatozoa and is not detectable in the principle piece. **j**–**l** Sperm cells from a fertile control stained with anti RSPH1 (green), a protein of the radial spoke's head, and anti-acetylated tubulin (red) antibodies. DNA was counterstained with Hoechst 33342. **l** corresponds to **j** and **k** overlay and shows that RSPH1 and tubulin staining superimpose in control sperm. Scale bar = 5 µm. **m**–**o** In sperm from the patient P_{43}-5, RSPH1 staining (green) is significantly different from control (**m**) with a marked diffuse staining. **p**–**r** In sperm from the patient P_{44}-2 the intensity of the RSPH1 staining is strongly reduced

within the normal values for the mouse. We next studied sperm morphology and observed that for *Cfap43*$^{-/-}$ males, 100% of sperm displayed a typical human MMAF phenotype with short, thick, and coiled flagella (Fig. 4b); although slightly deformed, sperm head exhibited an overall normal hooked form (Fig. 4e). In contrast, sperm from *Cfap44*$^{-/-}$ males had normal flagellum length, but most of them showed abnormal forms and irregular caliber of the midpiece (Fig. 4c–f). In sperm from *Cfap44*$^{-/-}$ mice, staining of Mpc1l, a sperm mitochondrial protein[12], was discontinuous or punctiform, strongly suggesting that the distribution of the mitochondria surrounding the axoneme was irregular, consistent with the observed midpiece abnormalities (Fig. 4h–m). For both *Cfap43*$^{-/-}$ and *Cfap44*$^{-/-}$ sperm, the observed structural flagellum defects were associated with severe motility deficiencies (Fig. 4g) with sperm from *Cfap43*$^{-/-}$ males showing complete immotility whereas those from *Cfap44*$^{-/-}$ sometime presented some small vibrations (Supplementary Movie 1), in contrast to the motile sperm from *Cfap44*$^{+/-}$ heterozygotes (Supplementary Movie 2). Interestingly, we observed that in heterozygous animals, *Cfap43*$^{+/-}$ mice had an impaired progressive sperm motility and *Cfap44*$^{+/-}$ mice had an increased proportion or morphologically abnormal spermatozoa (Fig. 4g).

To better characterize the molecular defects induced by the absence of Cfap43 and Cfap44 in mouse sperm, we studied by IF the presence and localization of several proteins belonging to different substructures of the axoneme in sperm from both KO mouse models. The presence of the following proteins was investigated: Dnah5 and Dnali1 as markers of dynein arms, Rsph1 and Rsph4a as markers of radial spokes, Gas8 as a marker of nexin links, and Spef2 as a marker of the central pair (Supplementary Table 2—Mouse). In *Cfap43*$^{-/-}$-mutated animals, Spef2, Rsph1, and Rsph4a were clearly missing (Fig. 5d–f, j–l) as compared to controls (Fig. 5a–c, g–i, Supplementary Table 2, and Supplementary Fig. 4), indicating that the center of the axoneme, including the head of the radial spoke interacting with the central pair, was absent. In contrast, no obvious defects were observed in sperm from *Cfap44*$^{-/-}$, suggesting that all substructures were present (Supplementary Table 2—Mouse).

The impact of the absence of these proteins on the flagellum ultrastructure was analyzed on sperms from *Cfap43*$^{-/-}$ and *Cfap44*$^{-/-}$ by TEM. In the midpiece of WT mouse flagellum, the 9+2 axoneme is surrounded by nine ODFs and the mitochondria sheath[13]. The ODFs 1, 5, 6, and 9 are parallel and aligned with the central pair (Fig. 5m, see also Fig. 6a and Supplementary Fig. 5a for the DMT annotation). In addition, a transversal complex is composed of the central pair and DMTs 3 and 8. This complex includes ODFs 3 and 8 in the midpiece and is closely associated with the FS in the PP through two stalks emerging from the longitudinal column and linking DMTs 3 and 8[14]. In the PP, the axoneme is surrounded by seven ODFs (3 and 8 are missing) and by an FS containing two longitudinal columns. Contrary to what is usually observed in human sperm, this organization is highly reproducible and <5% of WT sections from mouse sperm flagellum present structural defects (Supplementary Fig. 6a). Analyses of longitudinal sections from *Cfap43*$^{-/-}$ sperm showed that the observed short tail actually corresponds to a large cytoplasmic mass containing unorganized structural components of the flagellum (Supplementary Fig. 7). When the axoneme was present, all transversal sections observed revealed a substantial structural disorganization with uneven distribution of the nine DMTs and the absence of the CPC (Fig. 5n). In contrast, in *Cfap44*$^{-/-}$ sperm, the 9+2 organization of the axoneme was preserved in 70% of cases (Fig. 6b, d, e, g–i). The observed defects included the absence of peripheral doublets, external shift, or mislocalization of the central pair and distorted circular

distribution of the DMTs with CPC misorientation (Fig. 6d–f). *Cfap44*$^{-/-}$ sperm also exhibited important defects of the ODFs and of the FS (Supplementary Fig. 6b). In the midpiece, the number of dense fibers was significantly higher and their localization and orientation were defective (Fig. 6b, c). In the PP, ODFs 3 and 8 were abnormally retained, preventing normal anchoring of the FS's stalks of on DMTs 3 and 8 (Fig. 6g–i). The longitudinal columns also presented several defects: they were misaligned and not facing the 3-central-8 complex, leading to notable asymmetry of the structure (Fig. 6k–n), which was



reinforced in numerous sperm by the presence of a third longitudinal column (Fig. 6m, n). Finally, the defects detected in *Cfap44*$^{-/-}$ MS by optic and fluorescence microscopy were specified by TEM analyses of sperm midpieces, evidencing uneven distribution and fragmentation of the mitochondria (Supplementary Fig. 8).

TbCFAP43 and TbCFAP44 are necessary axonemal proteins. To better characterize CFAP43 and CFAP44 localization and function and to overcome the absence of reliable antibodies in human and murine cells, we decided to use *Trypanosoma brucei* (*T. brucei*) as an additional model organism because this flagellated protozoan has largely contributed to elucidating the molecular pathogeny of human ciliopathies[15]. The *T. brucei* axoneme is similar to that of mammalian flagella, but its peri-axonemal structure, although comparable in function, is different. Instead of the FS and the ODFs found in mammals' flagella, it contains the paraflagellar rod (PFR), a complex structure connecting with the axonemal doublets 4–7[16,17] which plays a role in flagellum motility[18,19] and serves as a platform for metabolic and signaling enzymes[20–22]. In *T. brucei* the DMTs are numbered from the doublet opposite the PFR (no. 1), which is normally parallel to the central singlets. The doublets are then numbered clockwise, doublets 5 and 6 facing the PFR and doublets 4 and 7 being positioned on each side of the PFR (Supplementary Fig. 5b for DMT annotation and relative PFR localization).

BLASTp analysis on *T. brucei* genome database[23] using human CFAP43 and CFAP44 sequences identified *T. brucei* orthologs Tb927.7.3560 (named TbCFAP44 in this study) and Tb927.4.5380 (named TbCFAP43 in this study), respectively. The sequence identity, calculated with clustal omega[24], between the four proteins shows that CFAP44 is closest to TbCFAP44 with 25.6% sequence identity and CFAP43 closest to TbCFAP43 with 20.8% sequence identity. Previous functional genomics and proteomic studies identified TbCFAP44 and TbCFAP43 as flagellar proteins[25,26]. In addition, TbCFAP44 is the *Chlamydomonas* FAP44 ortholog, and is involved in flagellar motility (also referred to as *T. brucei* components of motile flagella 7 (TbCMF7 in ref. [27]). However, the function of TbCFAP44 and TbCFAP43 are currently unknown. To confirm that both *Trypanosoma* proteins are the orthologs of the mouse and human proteins, we compared their secondary structure. The secondary structure predicted by Porter 4.0[28] shows that the four proteins share a common general motif with β-strand domains in the first half of the protein and α-helical domains in the second half (C-Ter) (Supplementary Fig. 9). We note that in CFAP43 and TbCFAP43, all residues between amino acids 719 and 1,657 and 631 and 1,446, respectively, are predicted to form α-helices. Interestingly, analyses using Superfamily[29,30] indicated that long α-helices stretches such as those identified in CFAP43 (amino acids 719–1,657) have not yet been reported in any known fold. According to SCOP[31], the N-terminal part of all four proteins (almost half of the proteins) belongs to the class of "All β-proteins" and "7-bladed β-propeller fold." These N-terminal domains also belong to the WD40 repeat-like superfamily (Supplementary Table 3). This structural comparative analysis indicates that TbCFAP44 and TbCFAP43 and their human orthologs share unique structural similarities and therefore demonstrates that these proteins belong to the same family.

We localized TbCFAP44 and TbCFAP43 in bloodstream from *T. brucei* (BSF, present in the mammalian host), using 10TY1-tagged and 3myc-tagged proteins by generating *T. brucei* cell lines expressing endogenous levels of N-terminal 10TY1-tagged proteins ($_{10TY1}$TbCFAP44 and $_{10TY1}$TbCFAP43) and C-terminal 3myc-tagged proteins (TbCFAP44$_{myc}$ and TbCFAP43$_{myc}$). TY1 tag (EVHTNQDPLDGS, 1.8 kDa) is used as an epitope for immunolabeling[32] used for numerous immunolocalizations in trypanosomes, either as 1xTY1, 3xTY1, or, more recently, 10xTY1[33]. Both proteins, whatever the tag used, were found in the axoneme as substantiated by co-labeling with an antibody against the PFR structure (Supplementary Figs. 10a, b). We acknowledge that the tag might interfere with protein function and localization; however, this result suggests that the 10TY1 tag did not interfere with the trafficking of either proteins which extended throughout the whole flagellum length, as demonstrated by the 3myc and 10TY1 labeling preceding the PFR at the posterior end of the flagellum (next to the kinetoplast) and along the flagellum up to its anterior end. To further investigate this localization we used the 10TY1 tag which produces a strong IF staining[32] necessary for stimulated-emission-depletion (STED) confocal microscopy[34] and created two double-labeled cell lines $_{10TY1}$*TbCFAP44/TbCFAP43*$_{myc}$ and $_{10TY1}$*TbCFAP43/TbCFAP44*$_{myc}$. Triple labeling STED confocal microscopy of 10TY1-tagged proteins, tubulin, and PFR2 on permeabilized whole cells demonstrated that $_{10TY1}$TbCFAP44 and $_{10TY1}$TbCFAP43 are closely associated to the axoneme along the PFR (Fig. 7a–f, respectively); similar results were obtained on detergent-extracted cytoskeletons. To confirm this remarkable localization, TEM immunogold labeling was performed using differently tagged proteins. We used the myc tag, a small tag of 10 amino acids which is known to have minimum effect on localization. Concordant with the STED microscopy results, all immunogold beads were located only along the axoneme and facing the PFR, therefore indicating that TbCFAP44 and TbCFAP43 are in direct contact only with the doublets facing the PFR, which correspond to DMTs 5 and 6[35] (Fig. 7g–j). The intervals between two adjacent gold particles were measured and the minimal distance between two particles was around 25 nm for both proteins (Fig. 7g, h, inset). Finally, 10TY1-tagged proteins presented an identical subcellular localization in TEM (Supplementary Fig. 11)

To evaluate the impact of the absence of TbCFAP44 and TbCFAP43 on the structure of the trypanosome flagellum and its beating, we knocked down protein expression by inducible RNA interference (RNAi), either in the parental cell line (cell lines *TbCFAP44*^{RNAi}, *TbCFAP43*^{RNAi}) or in the cell lines expressing the myc-tagged proteins (cell lines *TbCFAP44*$_{myc}^{RNAi}$, *TbCFAP43*$_{myc}^{RNAi}$) (Fig. 8). Efficiency and specificity of RNAi knockdown of *TbCFAP44* and *TbCFAP43* were validated by

Fig. 4 Reproductive phenotype of homozygous and heterozygous *Cfap43* and *Cfap44* male mice. **a** Fertility of *Cfap43*$^{+/-}$, *Cfap44*$^{+/-}$ *Cfap43*$^{-/-}$, and *Cfap44*$^{-/-}$ males. Heterozygous and homozygous mutant males were mated with WT females and the numbers of pups per litter were measured. KO males were completely sterile. **b**, **c** Spermatocytograms showing the number of abnormal sperm in heterozygous and homozygous mutant males. **d**–**f** Images of typical sperm stained with Harris–Shorr from *Cfap43*$^{-/-}$ and *Cfap44*$^{-/-}$ males. Scale bars = 10 µm. **g** Total and progressive motilities of sperm from *Cfap43*$^{-/-}$ and *Cfap44*$^{-/-}$ males. **h**–**m** The mitochondria sheath is fragmented in *Cfap44*$^{-/-}$ males. Staining of WT sperm with an anti-MCPI1 (**h**, green), a mitochondrial transporter, and anti-acetylated tubulin (**i**, red) antibodies. **j** Overlay of MCPI1 and tubulin staining. Sperms were counterstained with Hoechst (blue). **k**–**m** Similar experiments on sperm from *Cfap44*$^{-/-}$ males. Scale bars = 10 µm. **a**–**c**, **g** Data represent mean ± SD; statistical differences were assessed with *t* test, *P* value as indicated

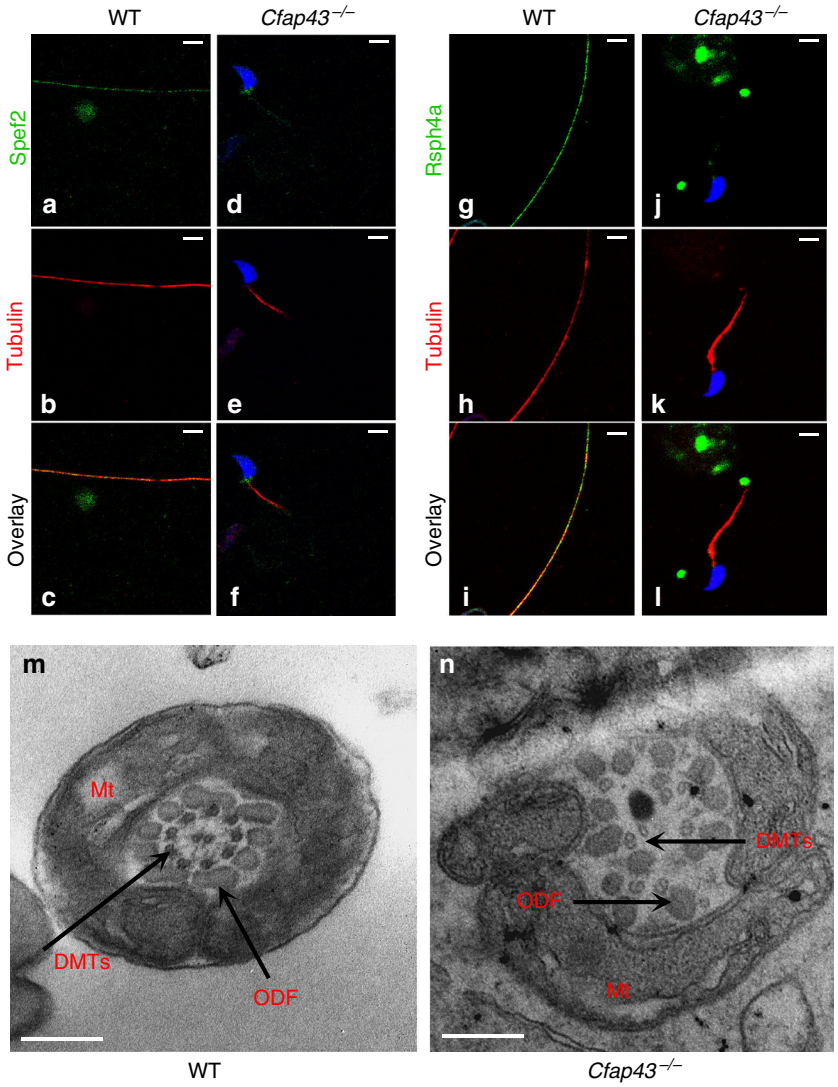


Fig. 5 The sperm axoneme from *Cfap43*$^{-/-}$ males is fully disorganized. **a** Central pair is absent in sperm from *Cfap43*$^{-/-}$ males. Staining of WT sperm with an anti-Spef2, a marker of the projection 1b of singlet C1 (**a**, green) and anti-acetylated tubulin (**b**, red) antibodies. **c** Overlay of Spef2 and tubulin staining. Sperms were counterstained with Hoechst (blue). **d**–**f** Similar experiments on sperm from *Cfap43*$^{-/-}$ males. Scale bars = 5 µm. **g**–**l** Head of radial spoke are absent in sperm from *Cfap43*$^{-/-}$ males. Staining of WT sperm with an anti-Rsph4a, another protein of the RS head (**g**, green) and anti-tubulin (**h**, red) antibodies. **i** Overlay of Rsph4a and tubulin staining. Sperm were counterstained with Hoechst (blue). **j**–**l** Similar experiments on sperm from *Cfap43*$^{-/-}$ males. Scale bars = 5 µm. **m**, **n** Transversal section of a sperm from WT (**m**) and *Cfap43*$^{-/-}$ (**n**) males observed by EM in the midpiece region. Note the specific arrangement of the ODFs around the axoneme in WT sperm and the complete disorganization of the DMTs and the absence of the central pair in *Cfap43*$^{-/-}$ sperm. Scale bars = 240 nm. Number of cross-section for WT and *Cfap43*$^{-/-}$ were 21 and 30, respectively. All cross-section were defective for *Cfap43*$^{-/-}$. DMTs, doublet of microtubules; ODF, outer dense fiber; Mt, mitochondria

RT-PCR (Supplementary Fig. 12) and by IF, showing a clear decrease of myc labeling in the new flagellum of *TbCFAP44*$_{myc}^{RNAi}$-induced and *TbCFAP43*$_{myc}^{RNAi}$-induced cells (Fig. 8e–l, respectively). Cell proliferation was assessed in parental, non-induced, and induced *TbCFAP44*^{RNAi} and *TbCFAP43*^{RNAi} cells (and in *TbCFAP44*$_{myc}^{RNAi}$ and *TbCFAP43*$_{myc}^{RNAi}$ cell lines, showing the same results). In all cases, both *TbCFAP44*^{RNAi}-induced and *TbCFAP43*^{RNAi}-induced cells showed abnormal beating (Supplementary Movies 2–5) and stopped proliferating after 24 h and eventually died (Fig. 8m). These growth defects were accompanied by a defect in cytokinesis producing multi-flagellated cells (Supplementary Fig. 13), a phenotype previously described when proteins directly or indirectly involved in flagellar motility are knocked down in BSF[36]. We observed that the absence of TbCFAP44$_{myc}$ and TbCFAP43$_{myc}$ did not impact flagella length (Fig. 8n). Overall, these results indicate that TbCFAP44 and TbCFAP43 are necessary for flagellar function and thus cell proliferation in *T. brucei* BSF.

To assess if the observed cellular phenotypes are caused by each protein independently or if the absence of one induces the absence of the other, we took advantage of the double-labeled cell lines $_{10TY1}$*TbCFAP44/TbCFAP43*$_{myc}$ and $_{10TY1}$*TbCFAP43/TbCFAP44*$_{myc}$. Inactivation of TbCFAP43 by RNAi did not alter the amount and localization of TbCFAP44; conversely, inactivation of TbCFAP44 had no impact on TbCFAP43 protein (Supplementary Fig. 14).

The impact of *TbCFAP44* and *TbCFAP43* knockdown was also investigated by TEM. In control longitudinal sections of parental cells, the flagellum exits the cell through the flagellar pocket (Fig. 8o), and, in cross-section, the canonical ultrastructure of the

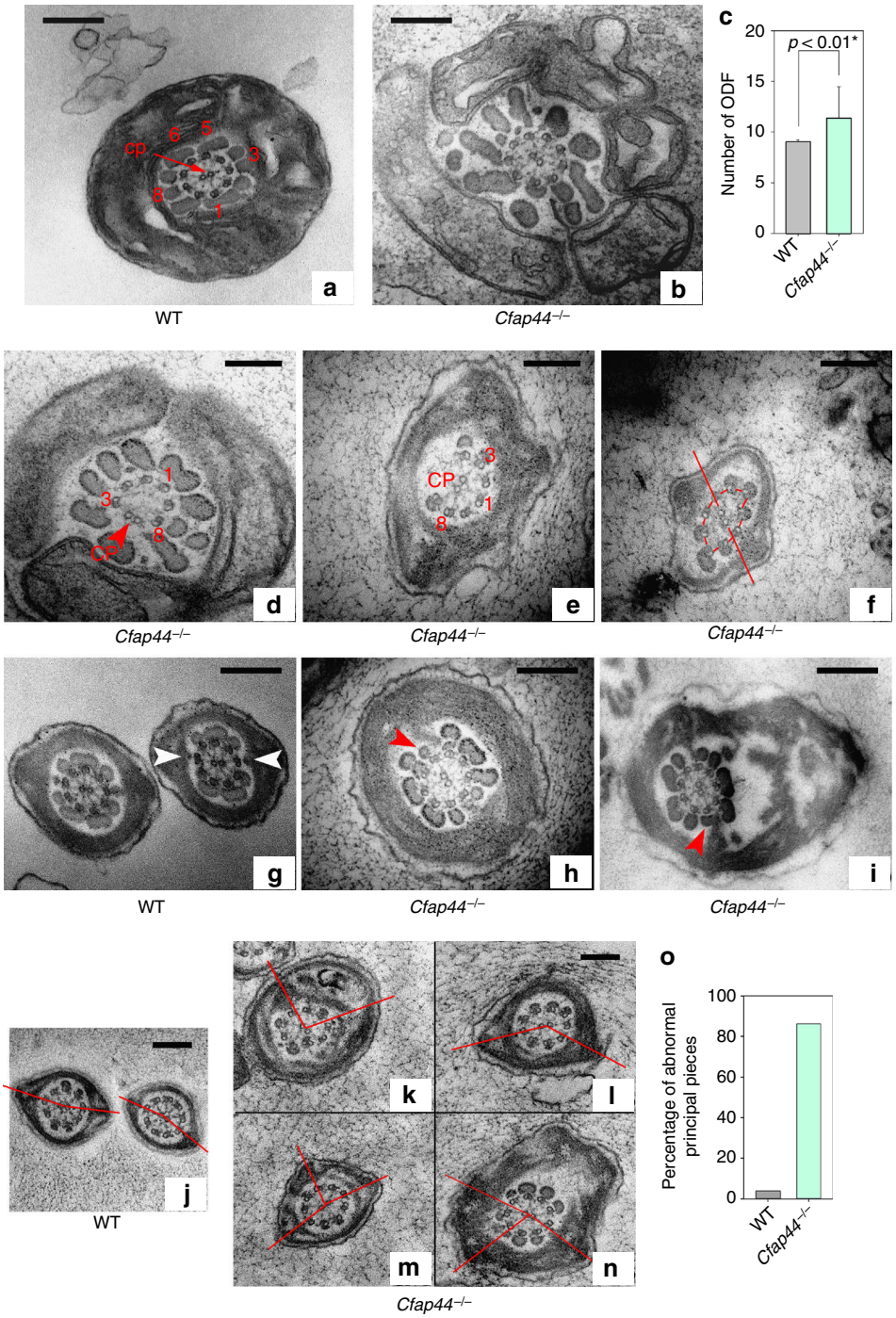


Fig. 6 Electron microscopy cross-sections of sperm from *Cfap44*$^{-/-}$ males reveal multiple structural axonemal defects. **a** Cross-section of the midpiece of a WT sperm, showing the arrangement of the ODFs around the axoneme. **b** Presence of extra ODFs in midpieces sections of sperm from *Cfap44*$^{-/-}$ males. The orientation of ODFs is also defective, leading to an increase of the midpiece diameter. **a**, **b** Same scale bars = 250 nm. **c** Graph showing the increased number of ODF in the mutant. Data represent mean ± SD; the statistical difference was assessed with *t* test, *P* value as indicated. Twenty-one cross-sections in the midpiece region were analyzed for WT and 21 for *Cfap44*$^{-/-}$. **d**–**f** Various structural defects of the axoneme in sperm from *Cfap44*$^{-/-}$ males. **d** Four DMTs (4–7) were missing. The central pair is shifted at the periphery (red arrowhead). **e** DMTs 5–7 were missing. Note the presence of a third longitudinal column (LC). **f** Irregular distribution of the DMTs associated with a rotation of the central pair (straight red line). Scale bars = 196 nm. **g** In WT sperm, the fibrous sheath is linked to the 3-central-8 complex by stalks emerging from the longitudinal columns (white arrowheads). **h**, **i** The presence of extra ODFs facing DMTs 3 and 8 (red arrowheads) prevents a normal anchoring of the fibrous sheath's stalks on DMTs 3 and 8. Scale bars = 270 nm. **j**–**o** LCs are not aligned with 3–8 CPC axis. In contrast to WT, where the 3-central-8 complex is aligned with LC to form almost a straight line (**j**, red line), LC are misaligned in sperm from *Cfap44*$^{-/-}$ males, leading to notable asymmetry of the structure (**k**, **m**, red lines). **l**, **n** The presence of a third LC increases asymmetry. Scale bars = 184 nm. **o** Bar graph showing the % of defects observed in the principal piece as described in **g**–**n**. One hundred cross-sections in the principal piece region were analyzed for WT and 50 for *Cfap44*$^{-/-}$

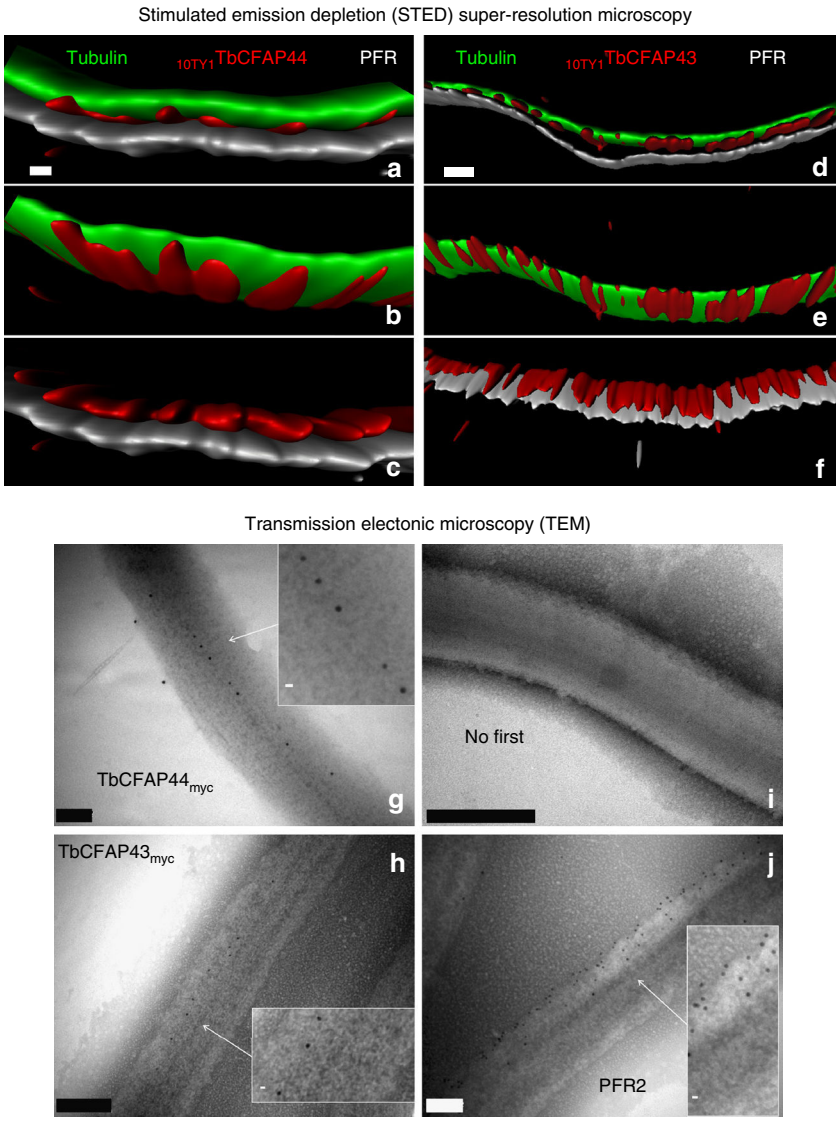


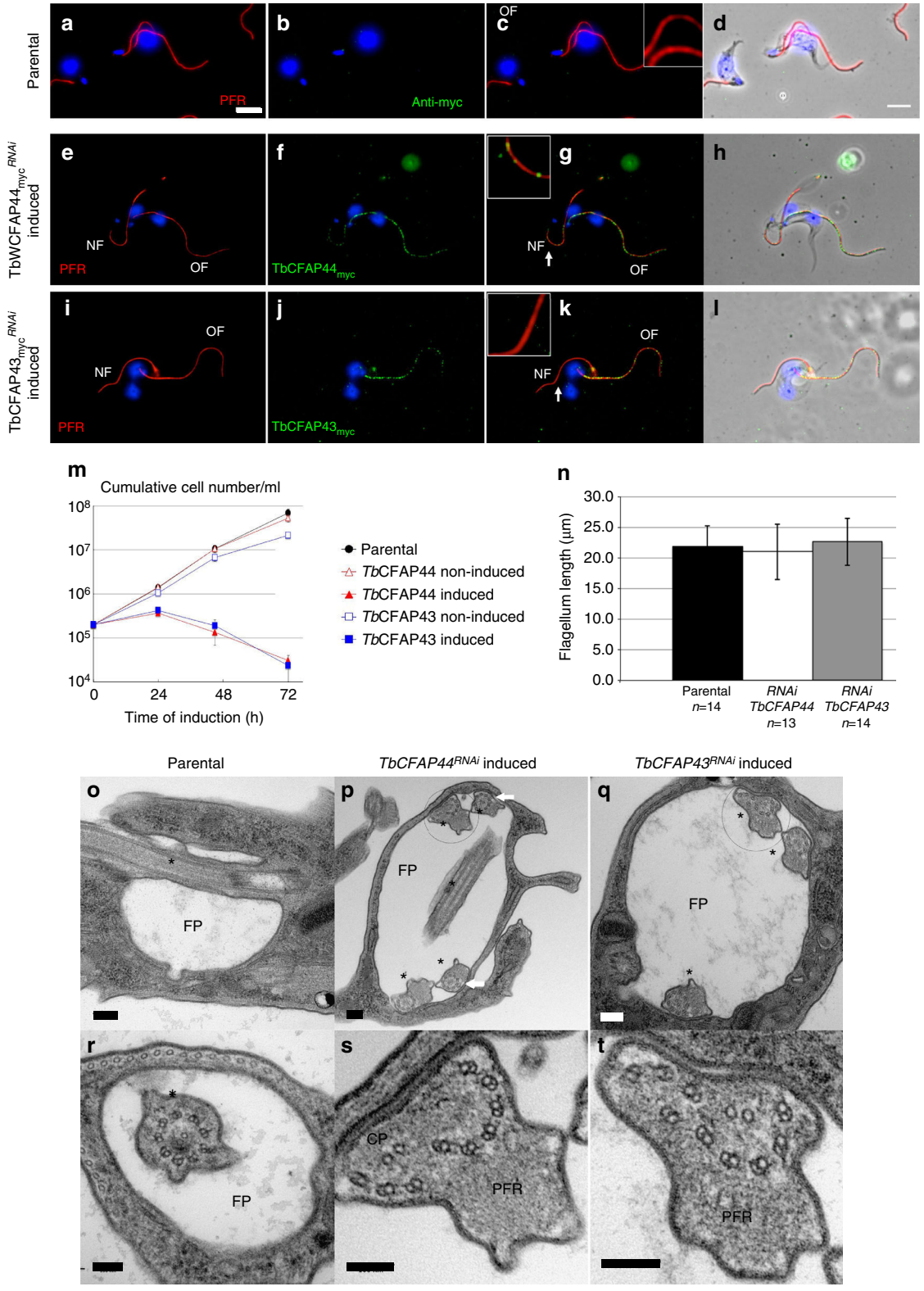
Fig. 7 Orthologs of *T. brucei* TbCFAP44 and TbCFAP43 are flagellar proteins apposed to DMTs 5–6. **a**–**f** STED confocal microscopy, deconvolution, and 3D reconstruction views of the co-labeling, on permeabilized whole cells, of tubulin (green), PFR2 (gray), and $_{10TY1}$TbCFAP44 (**a**–**c**) and $_{10TY1}$TbCFAP43 (**d**–**f**) showing that TbCFAP44 and TbCFAP43 are facing the PFR and are closely associated with the tubulin, which likely corresponds to DMTs 5–6. Scale bars=0.5 µm. **g**–**j** Electron immunogold labeling of cytoskeleton-extracted cells expressing TbCFAP44myc (**g**) and TbCFAP43myc (**h**). Controls with no primary antibody (**i**) and anti-PFR2 (**j**). Gold beads size was either 6 nm (**g**, **h**) or 10 nm (**i**, **j**). Scale bars=200 nm in **g**, **h**; 100 nm in **i**, **j**. Insets are enlargement of images of flagella taken from the main panels and scale bars=10 nm

axoneme, composed of nine DMTs and the CPC, is observed (Fig. 8r). In control cell line only a small proportion (0.6%) of axonemes ($n = 172$) were abnormal. In contrast, *TbCFAP44*^{RNAi}-induced and *TbCFAP43*^{RNAi}-induced cells were abnormal and displayed more than two flagella in one abnormally enlarged flagellar pocket (Fig. 8p, q). In cross-section, flagella exhibited numerous abnormal axonemes with 90° rotated CPC, a defect never observed in control trypanosome cell lines (Fig. 8p) and mainly displaced CPC and DMTs (Fig. 8s, t). Overall 27.3% of *TbCFAP44*^{RNAi} ($n = 139$) and 27.7% of the *TbCFAP43*^{RNAi} ($n = 150$) showed axonemal defects. We could not exclude some minor modifications in PFR; however, the disruption of the axoneme was so extreme after TbCFA44 or TbCFA43 RNAi knockdown that it was difficult to assess them. Altogether, these results confirm the essential role of TbCFAP43 and TbCFAP44 in the organization of the axoneme ultrastructure.

Discussion

Our work illustrates the efficiency of the combination of WES with an original workflow for the validation of the candidate genes. We identified two new genes responsible for MMAF syndrome and several new *DNAH1* mutations, reinforcing the importance of this gene in MMAF syndrome. The prevalence of *DNAH1* mutations was lower than what was reported previously, maybe due to a wider geographic recruitment of patients in the current study[5,6]. Altogether, *DNAH1*, *CFAP43*, and *CFAP44* mutations were identified in 28.2% of the analyzed subjects ($n = 78$) originating from North Africa, Europe, and the Middle East. These results underline the importance of these three genes in MMAF syndrome and will permit to improve the diagnosis efficiency of infertile MMAF patients. To investigate the possibility of a potential genotype–phenotype correlation, we examined the semen parameters of patients carrying mutations in *DNAH1*, *CFAP43*, and *CFAP44* and MMAF patients with yet

unidentified mutations. There were not differences between the first three groups except for a higher rate of spermatozoa with short and absent flagella and a lower motility rate evidenced in the mutated subjects compared to patients with no identified mutation, highlighting the severity of mutations in these three genes. The identified mutations are distributed throughout the whole CFAP43 and 44 genes indicating that the entirety of both proteins is necessary for preserving their functionality. Interestingly among *CFAP43*-mutated patients, only one (P43-9) carried a missense variant, with potentially a milder effect (Val347Ala). This patient has the lowest percentage of sperm with short flagella (22%) with 10% of morphologically normal spermatozoa (versus 0% for the other *CFAP43*-mutated subjects) and therefore presents a milder phenotype compared to the others with



homozygous truncating mutations. These data support the existence of a genotype–phenotype correlation as had been demonstrated for *DNAH1*-mutated subjects[5,6]. We have previously demonstrated that MMAF patients with *DNAH1* mutations had a good prognosis using ICSI[37], but it remains difficult to predict the success rate for the other MMAF genotypes. Additional correlation studies should now be performed to take into account the individual genotypes in the counseling of MMAF patients. Exome sequencing of MMAF patients permitted a diagnosis efficiency of 27% for the three validated genes. Subsequent analysis of exome data from the remaining 56 subjects permitted to identify 7 additional candidate genes with bi-allelic mutations found in at least two individuals in a total of 23 subjects. The identified mutations and these additional genes are currently being investigated. If all these variants are confirmed to be deleterious, a diagnosis will be obtained for 45/78 patients (58%), confirming the interest of WES as a diagnostic tool for MMAF syndrome.

Identification of disease-causing genes has recently been catalyzed by high-throughput genome-wide sequencing technologies. New challenges are now arising, in particular, the validation of candidate genes in the investigated phenotypes. CRISPR/Cas9 is a very efficient and fast technique to create KO models[38], compatible with the increased number of mutations/genes found by high-throughput sequencing. To confirm the pathogenicity of the identified mutations, we produced two knockout mice lines using CRISPR/Cas9 technology. *Cfap43*$^{-/-}$ males were infertile and 100% of their spermatozoa were immotile and morphologically abnormal with short and coiled tail, a phenotype very similar to human MMAF (Fig. 4e). *Cfap44*$^{-/-}$ male mice were also infertile due to flagellar immotility, yet presenting subtler flagellar defects than those observed in patients with *CFAP44* mutations (Fig. 4f). Such a phenotypic discrepancy between mouse and human is not uncommon. For example, *Dnah1* KO mice display asthenozoospermia without morphological defects of the flagellum[39], whereas in humans *DNAH1* truncating mutations induce a MMAF phenotype[5]. Nevertheless, TEM analysis of *Cfap44*$^{-/-}$ sperm evidenced ultrastructural defects similar to those found in *CFAP44* patients, confirming that CFAP44 is necessary for axonemal organization and function both in mouse and human. The reason why this axonemal disorganization does not impact the overall flagellum morphology and length in mice is not known. It is however highly unlikely that a functional truncated protein could persist and have a rescue effect in KO mice since they are predicted at best to produce a truncated peptide of 49 out of the 1,854 amino acids of the full-length protein (Supplementary Fig. 3). Overall, the tandem use of WES and CRISPR/Cas9 technology in mouse was a very efficient strategy to identify and validate the mutations responsible for infertility phenotypes.

The structure and organization of motile cilia or flagellum was mainly deciphered from models including the green alga *Chlamydomonas reinhardtii*, the protozoa *T. brucei* and *Tetrahymena*, sea urchins, zebrafish, *Xenopus*, and mouse[15,40]. Such a wide selection of distant models is possible because motile cilia or flagella are built on a canonical 9+2 axoneme which forms a highly organized protein network remarkably conserved during evolution[41]. Here we chose to use *T. brucei* as this model has two advantages, first, in contrast to *Chlamydomonas* and other models, it shares some specific structural axonemal characteristics with mammalian flagellum such as a full RS3 and a fixed orientation of the central pair during flagellum beating, and second, it allows forward and reverse genetics for an easier characterization of gene function. BlastP enabled us to identify CFAP43 and CFAP44 likely orthologs in *T. brucei* with good sequence identity. Interestingly, in-depth structural analyses showed that these proteins share some unique structural similarity, in particular, a common general motif with β-strand domains in the N-terminal region followed by α-helical domains in the C-terminal region. These results confirm the relatedness of the identified orthologs and suggest that CFAP43 and CFAP44 proteins may have a similar function as reported for others proteins with similar structures[42]. Consistent with this, we showed that in *T. brucei*, inactivation of one protein by RNAi did not alter the amount and localization of the other protein, indicating that (i) the phenotype observed when invalidating each protein is independent of the other protein, (ii) the presence of one cannot compensate for the absence of the other, and (iii) if the two proteins interact with one another, they do not depend on each other for adequate positioning and docking.

In all studied models (i.e. humans, mouse, and *T. brucei*), the invalidation of each protein induced important axonemal disorganization. It is important to underline that the destabilization of the axoneme can be triggered by the absence of proteins located either in axonemal or peri-axonemal structures[43,44], making it difficult to assess the function from the observed defects. Only the localization at the ultrastructural level would-provide a clue on protein function. For this purpose, we generated several *T. brucei* cell lines expressing endogenously tagged proteins, and studied their localization by STED and immuno-EM. Using two different tags, we showed that both proteins have a restricted location within the flagellum, closely associated with the part of the axoneme facing the PFR (Fig. 7) and most probably corresponding to DMTs 5 and 6[35]. This localization rules out a potential role of both proteins in the intraflagellar transport

Fig. 8 *T. brucei* CFAP44 and CFAP43 are necessary for cell survival and proper axonemal organization. **a**–**d** Immunofluorescence on detergent-extracted cells of parental *T. brucei* stained with anti-PFR (red) and anti-myc (green) antibodies. **e**–**h** Immunofluorescence on detergent-extracted cells expressing TbCFAP44$_{myc}$ and RNAi induced (24 h) for *TbCFAP44* (*TbCFAP44*$_{myc}^{RNAi}$). Cells showed no or weak myc labeling (green) on the new flagellum (NF), while the old flagellum (OF) remained labeled. PFR is labeled in red. Note: cells with a maximum of two flagella were chosen for clear imaging. **i**–**l** Similar RNAi experiments as performed in **e**–**h** for *TbCFAP43* RNAi in cells expressing TbCFAP43$_{myc}$ (*TbCFAP43*$_{myc}^{RNAi}$). Cells showed no or weak myc labeling (green) on the new flagellum (NF), while the old flagellum (OF) remained labeled. PFR is labeled in red. Nuclei and kinetoplasts (mitochondrial genome) are labeled with DAPI (blue). Scale bars=5 µm. Insets are enlargement of images of flagella taken from the main panels and display areas indicated by white arrows (scale bar=1 µm). **m** Growth curves for parental cells, and *TbCFAP44*^{RNAi} and *TbCFAP43*^{RNAi} cells, non-induced or induced with tetracycline. The graph represents the cumulative number of cells per ml. Error bars represent the standard error from three independent experiments. **n** Flagellum length measurement of the new flagellum in cells bearing two flagella from parental cells, and from cells expressing TbCFAP44$_{myc}$ and TbCFAP44$_{myc}$ and induced (24 h) for *TbCFAP44* RNAi and *TbCFAP43* RNAi, respectively. Flagellum length was measured in cells showing a clear decrease in myc labeling in the new flagellum. **o**–**t** Electron micrographs of stained thin sections of parental cells (**o**, **r**) bearing one flagellar pocket (FP) and one flagellum (*), and of 24 h induced *Tb*CFAP44^{RNAi} (**p**, **s**) and *Tb*CFAP43^{RNAi} (**q**, **t**) cells. **s** and **t** are enlargements of **p** and **q**, respectively. Scale bars=100 nm in **o**–**q**; 200 nm in **r**–**t**. Note: in *TbCFAP44*^{RNAi} and *TbCFAP43*^{RNAi} cells, the flagellar pocket is enlarged and bears more than two flagella (**p**, **q**). Some of these flagella present axonemal defects including displaced and rotated CPC and shifted DMTs (**p**, white arrows, **s**, **t**)

(IFT) since IFT in *T. brucei* flagella was never described between the PFR and the axoneme and is circumscribed to two sets of doublet microtubules 3–4 and 7–8, located on each side of the PFR[45]. Moreover, the absence of cytoplasmic material accumulated at the basis of the flagellum and in the flagellar pocket and the normal length of the flagellum in both RNAi mutants (Fig. 8) do not support the IFT hypothesis[46].

Interestingly, it has been reported that two specific structures are present around DMTs 5 and 6: the first one is proposed to consolidate the interaction between the axoneme and the para-axonemal components[47] and the second one to specifically strengthen DMTs 5–6 interaction, known as the 5–6 bridge[11], made of inner and outer subunits. Both structures, the 5–6 bridge and the connecting proteins linking the axoneme to the PFR, are known to present an intervallic pattern[11,47]. This subcellular distribution is in agreement with our super-resolution and immuno-EM results which showed a minimum interval between two immunogold particles of around 25 nm for both TbCFAP44 and TbCFAP43, a distance similar to the interval between two outer 5–6 bridge subunits observed in the flagellum of sea urchin sperm[11]. All these connecting structures, located in the same area, could share physical interactions and the absence of one of the CFAP proteins could destabilize the whole complex, leading to both para-axonemal and axonemal defects. This echoes the anomalies evidenced in IF and TEM experiments in both mouse models and in man. This was particularly noticeable in *Cfap44*$^{-/-}$ sperm, where DMTs 5–6 were preferentially missing and para-axonemal structures such as dense fibers and FSs were defective (Fig. 6). Altogether, these results suggest that TbCFAP44 and TbCFAP43 may be the first two proteins specifically located next to DMTs 5–6, and therefore opening the way for a fine characterization of the protein complexes located in this area.

Apart from infertility, the studied patients did not present any obvious PCD-associated symptoms, such as cough, rhinitis, sinusitis, and rhinorrhea and chronic bronchitis. It is not very surprising because flagellar and cilia defects are not always associated[48]. We already have shown that it is also true for *DNAH1*-mutated patients[5,6,49,50]. These observations suggest that mutations in MMAF genes are only responsible for primary infertility without other PCD features and reinforce the presumption that the sperm flagellum is assembled and organized through a specific way, different from other cilia. These differences are notably illustrated by a different beating pattern and the presence of specific peri-axonemal structures, mainly the FS and the ODFs[51]. Importantly, we demonstrated here that in the trypanosome, CFAP43 and CFAP44 are associated to a subset of microtubule doublets facing the PFR. The PFR does not exist in mammalian sperm but its function is likely covered by the FS and the ODF. The fact that the absence of CFAP43/44 has no visible impact on motile cilia could therefore be explained by the functional specificity of the proteins: to interact/link the axoneme with flagella specific extra-axonemal structures, not present in motile cilia in respiratory tissues, brain, or in the oviduct. Further work should now be performed to elucidate the differential composition and organization of motile cilia and flagella. Last, analysis of KO heterozygous mice showed a small deterioration of sperm motility and morphology, suggesting that heterozygous mutations in key spermatogenesis genes might, alone or more likely through cumulative genetic and/or environmental factors, contribute to the less severe but much more frequent phenotype of mild to intermediate oligoasthenozoospermia.

Methods

Subjects and controls. We included 78 subjects presenting with asthenozoospermia due to a combination of morphological defects of the sperm flagella including: absent, short, bent, coiled flagella, and of irregular width without any of the additional symptoms associated with primary ciliary dyskinesia (PCD). The morphology of patients' sperm was assessed with Papanicolaou staining. Small variations in protocol might occur between the different laboratories. Subjects were recruited on the basis of the identification of >5% of at least three of the aforementioned flagellar morphological abnormalities (absent, short, coiled, bent, and irregular flagella).

The global average of all semen parameters are presented in Table 1 and were compared between the different genotype groups using a two-tailed*t* test. Forty-six patients were of North African origin and consulted for primary infertility at the Clinique des Jasmin in Tunis. Ten individuals originated from the Middle East (Iranians) and were treated in Tehran at the Royan Institute (Reproductive Biomedicine Research Center) for primary infertility and 22 patients were recruited in France: 21 at the Cochin Institute and 1 in Lille.

All patients were recruited between 2008 and 2016. All subjects had normal somatic karyotypes. Approximately half of the patients declared to be born from related parents. Sperm analysis was carried out in the source laboratories during the course of the routine biological examination of the patient, according to World Health Organization (WHO) guidelines[52]. Saliva and/or peripheral blood was obtained for all participants. During their medical consultation, all subjects answered a health questionnaire focused on PCD manifestations for infertility. Informed consent was obtained from all the subjects participating in the study according to the local protocols and the principles of the Declaration of Helsinki. In addition, the study was approved by local ethic committees. The samples were then stored in the CRB Germetheque (certification under ISO-9001 and NF-S 96-900) following a standardized procedure. Controls from fertile individuals with normal spermograms were obtained from CRB Germetheque. Consent for CRB storage was approved by the CPCP Sud-Ouest of Toulouse (coordination of the multi-sites CRB Germetheque).

WES and bioinformatics analysis. Genomic DNA was isolated from saliva using Oragen DNA Extraction Kit (DNAgenotech®, Ottawa, ON, Canada). Coding regions and intron/exon boundaries were enriched using the all Exon V5 Kit (Agilent Technologies, Wokingham, UK). DNA sequencing was undertaken at the Genoscope, Evry, France, on the HiSeq 2000 from Illumina®. Sequence reads were aligned to the reference genome (hg19) using MAGIC[53]. MAGIC produces quality-adjusted variant and reference read counts on each strand at every covered position in the genome. Duplicate reads and reads that mapped to multiple locations in the genome were excluded from further analysis. Positions whose sequence coverage was below 10 on either the forward or reverse strand were marked as low confidence, and positions whose coverage was below 10 on both strands were excluded. Single-nucleotide variations and small insertions/deletions (indels) were identified and quality-filtered using in-house scripts. Briefly, for each variant, independent calls are made on each strand, and only positions where both calls agree are retained. The most promising candidate variants were identified using an in-house bioinformatics pipeline, as follows. Variants with a minor allele frequency >5% in the NHLBI ESP6500 (Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA) or in 1,000 Genomes Project phase 1 datasets[54], or >1% in ExAC[55], were discarded. We also compared these variants to an in-house database of 94 control exomes obtained from subjects mainly originated from North Africa and Middle East corresponding to the geographical origin of most patients from this study and which is under-represented in SNP public databases. All variants present in homozygous state in this database were excluded. We used Variant Effect Predictor (VEP version 81[56]) to predict the impact of the selected variants. We only retained variants impacting splice donor/acceptor or causing frameshift, inframe insertions/deletions, stop-gain, stop loss, or missense variants, except those scored as "tolerated" by SIFT[57] (sift.jcvi.org) and as "benign" by Polyphen-2[58] (genetics.bwh.harvard.edu/pph2). Finally, identified mutations were validated by Sanger sequencing. PCR primers and protocols used for each patient are listed in the Supplementary Table 3. Sequencing reactions were carried out with BigDye Terminator v3.1 (Applied Biosystems). Sequences analysis were carried out on ABI 3130XL (Applied Biosystems). Sequences were analyzed using SeqScape software (Applied Biosystems).

RT-qPCR analysis. RT-qPCR was performed with cDNAs from various tissues of human and mouse including testes. A panel of eight organs was used for mouse experiments: testis, brain, lung, kidney, liver, stomach, colon, and heart. RNA extraction were performed from three DBA-C57 WT mice with the mirVana™ PARIS™ Kit (Life Technologies®). For human experiments, a panel of the three main ciliated tissues was used: testis, brain, and lung. Human RNAs were purchased from Life Technologies®. Reverse transcriptions were performed using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystem®). Each sample was assayed in triplicate for each gene on a StepOnePlus (Life Technologies®), with Power SYBR®Green PCR Master Mix (Life Technologies®). The PCR cycle was as follows: 10 min at 95 °C, 1 cycle for enzyme activation; 15 s at 95 °C, 60 s at 60 °C with fluorescence acquisition, 40 cycles for the PCR. RT-qPCR data were normalized using the reference housekeeping gene *ACTB* for human and mouse with the −ΔΔCt method[59]. The 2−ΔΔCt value was set at 0 in brain cells, resulting in an arbitrary expression of 1. Primers sequences and RT-qPCR conditions are indicated in the Supplementary Table 4. The efficacy of primers was checked using a standard curve. Melting curve analysis was used to confirm the presence of a single PCR product. Statistics were performed using a two-tailed*t* test on the Prism 4.0 software (GraphPad, San Diego, CA, USA) to compare the relative expression of *CFAP43* and *CFAP44* transcripts in several organs. Statistical tests with a two-tailed*P* values ≤0.05 were considered significant.

Immunostaining in human and mouse sperm cells. We performed IF staining on human and mouse spermatozoa as described by our laboratory[60]. In human, immunostaining could be carried out on samples from two of the patients with a stop-gain mutation, one mutated in each gene. IF experiments were performed using sperm cells from control individuals, from the patient P_{43}-5 homozygous for the c.2658C>T variant in *CFAP43* and from the patient P_{44}-2 homozygous for the c.3175C>T variant in *CFAP44*. Sperm cells were fixed in phosphate-buffered saline (PBS)/4% paraformaldehyde for 1 min at room temperature. After washing in 1 ml PBS, the sperm suspension was spotted onto 0.1% poly-L-lysine pre-coated slides (Thermo Scientific). After attachment, sperms were permeabilized with 0.1% (v/v) Triton X-100–DPBS (Triton X-100; Sigma-Aldrich) for 5 min at room temperature. Slides were then blocked in 5% corresponding normal serum–DPBS (normal goat or donkey serum; Gibco, Invitrogen) and incubated overnight at 4 °C with primary antibodies. For human experiments, the following primary antibodies were used: DNAI2, DNALI1, RSPH1, RPSH4A, SPAG6, GAS8, AKAP4, and anti-acetylated-α-tubulin. For mouse experiments: Rsph1, Rsph4a, Gas8, Spef2, Dnali1, Dnah5, Mpc1l, and anti-acetylated-α-tubulin. Primary antibodies references, provider, species, and dilutions used are listed in the Supplementary Table 5. Washes were performed with 0.1% (v/v) Tween-20–DPBS, followed by 1 h incubation at room temperature with secondary antibodies. Highly cross-adsorbed secondary antibodies (Dylight 488 and Dylight 549) were from Jackson Immunoresearch®. Appropriate controls were performed, omitting the primary antibodies. Samples were counterstained with 5 mg/ml Hoechst 33342 (Sigma-Aldrich) and mounted with DAKO mounting media (Life Technology). Fluorescence images were captured with a confocal microscope (Zeiss LSM 710).

Electron microscopy of human and mouse sperm cells. We performed TEM experiments on human and mouse (*Cfap43*$^{-/-}$ and *Cfap44*$^{-/-}$ and WT) spermatozoa. In human, TEM experiments were performed using sperm cells from fertile control individuals, from patient P_{43}-8 homozygous for the c.3352C>T variant in *CFAP43* and from patient P_{44}-3 homozygous for the c.2818dupG variant in *CFAP44*. Sperm cells were fixed with 2.5% glutaraldehyde in 0.1 M sodium phosphate (pH 7.4) during 2 h at room temperature. Cells were washed with buffer and post-fixed with 1% osmium tetroxide in the same buffer for 1 h at 4 °C. After washing with distilled water, cells were stained overnight at 4 °C with 0.5% uranyl acetate (pH 4.0). Cells were dehydrated through graded alcohol (30, 60, 90, 100, 100, and 100; 10 min for each bath) and infiltrated with a mix of 1:1 Epon/alcohol 100% for 2 h before two baths of fresh Epon for 2 h. Finally, cells were included in fresh Epon and polymerized during 2 days at 60 °C. Ultrathin sections (90 nm for human samples and 60 nm for mouse samples) of the cell pellet were done with an ultramicrotome (Leica). Sections were post-stained 10 min with 5% uranyl acetate (pH 4.0), washed four times with distilled water (1 h), and then stained with 0.4% lead citrate before being observed in a TEM at 80 kV (JEOL 1200EX), and images were acquired with a digital camera (Veleta, Olympus).

***Trypanosoma* cultures and transfection**. The trypanosome cell lines used in this study derived from the bloodstream form parental *T. brucei* 427 90-13 strain co-expressing the T7 RNA polymerase and tetracycline repressor[61]. Cells were cultured at 37 °C and 5% CO_2 and transfected and cloned as described in ref. [62] in medium supplemented with puromycin (0.1 µg/ml) or blasticidin (10 µg/ml) for constitutive expression of myc-tagged proteins and 10TY1-tagged proteins respectively, and with phleomycin (2.5 µg/ml) for RNAi-inducible cells. RNAi interference was induced with tetracycline (10 µg/ml).

Cell lines generated for this study. For WT level of expression of 3 x myc C-terminal tagged proteins ($TbCFAP44_{myc}$, $TbCFAP43_{myc}$), parental cells were transfected as in ref. [62] with a tagging cassette that was obtained by PCR using a set of long primers containing 80 nucleotides from the 5′-UTR and 3′-UTR flanking regions of the *Tb*CFAP44 and *Tb*CFAP43 open reading frame (ORF) and pMO-Tag23M vector as template[20] (Supplementary Table 6). For RNAi, fragments of *TbCFAP44* ORF (bp 1,880–2,309) and *TbCFAP43* ORF (bp 131–596) were cloned into p2T7tiB[63] and transfected into $TbCFAP44_{myc}$, $TbCFAP43_{myc}$ cell lines generating the cell lines $TbCFAP44_{myc}^{RNAi}$ and $TbCFAP43_{myc}^{RNAi}$ (Supplementary Table 6). For WT level of expression of 10 x TY1 N-terminal-tagged proteins, cell lines expressing $TbCFAP44_{myc}^{RNAi}$ and $TbCFAP43_{myc}^{RNAi}$ were transfected with a tagging cassette that was obtained by PCR using a set of primers amplifying the end of the 5′-UTR and the beginning of, respectively, the *TbCFAP43* and *TbCFAP44* ORF and the pPOTv4-10TY1-blast vector as template[34] (Supplementary Table 6).

***Trypanosoma* IF**. Cells were collected, washed, and processed for immunolabeling on cytoskeletons (from detergent-extracted cytoskeleton) as in ref. [33], except that the cytoskeletons were fixed in methanol at −20 °C for 30 min. Samples were incubated with the primary antibodies anti-myc and rabbit anti-PFR2 (PFR2 is a protein of the para-axonemal structure called PFR, 1:2,000 dilution in PBS) and with secondary antibodies anti-mouse FITC conjugated (Sigma F-2012, 1:400 dilution in PBS) and anti-rabbit Alexa-594 conjugated (Thermo Fisher A11012, 1:400 dilution in PBS). Nuclei and kinetoplasts were stained with DAPI (4',6-diamidino-2-phenylindole; 10 µg/ml). Images were acquired on a Zeiss Imager Z1 microscope, using a Photometrics Coolsnap HQ2 camera, with Zeiss ×100 or ×63 objectives (NA 1.4) using the Metamorph software (Molecular Devices), and processed with ImageJ.

For STED analysis on whole cells, cells were collected, washed, and loaded on poly-L-lysine-coated coverslips and air-dried. After a 5 min rehydration step in PBS, cells were fixed in methanol at −20 °C for at least 30 min. Samples were incubated with antibodies in four steps: (1) with the primary antibody anti-TY1 (BB2)[32], a mouse IgG1 monoclonal, (2) followed by the secondary antibody Alexa-594 coupled anti-mouse IgG1, (3) anti-tubulin (TAT1)[64], a mouse IgG2a monoclonal together with anti-PFR2 (rabbit polyclonal), (4) Alexa-488-coupledanti-mouse IgG2a and ATTO647N-coupled anti-rabbit secondary antibodies. Coverslips were mounted using ProLong Gold antifade reagent. Images were acquired on a Leica DMI6000 TCS SP8 X-STED microscope with a ×93 glycerol objective (NA 1.3), de-convolved with Huygens Pro 16.10, and 3D reconstructions were generated using Imaris X64 8.1.2. Primary antibodies' references, provider, species, and dilutions used are listed in Supplementary Table 6.

Electron microscopy of *Trypanosoma* cells. Cells were fixed in culture medium, by the addition of glutaraldehyde to a final concentration of 2.5%, for 60 min. They were pelleted (1,000 g, 10 min), resuspended in fixation buffer (2.5% glutaraldehyde, 2% PFA, 100 mM phosphate buffer, pH 7.4, 50 mM sucrose for 2 h). Fixed cells were washed in water for 10 min, post-fixed in 1% OsO_4 for 1 h, washed three times in water, and then samples were stained in 2% uranyl acetate in water at 4 °C overnight. Samples were next washed in water, dehydrated in ethanol, embedded in Spurr resin, and polymerized overnight at 60 °C. Sections (60–80 nm) were stained in 2% uranyl, then lead citrate, and visualized on a TECNAI 12 TEM.

Immunogold labeling on detergent-extracted cytoskeleton from *Trypanosoma* cells. Cell cultures were pelleted (1,000 g, 10 min), washed in vPBS (PBS, sucrose 15.7 g/l, glucose 1.8 g/l) and extracted with PIPES 100 mM, NP-40 (Igepal) 1%, $MgCl_2$ 1 mM, benzonase. Cytoskeletons were pelleted by centrifugation 10 min at 2,500 g and washed twice in PIPES 100 mM, 1%, $MgCl_2$ 1 mM, benzonase. Two hundred and fifty microliters of droplets of cytoskeletons were placed on parafilm. Butvar covered, charged, carbon-coated nickel grids were floated on each drop and incubated for 10 min to let the cytoskeletons adhere. The grids were then moved through two blocking droplets for 5 min each (blocking buffer—1% Fish skin gelatin, 0.01% Tween-20 in PBS) and then to droplets containing the primary antibody in blocking buffer (anti-TY1 1:10, anti-myc, anti-PFR2 mouse monoclonal L8C4) for 2 h at room temperature. The grids were washed four times in blocking buffer and incubated with gold-conjugated secondary antibodies—EM GAR10 1:20 (anti-mouse, 6 nm), in blocking buffer for 2 h at room temperature. After incubation, grids were washed 3× for 5 min in blocking buffer, 2× 5 min in 0.1% fish skin gelatin, 0.001% Tween-20 in PBS, 2× for 5 min in PBS, and then fixed in 2.5% glutaraldehyde in milliQ H_2O for 5 min. Samples were negatively stained with 10 µl auro-thioglucose for 20 s. Samples were visualized on a FEI Tecnai 12 electron microscope, camera ORIUS 1000 11M Pixel (resolution 3–5 nm). Images were acquired with Digitalmicrograph and processed with ImageJ.

Western blotting on detergent-extracted cytoskeleton from *Trypanosoma* cells. Proteins from detergent-extracted cytoskeleton (2.10^7) were separated on sodium dodecyl sulfate-polyacrylamide gel electrophoresis gels (6%) and semi-dry transferred (Bio-Rad) for 45 min at 25 V on PVDF membrane. After a 1 h blocking step in 5% milk in PBS-0.2% Tween-20, the membranes were incubated overnight at 4 °C with the anti-TY1 primary antibody (BB2, 1 :5,000) diluted in blocking buffer. After three washes in blocking buffer and one wash in 1 M NaCl, the membranes were incubated with the HRP-conjugated secondary antibody (Jackson, 1:10,000 dilution), washed twice 10 min in blocking buffer and twice 5 min in PBS. Blots were revealed using the Clarity Western ECL Substrate Kit (Bio-Rad) with the ImageQuant LAS4000. After membrane stripping, protein loading was controlled by probing tubulin (TAT1, 1:500) as described above.

CRISPR/Cas9 KO mice. All animal procedures were run according to the Swiss and French guidelines on the use of animals in scientific investigations with the approval of the French local Ethical Committee (ComEth Grenoble No. 318, ministry agreement number #7128 UHTA-U1209-CA) and the Direction Générale de la Santé (DGS) for the State of Geneva. All the procedures were done in Geneva until the birth of the modified litters. For each gene, three plasmids were injected directly into the nucleus of the zygotes. One plasmid expressed the Cas9 protein and the other two expressed two distinct RNA guides (single guide RNA, or sgRNA) targeting exons 2 and 21 of the *Cfap43* gene and exons 3 and 15 of the *Cfap44* gene. All plasmids (pGS-U6-sgRNA expression vector) were ordered from GeneScript (Piscataway, NJ, USA) with the different cDNA sequences already inserted. sgRNA sequences are indicated in the Supplementary Table 7. The Cas9 nuclease and sgRNAs were introduced into fertilized oocytes. Microinjected oocytes were introduced into pseudopregnant host females and carried to term. Edited founders were identified by Sanger sequencing from tail biopsies. Tail biopsies (2 mm in length) were digested in 200 µl lysis Direct PCR Lysis Reagent (Tail) (Viagen Biotech Inc., Los Angeles, CA, USA) and 0.2 mg of proteinase K for 12–15 h at 55 °C and 1 h at 85 °C. The DNA was directly used for PCRs. The two targeted exons were amplified using the following PCR protocol: 59 °C x 1 cycle,

58 °C x 1 cycle, and 57 °C x 35 cycles with 1 min elongation. Sanger sequencing was then performed to identify CRISP/Cas9-induced mutations and genotypes were determined according to the sequence electrophoregrams. Mice carrying desired modification events are bred with C57BL6/J to ensure germline transmission and eliminate any possible mosaicism. From the second generation of mice, genotyping was performed by high-resolution melting (HRM) using MeltDoctor™ HRM Master Mix with the following parameters: 10 min at 95 °C for enzyme activation, 15 s at 95 °C for denaturation, 1 min at 60 °C for annealing and extension, then for the melting curve, 1 min at 95 °C for denaturation, 1 min at 60 °C for annealing, 15 s at 95 °C for HRM and 15 s at 60 °C for annealing. List of primers used for mice genotyping with both methods is available in the Supplementary Table 8.

Heterozygous animals were mated to generate homozygous offspring. Approximately 25% of the offspring were homozygous for the mutated allele, indicating the absence of an increased embryonic or postnatal lethality in *Cfap43*$^{-/-}$*and Cfap44*$^{-/-}$ mice. Mice were housed with unlimited access to food and water and were sacrificed by cervical dislocation after 8 weeks old, which means that they were pubescent and that their reproductive organs were fully established.

The edited gene expression in mutant mice was validated by RT-PCR followed by Sanger sequencing of testicular transcripts. RT-PCR experiments were performed using testis RNA from WT, heterozygous, and homozygous animals. Reverse transcription was carried out with 5 µl of extracted RNA (~500 ng) using Macherey Nagel columns (Macherey Nagel, Hoerdt, France) according to the manufacturer's protocol. Hybridization of the oligo(dT) was performed by incubating for 5 min at 65 °C and quenching on ice with the following mix: 5 µl RNA, 3 µl of poly T oligo primers (dT) 12–18 (10 mM, Pharmacia), 3 µl of the four dNTPs (0.5 mM, Roche Diagnostics), and 2.2 µl of H_2O. Reverse transcription then was carried out for 30 min at 55° after the addition of 4 µl of 5× buffer, 0.5 µl RNase inhibitor, and 0.5 µl of Transcriptor reverse transcriptase (Roche Diagnostics). Two microliters of the obtained cDNA mix was used for the subsequent PCR and Sanger sequencing. Primers sequences and RT-PCR conditions are indicated in the Supplementary Table 9.

Phenotypic analysis of mouse mutant. To test the fertility, pubescent *Cfap43*$^{-/-}$*and Cfap44*$^{-/-}$ males (8 weeks old) were mated with C57BL6/J females. To determine sperm concentration, sperm samples were collected from the cauda epididymis of 8-week-old *Cfap43*$^{-/-}$*and Cfap44*$^{-/-}$ and sperm number was determined using a hemocytometer under a light microscope. For sperm morphology, sperm was displayed over a slide, dried at room temperature, and then fixed in 75% ethanol. Harris–Schorr staining was performed according to the WHO protocol. Schorr staining solution was obtained from Merck and at least 100 sperm per animal were analyzed. Mobility of sperm was assessed with computer-assisted motility analysis (CEROS I, Hamilton Thorn Research, Beverly, MA, USA) in an analysis chamber (100 µm depth, 30 µl volume, Leja Products B.V., Netherlands) at 37 °C. The settings employed for the analysis were as follows: acquisition rate, 60 Hz; number of frames, 100; minimum contrast, 25; minimum cell size, 10; low static size gate, 2.4; high static size gate, 2.4; low static intensity gate, 1.02; high static intensity gate, 1.37; minimum elongation gate, 12; maximum elongation gate, 100; magnification factor, 0.70. The motility parameters measured were curvilinear velocity, average path velocity (VAP), and straight line velocity (VSL). At least 100 motile sperm were analyzed for each assay. Motile sperm and progressive sperm were characterized by VAP >1 µm/s, by average path velocity >30 µm/s and straightness (VSL/VAP) >70%, respectively. For each genotype, five mice were used.

Statistical analyses were performed with the SigmaPlot software. Unpaired *t* test was used to compare the different genotypes. Statistical tests with a two-tailed*P* values ≤0.05 were considered significant.

RNA expression analysis. Total RNA was isolated from 10^8 cells of parental, non-induced, and tetracycline-induced RNAi cells using the TRIzol reagent according to the manufacturer's instructions (Life Technologies). The constitutively expressed housekeeper telomerase reverse transcriptase was used as an internal control[65]. RT-PCR was carried out with a SuperScript III One-step RT-PCR System with Platinum *Taq*high-fidelity polymerase (Life Technologies Ltd, Paisley, UK) following the manufacturer's protocol. Briefly, 100 ng of total RNA were mixed with primers and reverse transcriptase solution in 50 µl total volume and using the following cycle protocol: 30 min at 55 °C (reverse transcription); 2 min at 94 °C (inactivation of reverse transcriptase and activation *Taq* polymerase); followed by 25 cycles of regular PCR (denaturation: 94 °C for 15 s; annealing: 55 °C for 30 s; extension: 68 °C for 40 s); and finalized with a hold for 5 min at 68 °C. The RT-PCR products were resolved on 1% agarose running gel with BET and visualized by UV light. Primer sequence is provided in Supplementary Table 10.

Data availability. Data on genetic variants described here are available on ClinVar (https://www-ncbi-nlm-nih-gov.gate2.inist.fr/clinvar/) – submission ID SUB2319254444. All other data that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 1 June 2017 Accepted: 28 December 2017
Published online: 15 February 2018

References

1. Coutton, C., Fissore, R. A., Palermo, G. D., Stouffs, K. & Touré, A. Male infertility: genetics, mechanism, and therapies. *Biomed. Res. Int.* **2016**, 7372362 (2016).
2. Coutton, C., Escoffier, J., Martinez, G., Arnoult, C. & Ray, P. F. Teratozoospermia: spotlight on the main genetic actors in the human. *Hum. Reprod. Update* **21**, 455–485 (2015).
3. Ray, P. F. et al. Genetic abnormalities leading to qualitative defects of sperm morphology or function. *Clin. Genet.* **91**, 217–232 (2017).
4. Kherraf, Z.-E. et al. SPINK2 deficiency causes infertility by inducing sperm defects in heterozygotes and azoospermia in homozygotes. *EMBO Mol. Med.* **9**, 1132–1149 (2017).
5. Ben Khelifa, M. et al. 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 (2014).
6. Amiri-Yekta, A. et al. Whole-exome sequencing of familial cases of multiple morphological abnormalities of the sperm flagella (MMAF) reveals new DNAH1 mutations. *Hum. Reprod. (Oxf., Engl.)* **31**, 2872–2880 (2016).
7. Ivliev, A. E., 't Hoen, P. A., van Roon-Mom, W. M., Peters, D. J. & Sergeeva, M. G. Exploring the transcriptome of ciliated cells using in silico dissection of human tissues. *PLoS ONE* **7**, e35618 (2012).
8. Smith, T. F. Diversity of WD-repeat proteins. *Subcell. Biochem.* **48**, 20–30 (2008).
9. Smith, T. F. *Diversity of WD-Repeat Proteins* (Landes Bioscience, Austin, 2013).
10. Afzelius, B. Electron microscopy of the sperm tail; results obtained with a new fixative. *J. Biophys. Biochem. Cytol.* **5**, 269–278 (1959).
11. Lin, J., Heuser, T., Song, K., Fu, X. & Nicastro, D. One of the nine doublet microtubules of eukaryotic flagella exhibits unique and partially conserved structures. *PLoS ONE* **7**, e46494 (2012).
12. Vanderperre, B. et al. MPC1-like is a placental mammal-specific mitochondrial pyruvate carrier subunit expressed in postmeiotic male germ cells. *J. Biol. Chem.* **291**, 16448–16461 (2016).
13. Fawcett, D. W. The mammalian spermatozoon. *Dev. Biol.* **44**, 394–436 (1975).
14. Lindemann, C. B., Orlando, A. & Kanous, K. S. The flagellar beat of rat sperm is organized by the interaction of two functionally distinct populations of dynein bridges with a stable central axonemal partition. *J. Cell. Sci.* **102**, 249–260 (1992).
15. Vincensini, L., Blisnick, T. & Bastin, P. 1001 model organisms to study cilia and flagella. *Biol. Cell Auspices Eur. Cell Biol. Organ.* **103**, 109–130 (2011).
16. de Souza, W. & Souto-Padrón, T. The paraxial structure of the flagellum of trypanosomatidae. *J. Parasitol.* **66**, 229–236 (1980).
17. Sherwin, T. & Gull, K. The cell division cycle of *Trypanosoma brucei brucei*: timing of event markers and cytoskeletal modulations. *Philos. Trans. R. Soc. Lond. Ser. B* **323**, 573–588 (1989).
18. Santrich, C. et al. A motility function for the rod of Leishmania parasites revealed by PFR-2 gene knockouts. *Mol. Biochem. Parasitol.* **90**, 95–109 (1997).
19. Bastin, P., Sherwin, T. & Gull, K. Paraflagellar rod is vital for trypanosome motility. *Nature* **391**, 548 (1998).
20. Oberholzer, M., Morand, S., Kunz, S. & Seebeck, T. A vector series for rapid PCR-mediated C-terminal in situ tagging of *Trypanosoma brucei* genes. *Mol. Biochem. Parasitol.* **145**, 117–120 (2006).
21. Pullen, T. J., Ginger, M. L., Gaskell, S. J. & Gull, K. Protein targeting of an unusual, evolutionarily conserved adenylate kinase to a eukaryotic flagellum. *Mol. Biol. Cell* **15**, 3257–3265 (2004).
22. Ridgley, E., Webster, P., Patton, C. & Ruben, L. Calmodulin-binding properties of the paraflagellar rod complex from *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **109**, 195–201 (2000).
23. Aslett, M. et al. TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res.* **38**, D457–D462 (2010).
24. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
25. Subota, I. et al. Proteomic analysis of intact flagella of procyclic *Trypanosoma brucei* cells identifies novel flagellar proteins with unique sub-localization and dynamics. *Mol. Cell. Proteom.* **13**, 1769–1786 (2014).
26. Broadhead, R. et al. Flagellar motility is required for the viability of the bloodstream trypanosome. *Nature* **440**, 224–227 (2006).
27. Baron, D. M., Ralston, K. S., Kabututu, Z. P. & Hill, K. L. Functional genomics in *Trypanosoma brucei* identifies evolutionarily conserved components of motile flagella. *J. Cell Sci.* **120**, 478–491 (2007).
28. Mirabello, C. & Pollastri, G. Porter, PaleAle 4.0: high-accuracy prediction of protein secondary structure and relative solvent accessibility. *Bioinformatics (Oxf., Engl.)* **29**, 2056–2058 (2013).
29. Wilson, D. et al. SUPERFAMILY—sophisticated comparative genomics, data mining, visualization and phylogeny. *Nucleic Acids Res.* **37**, D380–D386 (2009).
30. Gough, J., Karplus, K., Hughey, R. & Chothia, C. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J. Mol. Biol.* **313**, 903–919 (2001).

31. Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536–540 (1995).
32. Bastin, P., Bagherzadeh, Z., Matthews, K. R. & Gull, K. A novel epitope tag system to study protein targeting and organelle biogenesis in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **77**, 235–239 (1996).
33. Albisetti, A. et al. Interaction between the flagellar pocket collar and the hook complex via a novel microtubule-binding protein in *Trypanosoma brucei*. *PLoS Pathog.* **13**, e1006710 (2017).
34. Dean, S. et al. A toolkit enabling efficient, scalable and reproducible gene tagging in trypanosomatids. *Open Biol.* **5**, 140197 (2015).
35. Langousis, G. & Hill, K. L. Motility and more: the flagellum of *Trypanosoma brucei*. *Nat. Rev. Microbiol.* **12**, 505–518 (2014).
36. Ralston, K. S., Kisalu, N. K. & Hill, K. L. Structure–function analysis of dynein light chain 1 identifies viable motility mutants in bloodstream-form *Trypanosoma brucei*. *Eukaryot. Cell* **10**, 884–894 (2011).
37. Wambergue, C. et al. Patients with multiple morphological abnormalities of the sperm flagella due to DNAH1 mutations have a good prognosis following intracytoplasmic sperm injection. *Hum. Reprod. (Oxf., Engl.)* **31**, 1164–1172 (2016).
38. Ran, F. A. et al. Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
39. Neesen, J. et al. Disruption of an inner arm dynein heavy chain gene results in asthenozoospermia and reduced ciliary beat frequency. *Hum. Mol. Genet.* **10**, 1117–1128 (2001).
40. Ostrowski, L. E., Dutcher, S. K. & Lo, C. W. Cilia and models for studying structure and function. *Proc. Am. Thorac. Soc.* **8**, 423–429 (2011).
41. Silflow, C. D. & Lefebvre, P. A. Assembly and motility of eukaryotic cilia and flagella. Lessons from *Chlamydomonas reinhardtii*. *Plant Physiol.* **127**, 1500–1507 (2001).
42. Alberts, B. et al. *Molecular Biology of the Cell* (Garland Science, New York, 2002).
43. Coutton, C., Escoffier, J., Martinez, G., Arnoult, C. & Ray, P. F. Teratozoospermia: spotlight on the main genetic actors in the human. *Hum. Reprod. Update* **21**, 455–485 (2015).
44. Yan, W. Male infertility caused by spermiogenic defects: lessons from gene knockouts. *Mol. Cell. Endocrinol.* **306**, 24–32 (2009).
45. Morga, B. & Bastin, P. Getting to the heart of intraflagellar transport using *Trypanosoma* and *Chlamydomonas* models: the strength is in their differences. *Cilia* **2**, 16 (2013).
46. Huet, D., Blisnick, T., Perrot, S. & Bastin, P. The GTPase IFT27 is involved in both anterograde and retrograde intraflagellar transport. *eLife* **3**, e02419 (2014).
47. Koyfman, A. Y. et al. Structure of *Trypanosoma brucei* flagellum accounts for its bihelical motion. *Proc. Natl. Acad. Sci. USA* **108**, 11105–11108 (2011).
48. Praveen, K., Davis, E. E. & Katsanis, N. Unique among ciliopathies: primary ciliary dyskinesia, a motile cilia disorder. *F1000prime Rep.* **7**, 36 (2015).
49. Wang, X. et al. Homozygous DNAH1 frameshift mutation causes multiple morphological anomalies of the sperm flagella in Chinese. *Clin. Genet*. https://doi.org/10.1111/cge.12857 (2016).
50. Sha, Y. et al. DNAH1 gene mutations and their potential association with dysplasia of the sperm fibrous sheath and infertility in the Han Chinese population. *Fertil. Steril.* **107**, 1312–1318.e2 (2017).
51. Oberholzer, M. et al. Trypanosomes and mammalian sperm: one of a kind? *Trends Parasitol.* **23**, 71–77 (2007).
52. World Health Organisation. *WHO Laboratory Manual for the Examination and Processing of Human Semen*, 5th ed. (World Health Organization, 2010).
53. SEQC/MAQC-III Consortium. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat. Biotechnol.* **32**, 903–914 (2014).
54. 1000 Genomes Project Consortium. et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
55. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
56. McLaren, W. et al. The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
57. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081 (2009).
58. Adzhubei, I. A. et al. A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
59. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(−Delta Delta C(T)) method. *Methods* **25**, 402–408 (2001).
60. Escoffier, J. et al. Subcellular localization of phospholipase Cζ in human sperm and its absence in DPY19L2-deficient sperm are consistent with its role in oocyte activation. *Mol. Hum. Reprod.* **21**, 157–168 (2015).
61. Wirtz, E., Leal, S., Ochatt, C. & Cross, G. A. A tightly regulated inducible expression system for conditional gene knockouts and dominant-negative genetics in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **99**, 89–101 (1999).
62. Dauchy, F.-A. et al. *Trypanosoma brucei* CYP51: essentiality and targeting therapy in an experimental model. *PLoS Negl. Trop. Dis.* **10**, e0005125 (2016).
63. LaCount, D. J., Barrett, B. & Donelson, J. E. *Trypanosoma brucei* FLA1 is required for flagellum attachment and cytokinesis. *J. Biol. Chem.* **277**, 17580–17588 (2002).
64. Woods, A. et al. Definition of individual components within the cytoskeleton of *Trypanosoma brucei* by a library of monoclonal antibodies. *J. Cell Sci.* **93** (Part 3), 491–500 (1989).
65. Brenndörfer, M. & Boshart, M. Selection of reference genes for mRNA quantification in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **172**, 52–55 (2010).

Acknowledgements

We thank Anne Bertrand from the GIN electron microscopy platform and the IAB microscopy platform and Alexei Grichine and Jacques Mazzega for their technical help. We thank Klaus Ersfeld (Bayreuth University) for the anti-myc antibody, Nicolas Biteau (Bordeaux University) for the anti-PFR2 antibody, Keith Gull (Oxford University) for the TAT1 antibody, Philippe Bastin (Institut Pasteur, Paris) for the anti-TY1 antibody, and Vanderperre Benoît and Martinou Jean Claude (UNIGE university, Geneva, CH) for the Mpcl1l antibody. We thank Denise Escalier and Gérard Gascon for discussions and expertise on the topic. The STED microscopy was done in the Bordeaux Imaging Centre a service unit of the CNRS-INSERM and Bordeaux University, member of the national infrastructure France BioImaging supported by the French National Research Agency (ANR-10-INBS-04). The help of Christel Poujol and Patrice Mascalchi is acknowledged. Some of the electron microscopy was done at the Bordeaux Imaging Centre. We thank our patients and control individuals for their participation. This work was mainly supported by the following grants: The "MAS-Flagella" project financed by French ANR and the DGOS for the program PRTS 2014 and the "Whole genome sequencing of patients with Flagellar Growth Defects (FGD)" financed by the fondation maladies rares (FMR) for the program Séquençage à haut débit 2012.

Author contributions

C.C., M.B., A.T., C.A., and P.F.R. analyzed the data and wrote the manuscript; Z.-E.K., A. A.-Y., L.S., P.L., E.E.K., J.F., and C.C. performed molecular work; T.K., J.-F.D., A.B., and N.T.-M. analyzed genetic data; C.C., A.S.V., P.L.T., C.W.-L., and S.P.B. performed IF experiments; S.C. performed protein in silico analysis, K.P.-G., A.S.V, A.S., and A.T. performed EM experiments, Z.-E.K., A.A.-Y., A.S.V., B.C., S.F.B.M., S.N., and J.E. performed mouse work, D.D., N.L., D.R.R., and M.B. performed *Trypanosoma* work. N.L. and M.B. performed STED analysis. S.F.B.M., G.M., A.D., S.H.H., V.M., L.H., O.M., M. M., H.L., M.K., S.H., V.S., P.-S.J., J.-P.W., H.G., E.D., and R.Z. provided clinical samples and data; M.B., A.T., C.A., and P.F.R. designed the study, supervised all molecular laboratory work, had full access to all of the data in the study. and takes responsibility for the integrity of the data and its accuracy. All authors contributed to the report.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-017-02792-7.

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

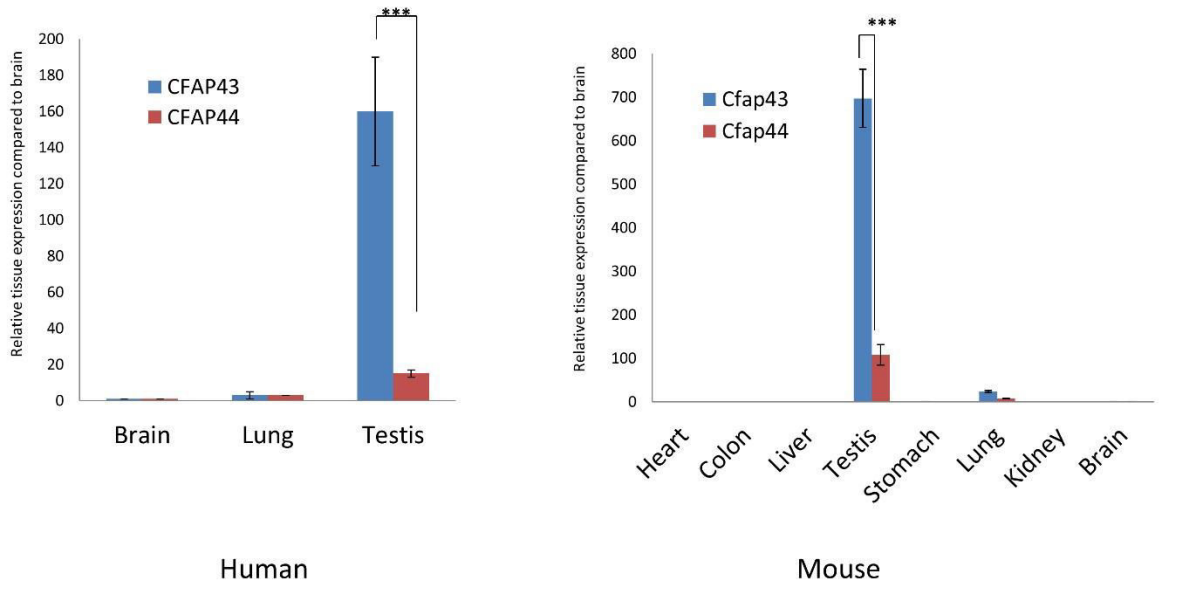
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Charles Coutton[1,2], Alexandra S. Vargas[1], Amir Amiri-Yekta[1,3,4], Zine-Eddine Kherraf[1,3], Selima Fourati Ben Mustapha[5], Pauline Le Tanno[1,2], Clémentine Wambergue-Legrand[1,3], Thomas Karaouzène[1,3], Guillaume Martinez[1,2], Serge Crouzy[6], Abbas Daneshipour[4], Seyedeh Hanieh Hosseini[7], Valérie Mitchell[8], Lazhar Halouani[5], Ouafi Marrakchi[5], Mounir Makni[5], Habib Latrous[5], Mahmoud Kharouf[5], Jean-François Deleuze[9], Anne Boland[9], Sylviane Hennebicq[1,10], Véronique Satre[1,2], Pierre-Simon Jouk[11], Nicolas Thierry-Mieg[12], Beatrice Conne[13], Denis Dacheux[14,15], Nicolas Landrein[14], Alain Schmitt[16,17,18], Laurence Stouvenel[16,17,18], Patrick Lorès[16,17,18], Elma El Khouri[16,17,18], Serge P. Bottari[1], Julien Fauré[19,20], Jean-Philippe Wolf[18,21], Karin Pernet-Gallay[20], Jessica Escoffier[1], Hamid Gourabi[4], Derrick R. Robinson[14], Serge Nef[13], Emmanuel Dulioust[18,21], Raoudha Zouari[5], Mélanie Bonhivers[14], Aminata Touré[16,17,18], Christophe Arnoult[1] & Pierre F. Ray[1,3]

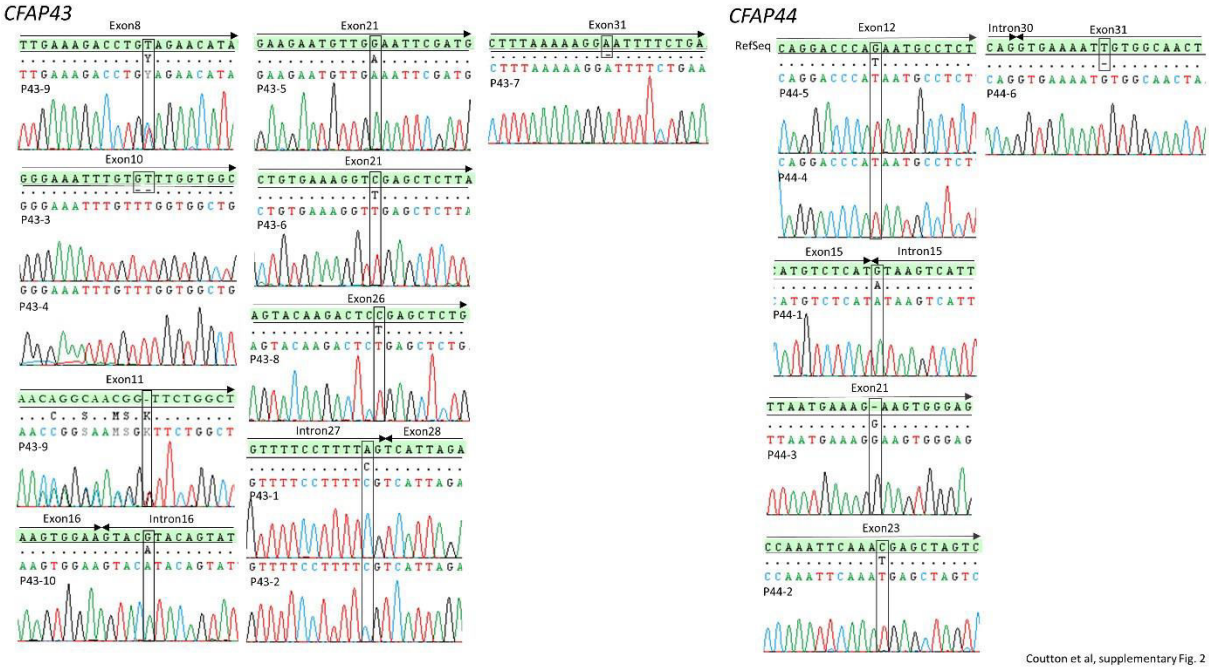
[1]Genetic Epigenetic and Therapies of Infertility, Institute for Advanced Biosciences, Inserm U1209, CNRS UMR 5309, Université Grenoble Alpes, 38000 Grenoble, France. [2]CHU de Grenoble, UM de Génétique Chromosomique, 38000 Grenoble, France. [3]CHU de Grenoble, UM GI-DPI, 38000 Grenoble, France. [4]Department of Genetics, Reproductive Biomedicine Research Center, Royan Institute for Reproductive Biomedicine, ACER, Tehran, 16635-148, Iran. [5]Polyclinique les Jasmins, Centre d'Aide Médicale à la Procréation, Centre Urbain Nord, 1003 Tunis, Tunisia. [6]Laboratoire de Chimie et Biologie des Métaux, Institut de Recherche en Technologie et Sciences pour le Vivant, CEA iRTSV/LCBM/GMCT, CNRS UMR 5249, Université Grenoble Alpes, 38054 Grenoble, France. [7]Department of Andrology, Reproductive Biomedicine Research Center, Royan Institute for Reproductive Biomedicine, ACECR, Tehran, 16635-148, Iran. [8]EA 4308, Department of Reproductive Biology and Spermiology-CECOS Lille, University Medical Center, Lille, 59037, France. [9]Centre National de Génotypage, Institut de Génomique, CEA, 91000Evry, France. [10]CHU de Grenoble, UF de Biologie de la procréation, 38000 Grenoble, France. [11]CHU de Grenoble, UF de Génétique Médicale, 38000 Grenoble, France. [12]Univ. Grenoble Alpes/CNRS TIMC-IMAG, 38000 Grenoble, France. [13]Department of Genetic Medicine and Development, University of Geneva Medical School, 1211 Geneva, Switzerland. [14]Microbiologie Fondamentale et Pathogénicité CNRS UMR 5234, University Bordeaux, 33000 Bordeaux, France. [15]Bordeaux-INP, Microbiologie Fondamentale et Pathogénicité, UMR-CNRS 5234, 33000 Bordeaux, France. [16]Institut National de la Santé et de la Recherche Médicale, INSERM U1016, Institut Cochin, 75014 Paris, France. [17]Centre National de la Recherche Scientifique, CNRS UMR8104, 75014 Paris, France. [18]Sorbonne Paris Cité, Faculté de Médecine, Université Paris Descartes, 75014 Paris, France. [19]CHU de Grenoble UF de Biochimie Génétique et Moléculaire, 38000 Grenoble, France. [20]Grenoble Neuroscience Institute, INSERM 1216, 38000 Grenoble, France. [21]Laboratoire d'Histologie Embryologie - Biologie de la Reproduction, GH Cochin Broca Hôtel Dieu, Assistance Publique-Hôpitaux de Paris, 75014 Paris, France. Mélanie Bonhivers, Aminata Touré, Christophe Arnoult, and Pierre F. Ray contributed equally to this work.

SUPPLEMENTARY INFORMATION

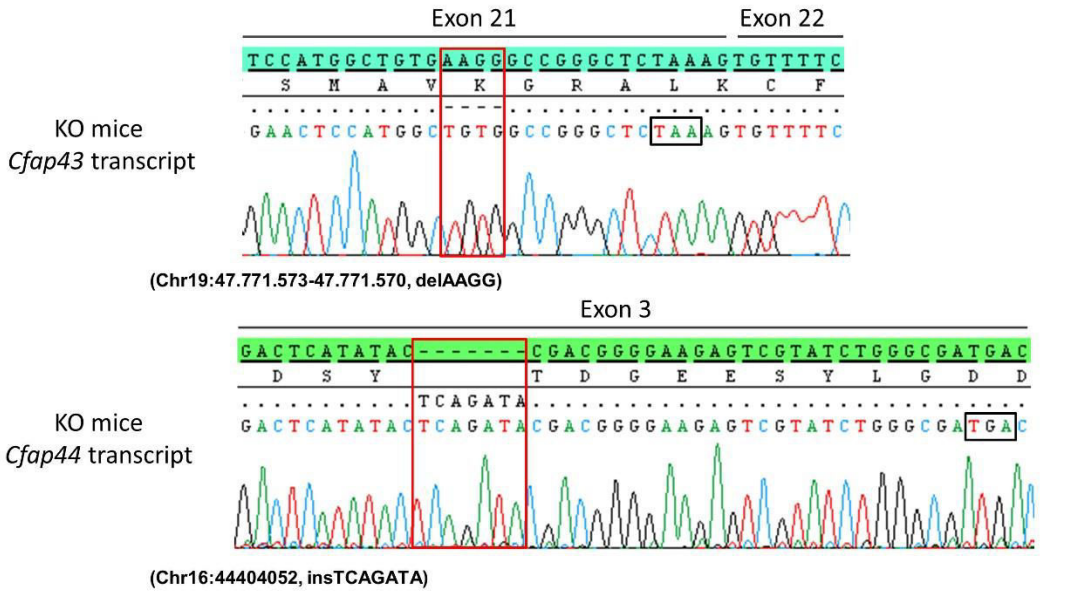
SUPPLEMENTARY FIGURES



Supplementary Figure 1. Relative mRNA Expression of human and mouse CFAP43 and CFAP44 transcripts. *CFAP43* and *CFAP4* mRNA levels in a panel of human and mouse normal tissues. Results are presented as the mean of triplicates (ratio target gene/ACTB) ± Standard Deviation (SD). RT-qPCR data were normalized using the reference gene ACTB with the -ΔΔCt method. Brain expression is arbitrarily set to 1. In human and mouse, *CFAP43* (blue columns) and *CFAP44* (red columns) have their strongest expression in testis compared to other organs. Unpaired t-test, ***P< 0.001.

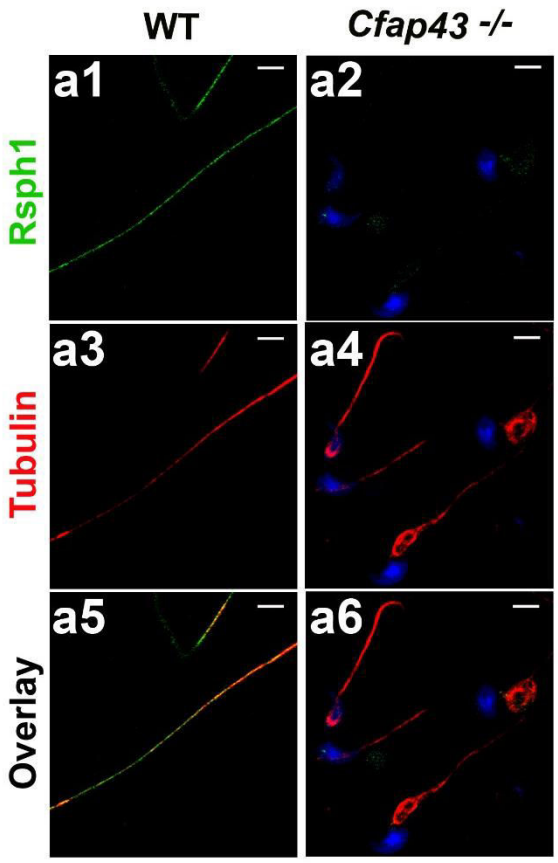


Supplementary Figure 2. Electrophoregrams of Sanger sequencing for all CFAP43 and CFAP44 mutated patients compared to reference sequence.

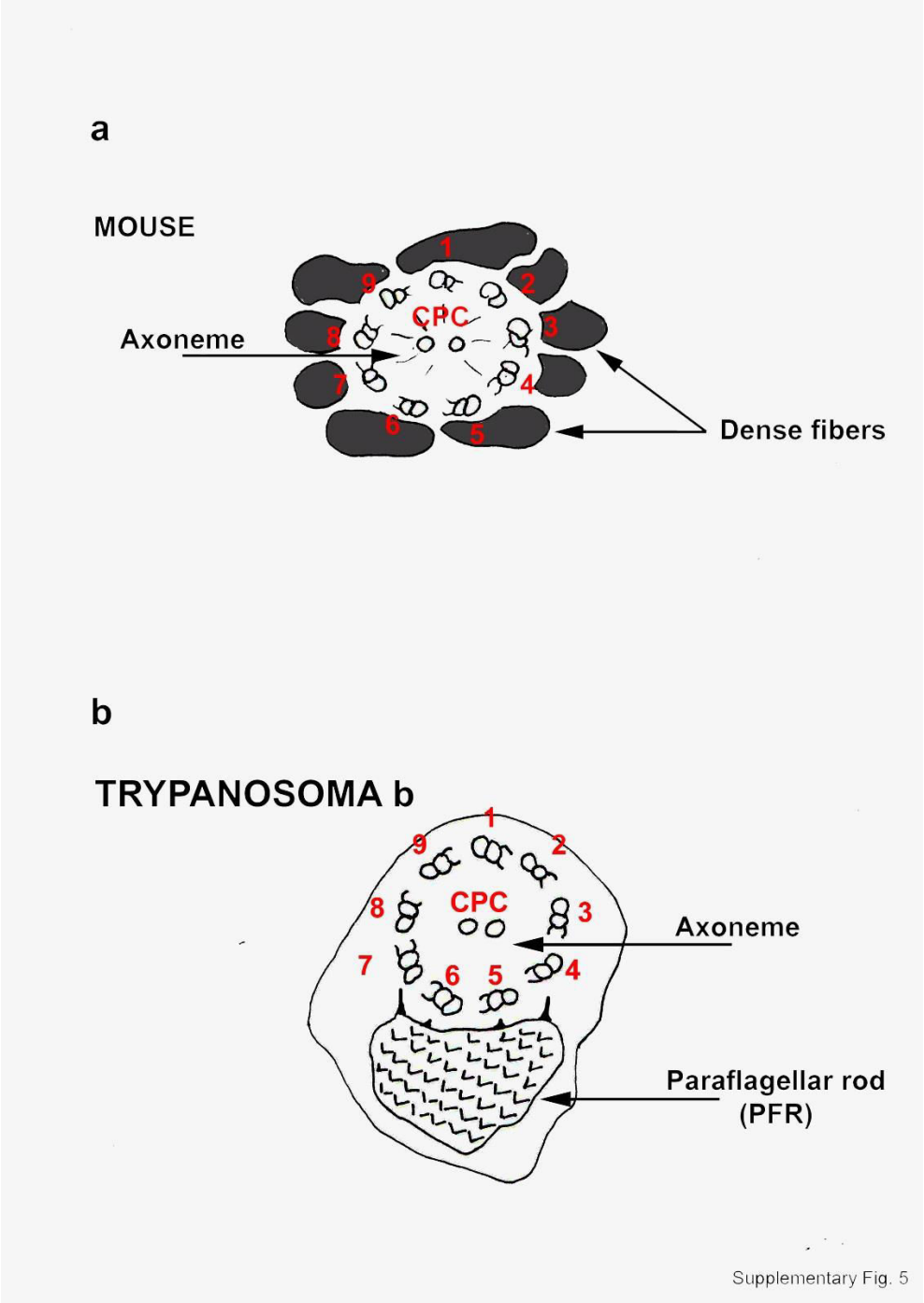


Supplementary Fig. S3

Supplementary Figure 3. Identification of indels in mice generated by CRISPR/Cas9 technology. *Cfap44* and *Cfap43* KO mice were generated by targeting exon 3 and 21, respectively. Sequencing of transcripts obtained from the testis of mutant mice allowed us to identify a frameshift mutation InsTCAGATA in *Cfap44* and delAAGG in *Cfap43* (Red boxes) leading to a premature stop codon (Black box).

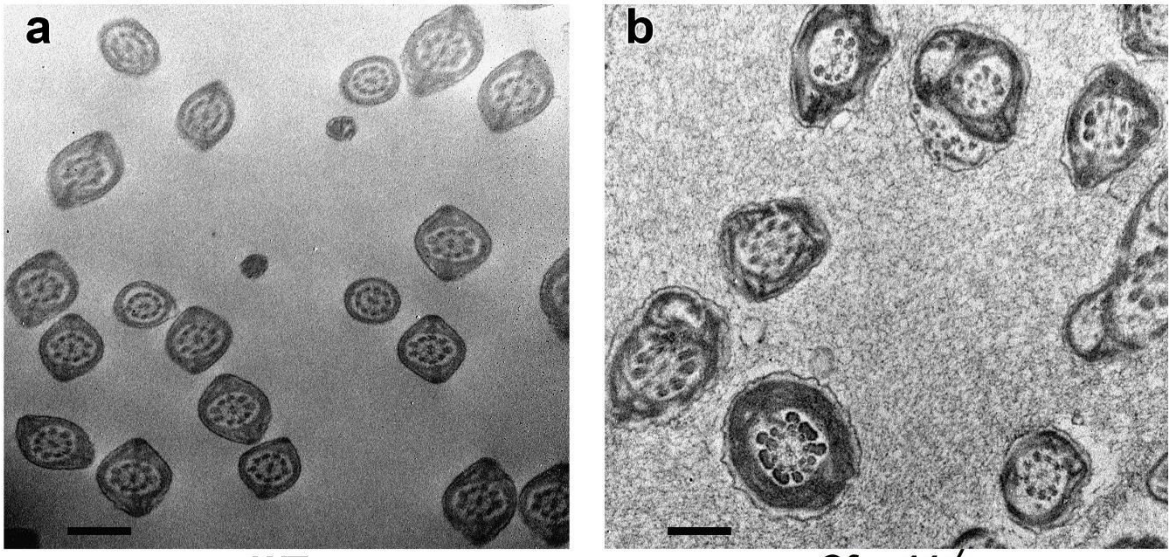


Supplementary Figure 4. Head of radial spoke Rsph1 are absent in sperm from Cfap43$^{-/-}$ males. (a1,a3) Staining of WT sperm with an anti-Rsph1 (**a1**, green) and anti-tubulin (**a3**, red) antibodies. (**a5)** overlay of Rsph1 and tubulin staining. Sperm were counterstained with Hoechst (blue). (**a2, a4, a6)** Similar experiments on sperm from *Cfap43*$^{-/-}$ males. Scale bars = 5 µm.



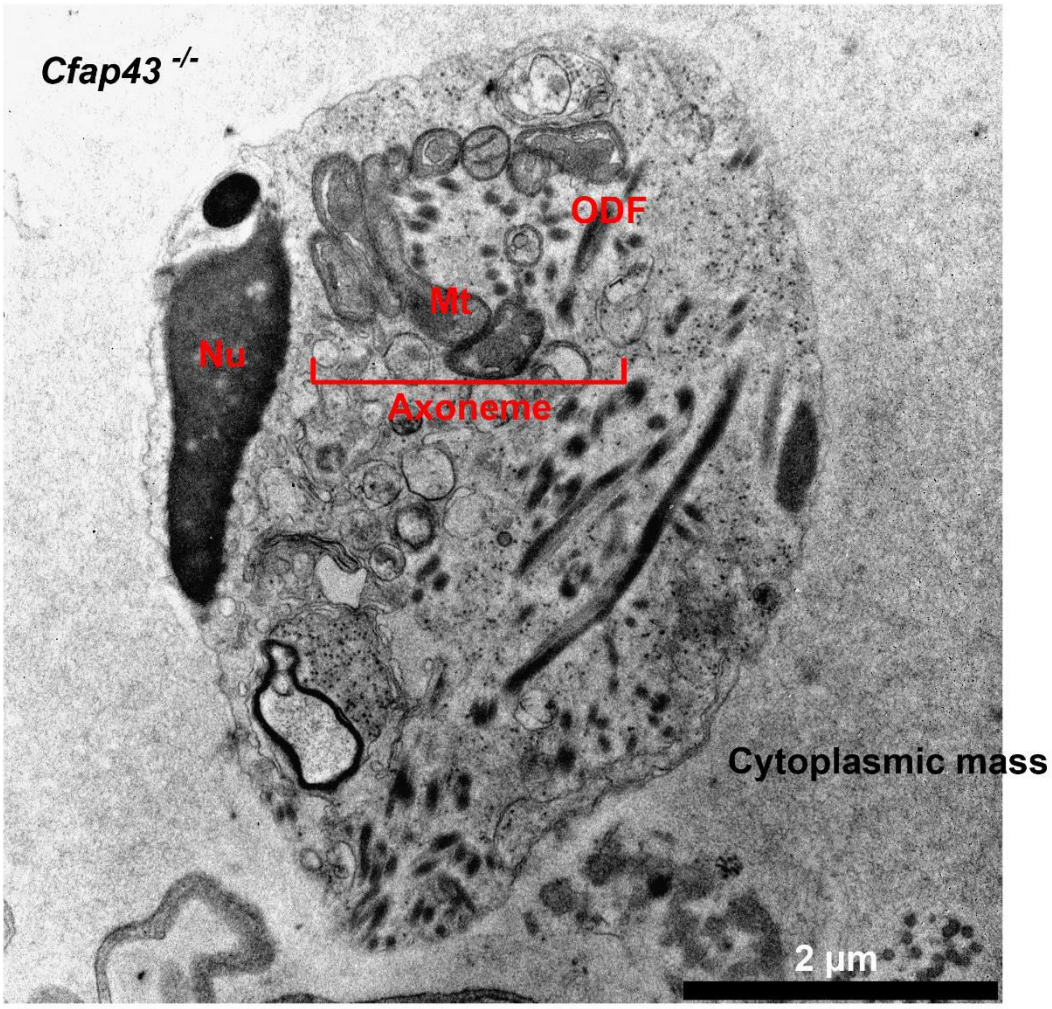
Supplementary Figure 5. Drawing of the ultrastructure of mouse and *Trypanosoma b.* axonemes, showing the annotations of the DMTs.

a. In mouse, drawing representing a section of the midpiece of the flagellum. The mitochondria and the plasma membrane are not represented. The axoneme is surrounded by 9 dense fibers.
b. In *Trypanosoma b.*, drawing representing a section of the flagellum and showing the organization of the axoneme and the PFR.

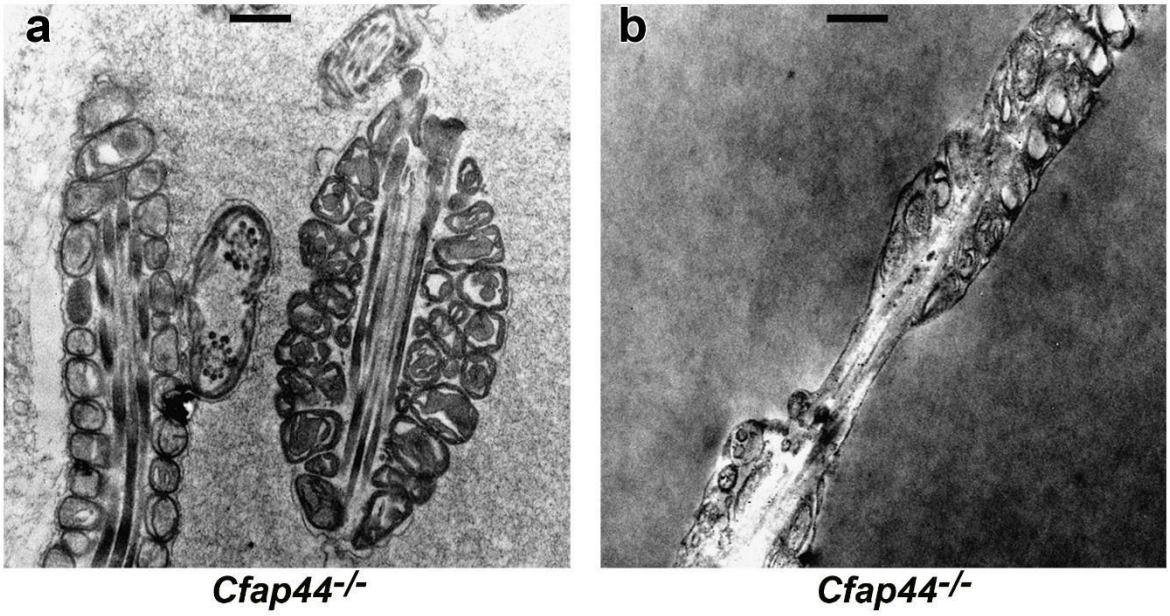


Supplementary Figure 6. Electron microscopy images of the ultrastructure of sperm from WT and Cfap44$^{-/-}$ male at low magnification.

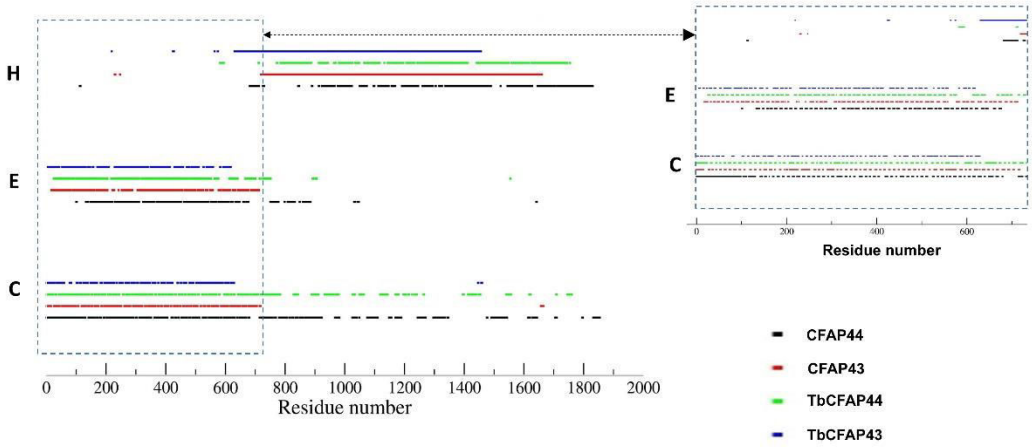
(a) Note the uniform structure of the WT sperm. **(b)** In contrast, in *Cfap44*$^{-/-}$ sperm, the overall shapes of the flagella are heterogeneous. Scale bars = 402 nm.



Supplementary Figure 7. TEM of sperm from Cfap43-/- exhibiting a disorganized cytoplasmic mass. Electron microscopy images of sperm from *Cfap43*$^{-/-}$ males show the absence of organized flagellum in short tail sperm. The short tail rather corresponds to an unorganized cytoplasmic mass containing the different components of the flagellum. scale bar = 2 µm.

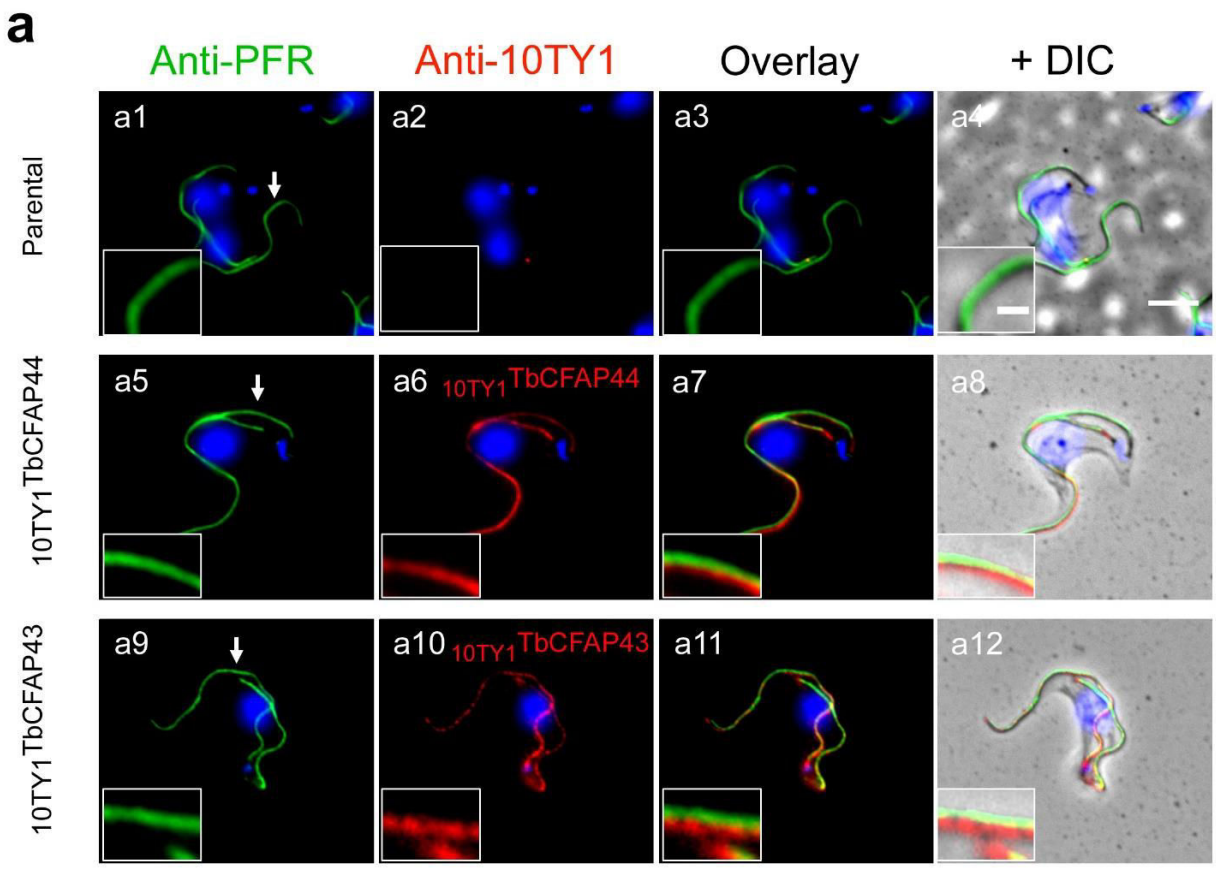


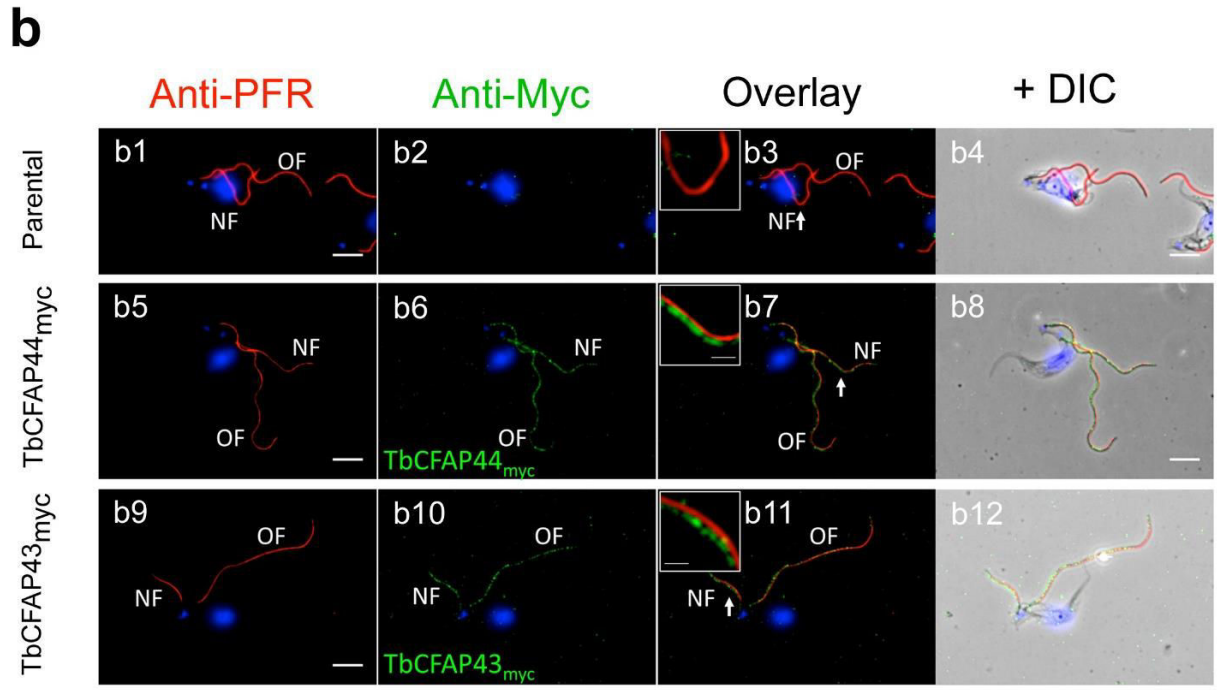
Supplementary Figure 8. Defective mitochondria organization in sperm from Cfap44$^{-/-}$
Longitudinal sections obtained by electron microscopy of sperm from *Cfap44*$^{-/-}$ males showing that the mitochondria are fragmented (**a**) and irregularly layered (**b**). Left image scale bar = 510 nm and right image scale bar = 395 nm.



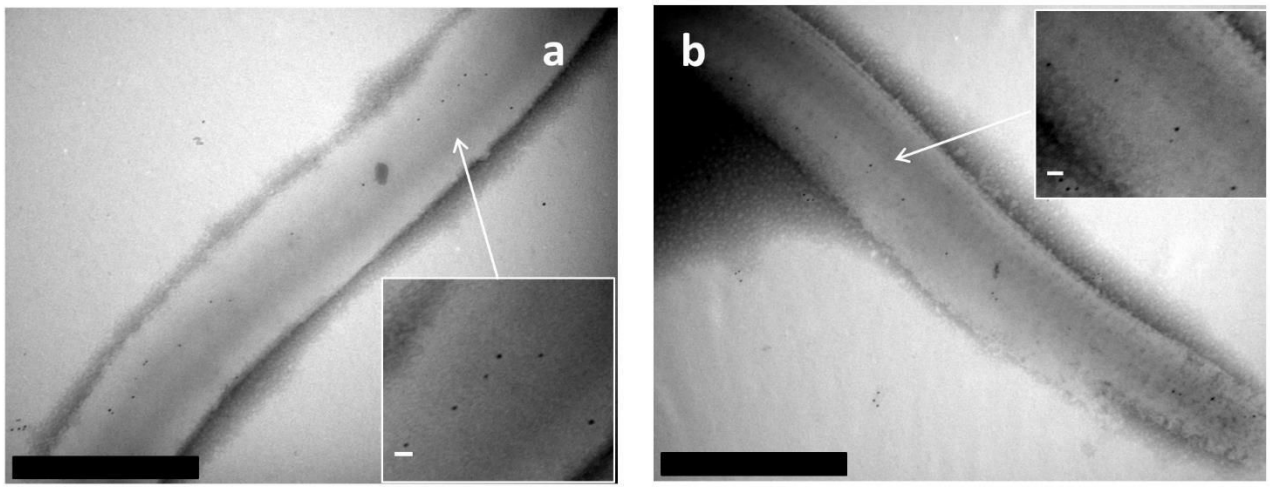
Coutton et al, Supplementary Fig. 9

Supplementary Figure 9. Secondary structure prediction. Elements of the secondary structures of the CFAP44, TbCFAP44, CFAP43, TbCFAP43 were predicted by Porter 4.0. C, E and H correspond to coils, extended conformations (β-sheets) and helices (α-helices), respectively. C, E, H are shown for all residues and a detail of the N-ter β-domain is highlighted in the inset.


a
Anti-PFR
Anti-10TY1
Overlay
+ DIC
Parental
10TY1 TbCFAP44
10TY1 TbCFAP43
a1
a2
a3
a4
a5
a6
10TY1 TbCFAP44
a7
a8
a9
a10
10TY1 TbCFAP43
a11
a12

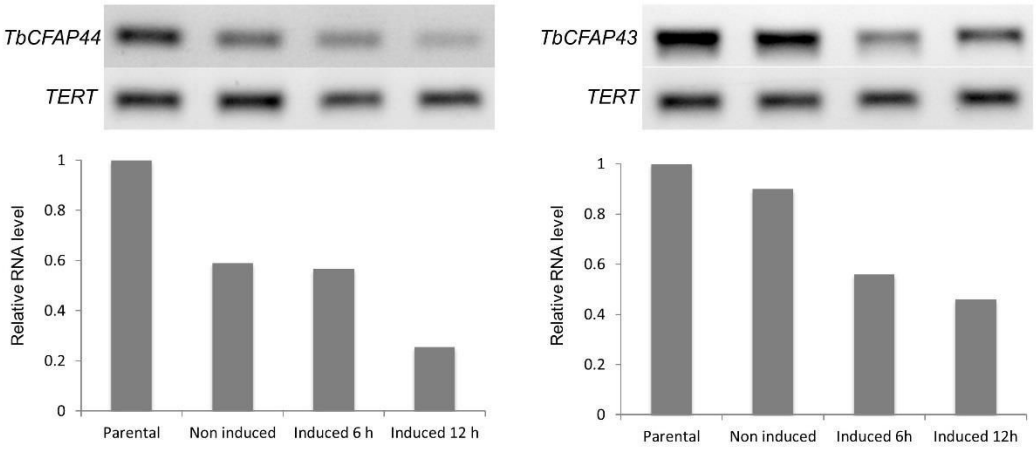

b
Anti-PFR
Anti-Myc
Overlay
+ DIC
Parental
TbCFAP44myc
TbCFAP43myc
b1
OF
NF
b2
b3
OF
NF
b4
b5
NF
OF
b6
NF
OF
TbCFAP44myc
b7
NF
OF
b8
b9
OF
NF
b10
OF
NF
TbCFAP43myc
b11
OF
NF
b12

Supplementary figure 10. Orthologs of T. brucei TbCFAP44 and TbCFAP43 are flagellar proteins. (a) Localization of $_{10TY1}$TbCFAP44 and $_{10TY1}$TbCFAP43 were observed by IF. **(a1-a4)** Immunofluorescence on detergent extracted cells of parental *T. brucei* (non expressing myc-tagged proteins) stained with anti-PFR (green) and anti-TY1 (red) antibodies. No TY1 staining was observed on parental cells, confirming the specificity of the anti-TY1 antibody. **(a5-a8)** Immunofluorescence on detergent extracted cells constitutively expressing 10TY1-tagged TbCFAP44 and non-induced for *TbCFAP43*^{RNAi}. The flagella were labelled with the anti-PFR (green), and 10TY1-tagged protein with anti-TY1 (green). Note the red staining of the entire flagellum showing that TbCFAP44 localisation is restricted to the flagellum. **(a9-a12)** Similar experiments as performed in **a5**-**a8** for TbCFAP43. Insets are enlargement of images of flagella taken from the main panels and display areas indicated by white arrows (scale bar 1 µm). Nuclei and kinetoplasts (mitochondrial genome) are labelled with DAPI (blue). Scale bars represent 5 µm. (b). Similar experiments performed with TbCFAP44$_{myc}$ and TbCFAP43$_{myc}$ proteins and stained with anti-PFR (red) and anti-myc (green) antibodies. An identical localization of both proteins is observed with both tags.



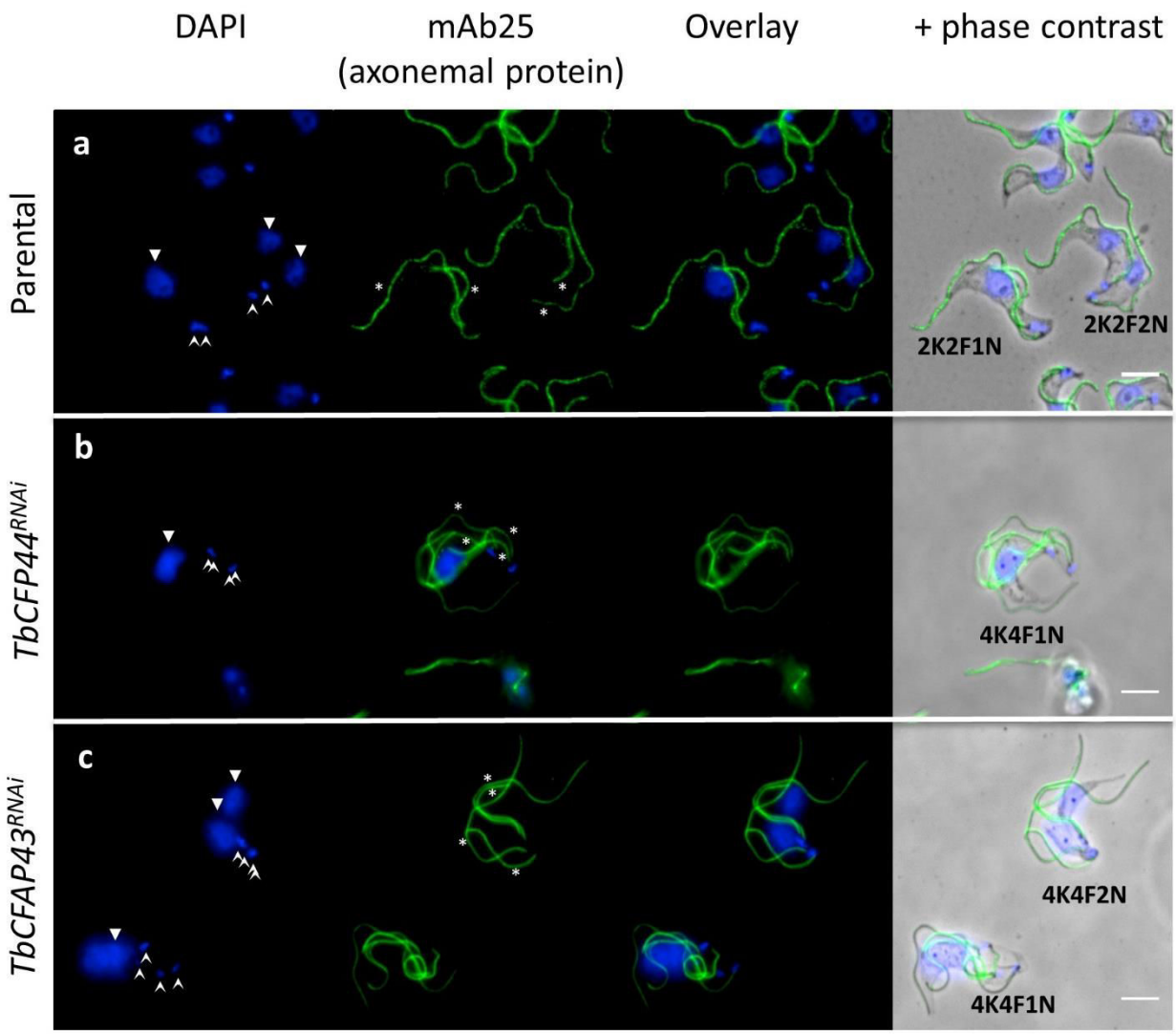
$_{10TY1}$TbCFAP44 $_{10TY1}$TbCFAP43

Supplementary Figure 11. Ultrastructural localization of ^{10TY1}TbCFAP$_{43}$ and ^{10TY1}TbCFAP$_{44}$. Electron immuno-gold labelling of detergent-extracted cytoskeleton cells expressing *$_{10TY1}$TbCFAP44* (a) and of *$_{10TY1}$TbCFAP43* (b). Gold beads size was 6 nm Scale bars represent 500 nm. Insets are enlargement of images of flagella taken from the main panels. Scale bars represent 10 nm.

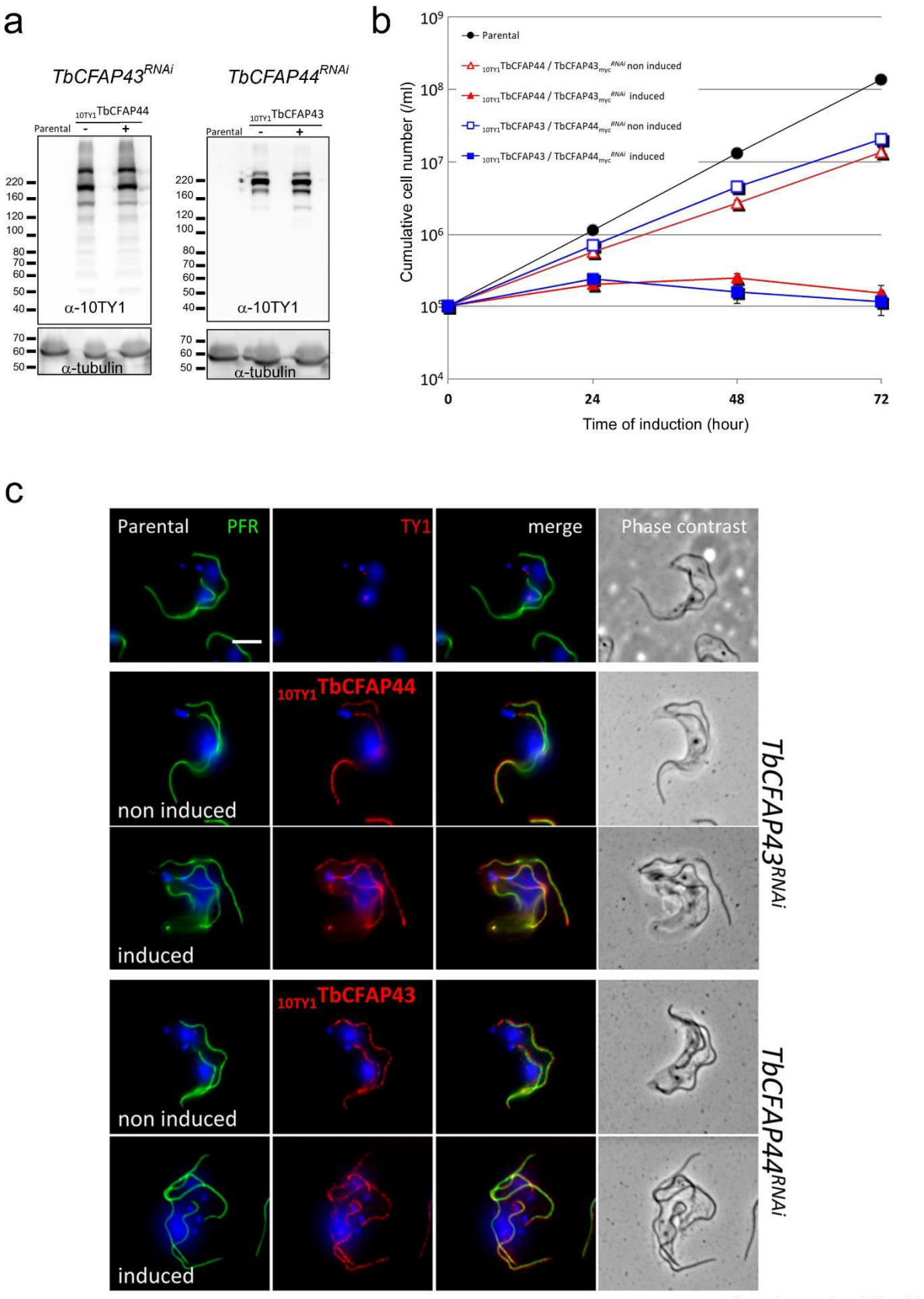


Supplementary Fig. 12

Supplementary Figure 12. Quantification of TbCFAP43 and TbCFAP44 RNAi knockdown. Total RNA was isolated from parental and non-induced cells, and from 6 and 12h RNAi induced cells. mRNA decrease upon RNAi induction was tested by semi-quantitative RT-PCR using primer sets specific to *TbCFAP44* (left panel) or *TbCFAP43* (right panel) and quantified (lower panel) relative to the telomerase reverse transcriptase (TERT) level.



Supplementary figure 13. Knockdown of TbCFAP43 and TbCFAP44 induces cytokinesis defects and produces multiflagellated cells. Parental cells **(a)**, and cells induced 24h for the RNAi of *TbCFPA44* **(b)** and *TbCFAP43* **(c)** were detergent extracted and methanol fixed. The flagella (F, asterisk) were immuno-labelled with mAb25, an anti-axonemal protein (green). Parental cells have 2 flagella whereas mutant cells have 4. The nuclei (N, triangle) and the kinetoplasts (K, arrow head) were labelled with DAPI. Scale bars 5 µm. 4K4F2N means cell with 4 kinetoplasts, 4 flagella and 2 nuclei.



Supplementary Fig. 14

Supplementary Figure 14. Expression of *TbCFAP43* does not rescue cell death induced by knock-down of *TbCFAP44* and vice versa. (a) Western-blot analysis of the expression of endogenously tagged $_{10TY1}$*TbCFAP44* and $_{10TY1}$*TbCFAP43* in the background of the cell line *TbCFAP43*$_{myc}$/*TbCFAP43* RNAi and *TbCFAP44*$_{myc}$/*TbCFAP44* RNAi respectively. RNAi was non induced (-) or induced (+) for 24 h. Expected MM of $_{10TY1}$*TbCFAP44* is 213.7 KDa, and of $_{10TY1}$*TbCFAP43* is 180 KDa. Loading was controled using anti-tubulin. Detergent-extracted cytoskeleton (2.10^7) were loaded on a 6% SDS-PAGE. **(b)** Growth curves for parental cells, and $_{10TY1}$*TbCFAP44*/*TbCFAP43*$_{myc}^{RNAi}$ and $_{10TY1}$*TbCFAP43*/*TbCFAP44*$_{myc}^{RNAi}$ cells, non-induced or induced with tetracycline. Cells were counted every 24 h. The graph represents the cumulative number of cells per ml. Error bars represent the standard error from 3 independent experiments. **(c)** Immunofluorescence on detergent extracted cells of parental *T. brucei* cell, $_{10TY1}$*TbCFAP44*/*TbCFAP43*$_{myc}^{RNAi}$ and $_{10TY1}$*TbCFAP43*/*TbCFAP44*$_{myc}^{RNAi}$ cells (non-induced or induced with tetracycline) and labelled with anti-PFR (green) and anti-TY1 (red) antibodies. Anti-TY1 labelling is still present on all flagella in RNAi induced cells. Nuclei and kinetoplasts (mitochondrial genome) are labelled with DAPI (blue). Scale bars represent 5 µm.

SUPPLEMENTARY TABLES

Supplementary Table 1. Sperm axonemal abnormalities observed by transmission electron microscopy (TEM) in individuals P_{43}-8 carrying mutations in CFAP43 and in individual P_{44}-3 carrying mutations in *CFAP44*.

Individual	Normal (9+2)	Abnormal	CPC defects (9+0)	CPC defects (9+1)	CPC defects (8+1)	CPC and DMTs defects
CFAP 43 (P_{43}-8)	4.6*	96.4	81.8	0	0	13.6
CFAP 44 (P_{44}-3)	4.8*	95.2	66.7	9.5	0	19
Control	100	0	0	0	0	0

Number of section analyzed: 23 sections for fertile control, 21 sections for CFAP44 mutated patient (1 patient) and 22 sections for CFAP43 patient (1 patient).

CPC central pair complexe; DMTs Doublets of microtubules

* Peri-axonemal structures (i.e. longitudinal columns) were mislocalized in all cases; a defect also observed in 4.3% of control sections.

Supplementary Table 2. Absence of CFAP43 or CFAP44 modulate intensities of staining of different axonemal substructures in mouse and human flagella.

The following antibodies were used to identify sub-components of the axoneme: α-RSPH1 and α-RSPH4A for radial spokes, α-GAS8 for N-DRC, α-DNALI1 for IDA, α-DNAI2 and α-DNAH5 for ODA, α-SPAG6 and α-SPEF2 for CP and α-AKAP4 for fibrous sheath. ++ intense staining, + intermediate staining, +/- weak

HUMAN	**Antibody**	**Fertile control**	***CFAP43 mutated* patients**	***CFAP44 mutated* patients**
Radial spoke	α-RSPH1	++	atypical	+/--
	α-RSPH4A	+	atypical	+/--
Central pair complex	α-SPAG6	++	-	Atypical weak
N-DRC	α-GAS8	+	+	+
IDA	α-DNALI1	++	++	++
ODA	α-DNAI2	++	++	++
Fibrous sheath	α-AKAP4	++	++	++

Supplementary Table 2 (continued)

MOUSE	Antibody	WT	*Cfap43*$^{-/-}$	*Cfap44*$^{-/-}$
Radial spoke	α-RSPH1	++	-	++
	α-RSPH4A	++	-	++
Central pair complex	α-SPEF2	+	-	+
N-DRC	α-GAS8	+	+	+
IDA	α-DNALI1	++	++	++
ODA	α-DNAH5	++	++	++

Supplementary Table 3. Protein folds found in the 4 target sequences.
WD40 folds are found for all N-ter domains of the proteins using the Superfamily program with very good confidence (e-value corresponds to the probability of finding such folds by chance in the sequence).

Protein	Family	Domains (aa number)	E-value
CFAP44	WD40 repeat-like WD40 repeat-like WD40 repeat-like	234-285,379-398,460-632 565-661,740-890 146-396	2.93e-40 3.71e-21 2.23e-20
CFAP43	WD40 repeat-like	110-216,347-503,626-713	3.4e-30
Prot-TbCFAP44	WD40 repeat-like WD40 repeat-like	75-249,276-338,392-462 389-567,672-748	1.46e-36 1.47e-15
Prot-TbCFAP43	WD40 repeat-like Tricorn protease domain 2	228-279,319-466,528-616 34-281	2.0e-12 1.8e-08

Supplementary Table 4. Primers used for *CFAP43* and *CFAP44*'s patients Sanger sequencing verification.

Primer name	Primer sequence (5'-3')	Amplicon Size (bp)
CFAP43-Ex8-F	AGGACAGTGCATCTATGCAAAC	516 bp
CFAP43-Ex8-F	TGCAGCATACTCTGCTCCTC	
CFAP43-Ex10-F	TTCAAAAACAACTGCCACCA	409 bp
CFAP43-Ex10-R	GCACACTGGAAAGGGAGGTA	
CFAP43-Ex11-F	TCCTGTGCCATCTTATGCAG	533 bp
CFAP43-Ex11-R	TGCAAGGACAGACCTGGGTA	
CFAP43-Ex16-F	AAGCCGTGTGGTAGAATTGC	464 bp
CFAP43-Ex16-R	ATCCTAAACCTGCCCAATCC	
CFAP43-Ex21-F	AGTGCAAGGTGGCTCAAAGT	439 bp
CFAP43-Ex21-R	GACTGAGCACAGAGGGAAGG	
CFAP43-Ex26-F	TCTAATGTTGGGTGGGAAGC	386 bp
CFAP43-Ex26-R	GCCTCCTCACTTAGCCTCCT	
CFAP43- Ex28-F	CAAGTGGGAGCATCTCTGGT	434 bp
CFAP43- Ex28-R	GTAAAGGGGAAGGTGCACAA	
CFAP43-Ex31-F	AGGTTTCCAAACCAAGTTGC	382 bp
CFAP43-Ex31-R	ACCGTGCTTACACAGATCCTT	
CFAP44-Ex12-F	ACTTGCGCAAAATCACACAG	421 bp
CFAP44-Ex12-R	GGAAAATCCCCAAATGCTCT	
CFAP44-Ex15-F	AATGGGGAAAATTGAGCAAG	401 bp
CFAP44-Ex15-R	CTGACCCTCACTGCAGGAAC	
CFAP44-Ex21-F	TTGAAACAGAGCCAATTCCA	530 bp
CFAP44-Ex21-R	GATGCCCAAAATAAATGGCTA	
CFAP44-Ex23-F	CAATCCAACTTTCATGACATCC	484 bp
CFAP44-Ex23-R	AGAGGGGCAAGAGATTGAGC	
CFAP44-Ex31-F	GGACATGTCTTCTGGGCTCT	443 bp
CFAP44-Ex31-R	TGAAAAGCTTGGGTGTGATG	

Supplementary Table 5. Primers used for RT-qPCR of *CFAP43 and CFAP44* in human and mouse.

Primer names	Primer Sequence (5'-3')	Tm (C°)
CFAP43_5F	CTGGTAGGCAAAGAGGCAGA	60
CFAP43_6R	TCACAGCCAATGTACAAGTC	
Cfap43_2F	ATGGCAACTAATGTGCCCAGT	60
Cfap43_3R	GCAAAGTAGAGTGTAGTCCAGGA	
CFAP44_5F	CTACCTGCGAAGTAGCAGTG	60
CFAP44_6R	CGAAGGACTCTGTATGGTCT	
Cfap44_15F	TTGCCAGTTGATGTGGTCTC	60
Cfap44_16R	ACACGTTCGGGATTGCAAAC	
ACTB_F	CCAACCGCGAGAAGATGA	60
ACTB_R	CCAGAGGCGTACAGGGATAG	
Actb_F	ACCAGAGGCATACAGGGACA	60
Actb_R	CTAAGGCCAACCGTGAAAAG	

Supplementary Table 6. List of primary antibodies used in immunofluorescence experiments in human, mouse and *Trypanosoma*

Primary antibodies	Reference	Species	Protein localization	Dilution
Human experiments				
DNAI2	Abnova H00064446-M01	mouse	ODA	1/400
DNALI1	Sigma-Aldrich® HPA028305	rabbit	IDA	1/100
RSPH1	Sigma-Aldrich® HPA017382	rabbit	RS, head	1/100
RSPH4A	Sigma-Aldrich® HPA031196	rabbit	RS, head	1/100
GAS8	Sigma-Aldrich® HPA041311	rabbit	N-DRC	1/100
SPAG6	Sigma-Aldrich® HPA038440	rabbit	CPC	1/500
AKAP4	Sigma-Aldrich® HPA020046	rabbit	FS	1/100
Acetylated tubulin	Sigma-Aldrich® T7451	mouse	Microtubules	1/2000
Mouse experiments				
Rsph1	Sigma-Aldrich® HPA017382	rabbit	RS, head	1/100
Rsph4a	Sigma-Aldrich® HPA031196	rabbit	RS, head	1/50
Gas8	Sigma-Aldrich® HPA041311	rabbit	N-DRC	1/100
Spef2	Sigma-Aldrich® HPA040343	rabbit	CPC	1/1000
Dnali1	Sigma-Aldrich® HPA028305	rabbit	IDA	1/100
Dnah5	Sigma-Aldrich® HPA037470	rabbit	ODA	1/100
Mpc1l	Described in (1)	rabbit	Mitochondria	1/500
Acetylated tubulin	Sigma-Aldrich®® T7451	mouse	Microtubules	1/500
***T brucei* experiments**				
myc	Santa Cruz monoclonal 9E10, sc-40	mouse	Myc-tagged proteins	1/10
PFR2	Gift of Dr N. Biteau (2)	rabbit	PFR / flagellum	1/2000; 1/400 (STED)
mAb25 (TbSAXO)	Described in (3)	mouse	Microtubules/axoneme	1/5
BB2 (TY1)	Described in (4)	mouse	TY1-tagged proteins	1/500 (IF); 1/10 (iEM)
TAT1 (tubulin)	Described in (4)	mouse	Cytoskeleton / flagellum	1/50 (IF); 1/10 (iEM)
L8C4 (PFR2)	Described in	mouse	PFR / flagellum	neat

Abbreviations are as follows : ODA, Outer Dynein Arms ; IDA, Inner Dynein Arms ; RS, Radial Spokes ; FS, Fibrous sheath ; CPC, Central Pair Complex;

(1) Vanderperre et al, (2016) MPC1-like Is a Placental Mammal-specific Mitochondrial Pyruvate Carrier Subunit Expressed in Postmeiotic Male Germ Cells.J Biol Chem. 2016 Aug 5;291(32):16448-61. doi: 10.1074/jbc.M116.733840.

(2) Laboratoire de Microbiologie Fondamentale et Pathogénicité UMR5234-CNRS F-33076 BORDEAUX

(3) D. Dacheux et al., (2012) A MAP6-Related Protein Is Present in Protozoa and Is Involved in Flagellum Motility," PLoS One 7 (2): e31344, doi:10.1371/journal.pone.0031344.

(4) Bastin et al., "A Novel Epitope Tag System to Study Protein Targeting and Organelle Biogenesis in Trypanosoma Brucei," *Mol Biochem Parasitol* 77, no. 2 (1996): 235–39.

(5) Linda Kohl, Trevor Sherwin, and Keith Gull, "Assembly of the Paraflagellar Rod and the Flagellum Attachment Zone Complex During the Trypanosoma Brucei Cell Cycle," *Journal of Eukaryotic Microbiology* 46, no. 2 (March 1, 1999): 105–9, doi:10.1111/j.1550-7408.1999.tb04592.x.

Supplementary Table 7. Endogenous tagging and RNAi primer sequences for *T. brucei* experiments

3xmyc C-terminal endogenous tagging primers
TbCFAP44
5'-
GGAAATTGATGCACTGAGAACAGAAGTTGCACTCCTACGTACGAAGGGTGGACACGTGTATGCCGCTGCTAT
GGCAGCCGGCAGGgtaccgggccccccctcgag-3'

5'-
CACTCCTTCCCAAATACATACTGGTATTGAAGCGCTTCGCTGTTCGTGATGGTGCCAAGGCACAACATAGAGAG
AAACTAGCAGCtggcggccgctctagaactagtggat-3'

TbCFAP43
5'-
ACTTGTTCGCCTTAAAAACGAGGTAGATCGCTTGCGTGAGGCCACATTTCCCTCATTTGCTGTTGTTACGAGGA
GGACCGCCAGAggtaccgggccccccctcgag-3'

5'-
CTTCCTTCTGCCACTTGTCCGAATAACGCCGACTTCTGCCTCACTTGCCCCCCCTCGCCAGTGTCAGCAGGGCAC
CGCCAtggcggccgctctagaactagtggat-3'

RNAi primers:
TbCFAP44 (target bp 1880-2309 within the ORF)
5'-GCCGGCCGCTCTAGAgggccggtggagtagagct-3'
5'-TAAGCTTGCTCTAGAggttggagttgggcaacagg-3'

TbCFAP43 (target bp 131-596 within ORF)
5'-GCCGGCCGCTCTAGActtgtgaaggccgc-3'
5'-TAAGCTTGCTCTAGAgaccctcagtgccccatacg-3'

RT-PCR primers
TbCFAP44
5'-GGCTATTTTCGCTTCTGGCG -3'
5'-GCTTCGTCACAGCGTTGTTT -3'

TbCFAP43
5'-AGAGACTGAACAGGCGCTTC -3'
5'-TCGCTCACCATAGCGTTCTC -3'
TbTERT
5'-GAGCGTGTGACTTCCGAAGG-3'
5'-AGGAACTGTCACGGAGTTTGC -5'

Supplementary Table 8. Guide RNA sequences for CRISPR/Cas9 mice generation.

Gene	Exon targeted	Guide RNA sequence (5'-3')
Cfap43	exon 2	AATGTGCCCAGTGAAGTCG
	exon 21	CTCCATGGCTGTGAAGGGC
Cfap44	exon 3	CAGCTGACTCATATACCGA
	exon 15	AGAAAGAATACAAGCCGA

Supplementary Table 9. Primers for CRISPR/Cas9 mice genotyping.

Gene	Method	PCR Primer Name	PCR Primer sequence	Tm (°C)
Cfap43	Sanger sequencing	Wdr96m_2F	CAAAGATGGGTGCAAGGATT	56
		Wdr96m_2R	AGCAGGGGGAAGAATGAACT	56
		Wdr96m_21F	GTGCAGGGGAATCCTGTAAA	58
		Wdr96m_21R	AGCAGGGGGAAGAATGAACT	59
	HRM	96HRM_21F	TGTGGAGATGCACAACTTGG	58
		96HRM_21R	GTCGGAGGCTGTGTGTCTCT	61
Cfap44	Sanger sequencing	Wdr52mT1_3F	TCTATACACGAAGCTTCACAAGACA	60
		Wdr52mT1_3R	AGAAACAGAATGTAGAACCCCTGA	59
	HRM	52HRM_3F	CGAAGCTTCACAAGACACCA	58
		52RHRM_3R	GCAGGAGCCTGAACGTACTC	60

Supplementary Table 10. List of primers used for RT-PCR experiments for *Cfap43 and Cfap44.*

Primer name	Exons	Primer sequence (5'-3')	Tm
Cfap43_E20-21F	20-21	GGAGGAAGTAGCGAAGATAA	60
Cfap43_E22R	22	CAACTCCTGTAGCTCCTCTT	60
Cfap44_E2F	2	ATATGAAGGAGCCAGATGAC	60
Cfap44_E3R	3	CGTACTCTTGAGATGAATGC	60

3- Article 5 : Whole-exome sequencing identifies mutations in *FSIP2* as a recurrent cause of multiple morphological abnormalities of the sperm flagella

Guillaume Martinez, **Zine-Eddine Kherraf**, Raoudha Zouari, Selima Fourati Ben Mustapha, Antoine Saut, Karin Pernet-Gallay, Anne Bertrand, Marie Bidart, Jean Pascal Hograindleur, Amir Amiri-Yekta, Mahmoud Kharouf, Thomas Karaouzène, Nicolas Thierry-Mieg, Denis Dacheux-Deschamps, Véronique Satre, Mélanie Bonhivers, Aminata Touré, Christophe Arnoult, Pierre F. Ray & Charles Coutton.

Human Reproduction. Accepté.

Contexte et principaux résultats de l'étude :

L'analyse des données issues du séquençage exomique de la cohorte de 78 patients MMAF nous a permis d'identifier un nouveau gène candidat. En effet, des mutations perte de fonction homozygotes du gène *FSIP2* ont été identifiées chez 4 sujets non apparentés. Après les quatre gènes précédemment décrits (*DNAH1*, *CFAP43*, *CFAP44* puis *WDR66*), il s'agit du 5ème gène candidat identifié dans cette cohorte. Aucune des 4 mutations identifiées n'est décrite dans les bases de données. Nous avons validé ces mutations par séquençage Sanger et avons montré par qRT-PCR à partir d'un panel de tissus humains et murins que le gène *FSIP2* est spécifiquement exprimé dans le testicule. De plus, la protéine codée par ce gène est décrite comme interagissant avec la gaine fibreuse, une structure para-axonémale spécifique du flagelle spermatique.

Nous avons réalisé une analyse approfondie de plusieurs protéines flagellaires par immunofluorescence et microscopie confocale sur des spermatozoïdes provenant des patients mutés pour *FSIP2*. De manière intéressante, nous avons observé que la protéine AKAP4 était absente des flagelles des patients présentant une mutation de *FSIP2* alors qu'elle était présente chez les autres patients MMAF mutés pour un autre gène (*DNAH1* ou *CFAP43* ou *CFAP44*). AKAP4 est une protéine de la gaine fibreuse qui porte un domaine de liaison à FSIP2 et a été précédemment décrite comme étant impliqué dans le phénotype MMAF chez la souris et l'Homme. Ce résultat confirme donc la forte interaction entre FSIP2 et AKAP4. L'étude de l'ultrastructure flagellaire de ces spermatozoïdes par microcopie électronique a révélé une dysplasie de la gaine fibreuse et une désorganisation profonde de la structure axonémale ce qui suggère le rôle important de ce nouveau gène dans l'assemblage et l'élongation du flagelle spermatique.

Dans cet article, nous avons donc identifié un cinquième gène impliqué dans le phénotype MMAF. Ce travail permet par ailleurs d'établir que les altérations des protéines de la gaine fibreuse comme les altérations des protéines de l'axonème entrainent la survenue d'un phénotype spermatique MMAF.

Whole-exome sequencing identifies mutations in FSIP2 as a recurrent cause of multiple morphological abnormalities of the sperm flagella.

Running title: Loss of function mutations in *FSIP2* cause MMAF phenotype

Guillaume Martinez[1,2], Zine-Eddine Kherraf[1,3], Raoudha Zouari[4], Selima Fourati Ben Mustapha[4], Antoine Saut[1,2], Karin Pernet-Gallay[5], Anne Bertrand[5], Marie Bidart[6], Jean Pascal Hograindleur[1], Amir Amiri-Yekta[1,6,7], Mahmoud Kharouf[4], Thomas Karaouzène[1,8], Nicolas Thierry-Mieg[8], Denis Dacheux-Deschamps[9,10], Véronique Satre[1,2], Mélanie Bonhivers[9,10], Aminata Touré[11,12,13], Christophe Arnoult[1], Pierre F. Ray[1,3,$*] and Charles Coutton[1,2,$]

[1] University Grenoble Alpes, INSERM U1209, CNRS UMR 5309, Institute for Advanced Biosciences, Team Genetics Epigenetics and Therapies of Infertility, 38000 Grenoble, France.

[2] CHU Grenoble Alpes, UM de Génétique Chromosomique, Grenoble, France.

[3] CHU de Grenoble, UM GI-DPI, Grenoble, F-38000, France

[4] Clinique des Jasmins, 23, Av. Louis BRAILLE 1002 Belvedere, Tunis, Tunisia

[5] Grenoble Neuroscience Institute, INSERM 1216, Grenoble, F38000, France

[6] CHU Grenoble Alpes, UM de Biochimie Génétique et Moléculaire, Grenoble, France.

[7] Department of Genetics, Reproductive Biomedicine Research Center, Royan Institute for Reproductive Biomedicine, ACECR, Tehran, Iran.

[8] Univ. Grenoble Alpes / CNRS, TIMC-IMAG, F-38000 Grenoble, France

[9] Université de Bordeaux, Microbiologie Fondamentale et Pathogénicité, CNRS UMR 5234, Bordeaux, France.

[10] Institut Polytechnique de Bordeaux, Microbiologie Fondamentale et Pathogénicité, UMR-CNRS 5234, F-33000 Bordeaux, France.

[11] INSERM U1016, Institut Cochin, Paris 75014, France.

[12] Centre National de la Recherche Scientifique UMR8104, Paris 75014, France.

[13] Faculté de Médecine, Université Paris Descartes, Sorbonne Paris Cité, Paris 75014, France.

[$] Equal contribution, the authors should both be regarded as last authors.

* Corresponding author: Pierre F Ray, UM GI-DPI, IBP, CHU Grenoble Alpes, 38043 Grenoble cedex 9, France. Tel: +33-4-76-76-55-73; E-mail: pray@chu-grenoble.fr

Abstract

STUDY QUESTION: Can whole-exome sequencing (WES) of infertile patients identify new genes responsible for multiple morphological abnormalities of the sperm flagella (MMAF)?

SUMMARY ANSWER: WES analysis of a large cohort of 78 infertile men with a MMAF phenotype permitted the identification of 4 homozygous mutations in the Fibrous Sheath Interacting Protein 2 (*FSIP2)* gene in 4 unrelated individuals.

WHAT IS KNOWN ALREADY: The use of high-throughput sequencing techniques revealed that mutations in *DNAH1*, *CFAP43* and *CFAP44,* account for approximately one third of MMAF cases thus indicating that other relevant genes await identification.

STUDY DESIGN, SIZE, DURATION: This was a retrospective genetics study of 78 patients presenting a MMAF phenotype recruited in France, Tunisia and Iran between 2008 and 2015.

PARTICIPANTS/MATERIALS, SETTINGS, METHOD: WES was performed for all 78 subjects. Relative expression of the selected candidate gene was assessed by RT-qPCR in a panel of normal human and mouse tissues. Immunofluorescence (IF) and Transmission Electron Microscopy (TEM) analyses were performed to characterize the structural and ultrastructural anomalies present in sperm flagellum of patients. All identified variants were confirmed by Sanger sequencing.

MAIN RESULTS AND THE ROLE OF CHANCE: We identified four unrelated patients (4/78, 5.1%) with homozygous loss of function mutations in the *FSIP2* gene, which encodes a protein of the sperm fibrous sheath (FS) specifically expressed in human and mouse testis. None of these mutations were reported in control sequence databases. TEM analyses showed a complete disorganization of the FS associated with axonemal defects. IF analyses confirmed that the central-pair microtubules and the inner and outer dynein arms of the axoneme were abnormal in patients carrying *FSIP2* mutations. Importantly, and in contrast to what was observed in MMAF patients with mutations in other MMAF-related genes (*DNAH1*, *CFAP43* and *CFAP44*), mutations in *FSIP2* led to the absence of AKAP4, a protein which interacts with FSIP2 and which was previously described in mouse and in humans to be associated with a MMAF-like phenotype.

LIMITATIONS, REASONS FOR CAUTION: The low number of biological samples and the absence of a reliable anti-FSIP2 antibodies prevented to formally demonstrate that the FSIP2 protein was absent in sperm from mutated subjects.

WIDER IMPLICATIONS OF THE FINDINGS: Our findings indicate that FSIP2 is one of the main genes involved in MMAF syndrome. In humans, previous genes associated with a MMAF phenotype encode for axonemal-associated proteins (*DNAH1*, *CFAP43* and *CFAP44*). We show here that FISP2, a protein of the fibrous sheath, is necessary for its own assembly but also for the overall axonemal and flagellar biogenesis in man. Last, this work reinforces the demonstration that WES sequencing is a good strategy to reach a genetic diagnosis for patients with severe male infertility phenotypes.

STUDY FUNDING/COMPETING INTEREST(S): This work was supported by the following grants: the 'MAS-Flagella' project financed by the French ANR and the DGOS for

the program PRTS 2014 (**14-CE15**) and the 'Whole genome sequencing of patients with Flagellar Growth Defects (FGD)' project financed by the Fondation Maladies Rares for the program Séquençage à haut débit 2012. The authors have no conflict of interest.

Key words: Male infertility, genetic diagnosis, gene mutation, flagellum, fibrous sheath, MMAF, *FSIP2*, *AKAP4*, whole-exome sequencing.

Introduction

Teratozoospermia, a condition defined by the presence of sperm with abnormal morphology, regroups a large variety of phenotypes, many of which are among the most severe infertility phenotypes in human (Coutton *et al.*, 2015). Multiple morphological abnormalities of the sperm flagella (MMAF phenotype) is a subgroup of teratozoospermia characterized by the presence in the ejaculate of immotile spermatozoa with several abnormalities of the sperm flagellum including short, coiled, absent and flagella of irregular caliber (for illustration see Fig. 1) (Ray *et al.*, 2017). Similar phenotypes were previously studied and described as "dysplasia of the fibrous sheath (DFS)," "short tails," or "stump tails" (Coutton *et al.*, 2015). Mutations in *DNAH1*, *CFAP43*, *CFAP44* and *CFAP69* have recently been reported to be recurrent genetic causes of MMAF in human highlighting the genetic heterogeneity of this phenotype (Ben Khelifa *et al.*, 2014; Amiri-Yekta *et al.*, 2016; Sha *et al.*, 2017; Tang *et al.*, 2017, Dong et al., 2018). These genes encode different components of the axoneme, a highly conserved central cytoskeletal structure shared by both the sperm flagellum and other motile cilia (Inaba, 2007; Satir and Christensen, 2008). In mouse models, in addition to axonemal genes, mutations in several genes coding for accessory structures of the flagellum have also been associated with sperm tail defects (Lehti and Sironen, 2017). Interestingly, a first genetic diagnosis was proposed twelve years ago for a patients described with DFS. He presented a deletion removing part of *AKAP3* and *AKAP4,* two proteins of the sperm fibrous sheath (Baccetti *et al.*, 2005), one of the main peri-axonemal structures surrounding the flagellum axoneme (Inaba, 2003, 2011). No other mutations in *AKAP* genes nor in any other FS-related genes were however reported since this initial description of an AKAP3/4 deletion (Turner *et al.*, 2001; Coutton *et al.*, 2015; Ray *et al.*, 2017). More than 20 proteins are estimated to be located in, or closely associated to the fibrous sheath of mammalian spermatozoa (Eddy, 2007).

Altogether, these proteins are involved in the control of sperm flagellum elasticity, glycolytic activities and phosphorylation signaling pathways that regulate sperm motility (Miki *et al.*, 2002; Eddy, 2007; Inaba, 2011).

Here we analyzed the genetic data obtained by whole-exome sequencing (WES) from a large cohort of 78 infertile MMAF patients, and identified 4 unrelated subjects carrying bi-allelic loss of function mutations in the Fibrous Sheath Interacting Protein 2 (FSIP2), which encode a FS protein interacting with AKAP4. Importantly, we observed in *FSIP2* mutated individuals, that the AKAP4 protein was absent from all spermatozoa; a feature which was not observed in spermatozoa from other MMAF individuals of different genetic etiology. Overall, our work confirms what has been suggested long ago: that not only axonemal components, but also peri-axonemal proteins from the FS play an important role in the assembly and/or the stability of the axoneme (Escalier and David, 1984; Chemes *et al.*, 1987, 1998) and that genetic defects interfering with the competent synthesis of these proteins induce a MMAF phenotype

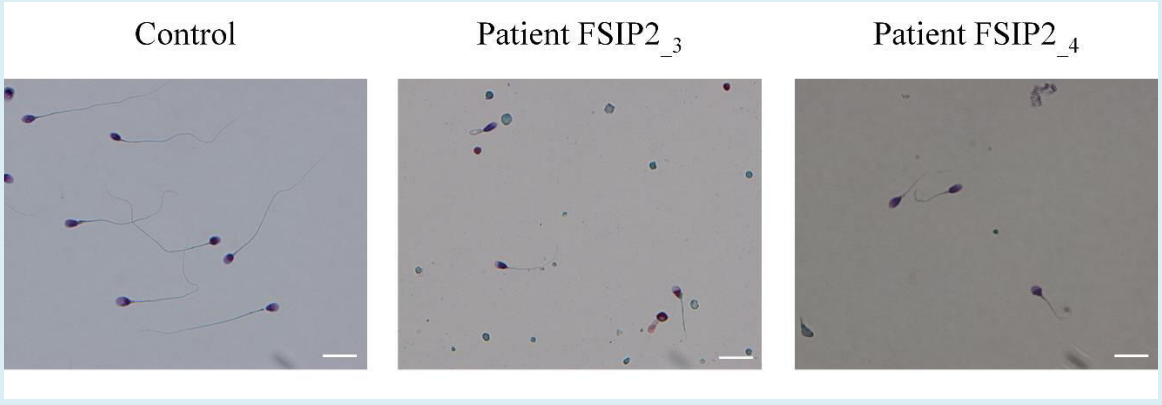


Figure 1. Morphology of spermatozoa from *FSIP2* mutated patients.
Morphology of spermatozoa from FSIP2_3 and FSIP2_4 mutated patients. All spermatozoa from mutated patients have shorter flagella than controls. Additional features of MMAF spermatozoa such as thick and coiled flagella are also observed. Scale bars 10 µm.

Material and Methods

Full materials and methods is provided online.

Subjects and controls

WES was performed on a large cohort of 78 MMAF patients as previously described in Coutton *et al.,* (2018). All patients presented with a typical MMAF phenotype defined by asthenozoospermia (total motility <40%) with >5% of at least three flagellar morphological abnormalities (absent, short, coiled, bent and irregular flagella). The four FSIP2 patients (FSIP2_1 to FSIP2_4) originated from North Africa (Tunisians) and were treated in Tunis at the Clinique des Jasmins for primary infertility from 2008 to 2015.

Whole-Exome Sequencing (WES) and Sanger sequencing

WES was performed according to our protocol described in Coutton *et al.,* (2018). All *FSIP2* mutations identified by WES were validated by Sanger sequencing. PCR primers and protocols used for each patient are listed in the Supplementary Table SI. Sequences analyses were carried out on an ABI 3130XL (Applied Biosystems). Sequences were analysed using the SeqScape software (Applied Biosystems).

Quantitative real-time RT-PCR analysis

RT-qPCR was performed with cDNAs from various tissues of human and mouse, including testes (Amsbio, Abingdon, UK). Each sample was assayed in triplicate for each gene on a StepOnePlus (LifeTechnologies®), with Power SYBR®Green PCR Master Mix (Life Technologies®). RT-qPCR data were normalized using the reference housekeeping gene *ACTB* for human with the -ΔΔCt method (Livak and Schmittgen, 2001). The 2-ΔΔCt value was set at

0 in brain cells, resulting in an arbitrary expression of 1. Primers sequences and RT-qPCR conditions are indicated in the Supplementary Table SII.

Immunostaining in human sperm cells

Immunofluorescence (IF) staining was performed on sperm cells from fertile control individuals, FSIP2_3 and FSIP2_4 patients and MMAF patients with mutation in other MMAF-causing genes (*CFAP43*, *CFAP44* and *DNAH1*).

Transmission electron microscopy of human sperm cells

Transmission electron microscopy (TEM) experiments were performed using sperm cells from fertile control individuals and from patient FSIP2_4 carrying the *FSIP2* c.16389_16392delAATA homozygous variant.

Results

Whole-exome sequencing (WES) identifies homozygous truncating mutations in *FSIP2* in four MMAF patients.

We previously described the genetic analysis of a cohort of 78 MMAF individuals permitting the identification of bi-allelic mutation in *DNAH1*, *CFAP43* or *CFAP44* in a total of 28 subjects (Coutton *et al.,* 2018). Additional work permitted the identification of homozygous loss of function mutations in the *CFAP69* gene of two unrelated patients from the cohort (Dong et al. 2018). Here we searched for novel genes associated with MMAF by performing an extensive reanalysis of the remaining exomes and applying stringent technical and biological filters. We identified four patients (4/78, 5.1%) with homozygous truncating mutations in the *FSIP2* gene. The *FSIP2* gene (NM_173651) encodes a 6907-amino acid protein (Q5CZC0) and is described as one of the main gene of the flagella FS. All identified variants: c.910delC, c.16389_16392delAATA, c.2282dupA and a complex variant c.1606_1607insTGT; 1607_1616delAAAGATTGCA are frameshift variants inserting a premature stop codon likely abrogating the production of a functional protein. Details of the variants is provided in Fig. 2 and in Supplementary Table SIII. All the identified variants were absent from control sequence databases (dbSNP, 1000 Genomes Project, NHLBI Exome Variant Server, gnomAD and in-house database), which is consistent with the expected key role of *FSIP2* in male reproduction and the negative selection of the mutations.

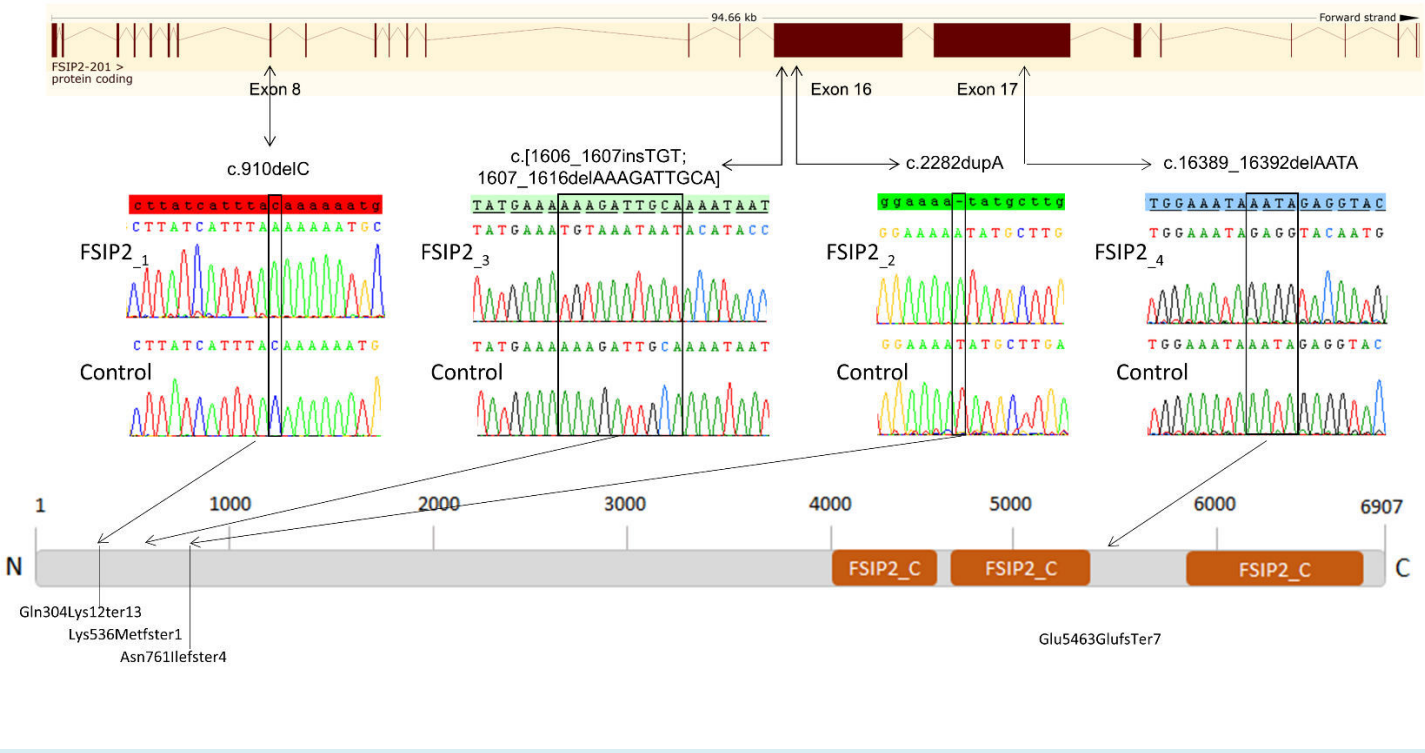


Figure 2. Location of *FSIP2* mutations in the intron-exon structure and in the protein representation. Electropherograms of Sanger sequencing for all FSIP2 mutated patients compared to reference sequence are reported. Orange square stands for FSIP2_C domains (Fibrous sheath-interacting protein 2, C-terminal, IPR031554) as predicted by InterPro server (https://www.ebi.ac.uk/interpro/). Mutations are annotated in accordance to the HGVS's recommendations. According to the InterPro server, three homologous domains called FSIP2_C (C-terminus of the FSIP2 protein) are present in the C terminal part of the protein. FSIP2_C domains are present in other proteins and have been found to be repeated up to 10 times but their function is not yet known (Finn et al., 2017).

The *FSIP2* gene was reported in public expression databases to be strongly expressed in the testis and to be connected with the sperm flagellum (Brown *et al.*, 2003). We performed RT-qPCR experiments which confirmed that, in mouse and human, *FSIP2* is exclusively expressed in the testis (Fig. S1).

In conclusion, taking into account: i.) the identification of several unrelated patients carrying homozygous *FSIP2* truncating mutations absent from all control sequences databases, ii.) the absence in these patients of other pathogenic variants in genes significantly expressed in testis and/or related to cilia/flagellum iii.) *FSIP2* specific expression pattern, and iv.) data

from the literature reporting the involvement of AKAP4, a protein interacting with FSIP2, in human and mouse MMAF-like phenotype, we conclude that the identified *FSIP2* mutations are responsible for the MMAF phenotype of patients FSIP_1-4.

Genotype-phenotype correlation

As observed for the other MMAF individuals of the cohort, analysis of sperm morphology from patients FSIP_1-4 revealed very few spermatozoa with normal flagella at the expense of spermatozoa with short, absent or irregular flagella (Fig. 1). The semen parameters of the overall MMAF cohort and the detailed semen parameters of FSIP2 patients (patients FSIP_1-4) are presented in Tables I and II. Interestingly, we observed that FSIP2 patients (FSIP_1-4) presented a significant lower percentage of spermatozoa with absent flagellum compared to MMAF patients with mutations in axonemal genes *DNAH1*, *CFAP43* or *CFAP44* (t-test; $P<0.05$) but no other differences were observed in the semen parameters between the two groups (Tables I and II).

***FSIP2* mutations lead to severe disorganization of the fibrous sheath and axonemal structure**

We next studied the ultrastructure of sperm cells from patient FSIP2_4 by transmission electron microscopy (TEM) (Fig. 3). Due to low sperm concentration, only few usable images could be obtained. Longitudinal sections showed severe FS defects which appeared completely disorganized and dysplastic (Fig. 3A, B). ODFs remained present whereas the mitochondrial sheath was totally absent from the midpiece and, as a consequence, the hypertrophic FS extended up to the sperm neck (Fig. 3B). Cross-sections permitting to observe the ultrastructure of the axonemal components were also analyzed (n=15). In the midpiece of spermatozoa from fertile controls, the axoneme is surrounded by 9 outer dense fibers ODFs and the mitochondrial

sheath (Fig. 3C); while in the principal piece, the axoneme is surrounded by 7 ODFs and flanked by the fibrous sheath composed of two longitudinal columns connected by circumferential ribs (ODFs 3 and 8 are replaced by the two longitudinal columns) (Fig. 3D). In *FSIP2* patients, 100% of the analyzed cross-sections showed redundant and randomly arranged FS without the expected organization in longitudinal columns and circumferential ribs (Fig. 3E-H). In some principal-piece cross-sections, we numbered nine ODFs indicating that the outer dense fibers 3 and 8 were not replaced by the longitudinal columns (Fig. 3H). ODFs organization was frequently altered with the presence of supernumerary or missing ODFs (Fig. 3E-G). Mitochondrial sheaths were not visible in any cross-sections. IF experiments with anti-HSP60, which detects a mitochondrial protein located in the midpiece (Magnoni et al., 2014), showed that mitochondria are still present but strongly disorganized and scattered all along the sperm flagellum (Fig. S2). This suggests that the absence of the mitochondrial sheath is not due to the absence of mitochondria but may be due to the impaired caudal migration of the annulus during spermiogenesis. The axoneme also showed a constant abnormal conformation with variable degrees of defects ranging from the absence of the central pair (CP) (9+0 conformation) to a complete disorganization (Fig. 3E-H). In addition to these obvious flagellar defects, we also observed some severe nuclear alterations particularly in chromatin texture (data not shown).

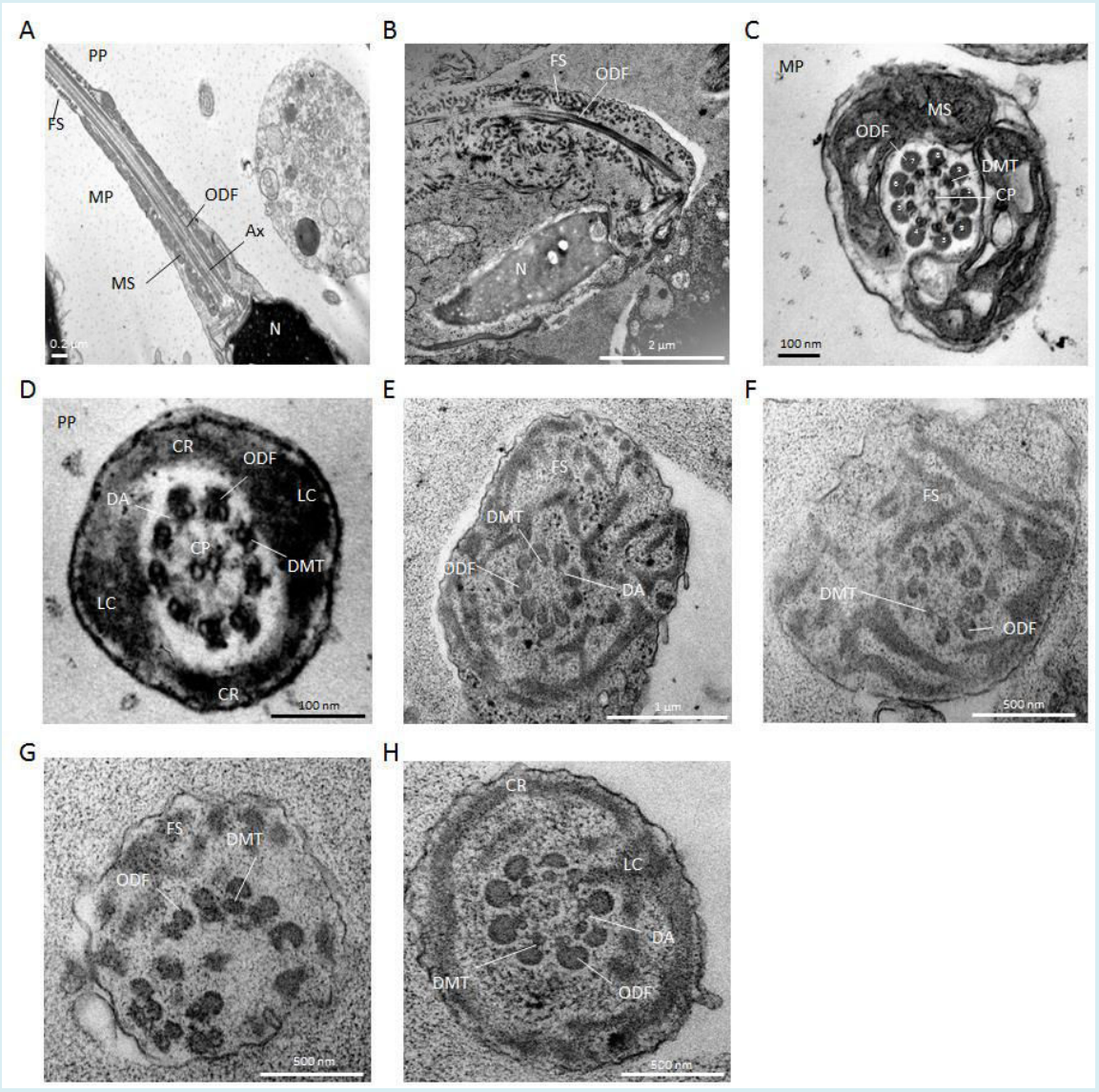


Figure 3. Transmission Electron Microscopy (TEM) analyses of sperm cells from *FSIP2* patient.

Longitudinal sections of sperm flagellum from control. In the midpiece (MP), outer dense fibers (ODF) are found longitudinally between the mitochondrial sheath (MS) and the axoneme **(Ax)**. The fibrous sheat (FS) is found in the principal piece (PP). **(B)** Longitudinal sections of sperm flagellum from control patient FSIP2_4. We can notice an abnormal flagellum with DFS which appears severly disorganized and extends up to the sperm neck. ODFs remains present whereas the mitochondrial sheath is totally absent from the midpiece likely due to an abnormal migration of the annulus . The central axoneme is not clearly visible. **(C)** Cross-sections of the mid-piece (MP) in control. The axoneme is surrounded by 9 outer dense fibers (ODFs) and the mitochondria sheath (MS). The axoneme is composed of nine peripheral microtubules doublets (DMTs) and a central pair of singlet microtubules (CP). **(D)** Cross-sections of the principal piece from fertile control. The axoneme is surrounded by 7 ODFs and by the fibrous sheath (FS) composed of two longitudinal columns (LC) connected by circumferential ribs (CR). **(E-H)** Cross-sections of sperm flagellum from patient FSIP2_4. The

fibrous sheath is dysplastic, thickened and dramatically disorganized with randomly oriented fibres **(E-G)**. When the circumferential ribs (CR) seem present, we observed only one nascent and abnormal longitudinal column (LC) adjacent to doublets 3 or 8 **(H)**. Various axonemal anomalies can be observed including the lack of the central pair **(E, F, H)** or complete axonemal desorganization **(G)**. Alterations of outer dense fibers is also seen with surpernumeray **(E)**, missing **(F)** or fully disorganized ODFs **(G)**. Dynein arms (DA) are rarely visible **(E, H).**

***FSIP2* mutations lead to the absence of AKAP4 protein and an altered distribution of the axonemal components.**

Despite repeated attempts, no antibodies specifically recognizing FSIP2 in human sperm cells could be obtained, precluding the immunodetection of FSIP2 protein on sperm samples from control individuals and its potential absence in FSIP_$_{1\text{-}4}$ patients. We therefore decided to analyze the distribution of various proteins of the sperm flagellum, and in particular of AKAP4, a protein described to localize to the FS and to directly interact with FSIP2 (Brown *et al.*, 2003). We performed some IF assays on sperm cells and observed that the AKAP4 staining was totally absent from patient FSIP2_$_3$ and FSIP2_$_4$ whereas AKAP4 immunostaining strongly decorated the principal piece of the sperm flagellum of control sperm cells from fertile individual (Fig. 4). In order to determine if the loss of AKAP4 was a constant feature of the MMAF phenotype, we analyzed MMAF patients carrying mutations in *CFAP43*, *CFAP44* and *DNAH1*. In these patients, AKAP4 immunostaining was comparable with what was observed on fertile controls (Fig. 4), indicating that the molecular structure of the FS was not as severely altered in the sperm from MMAF patients with mutations in axoneme-related genes. We decided to further investigate the flagellum structure of FSIP2 patients by performing immunostaining with antibodies targeting different axonemal proteins. We analyzed SPAG6, a component of the central pair complex. In fertile control, we observed that SPAG6 staining was concentrated in the principal piece of the spermatozoa whereas the marking was absent or strongly reduced in FSIP2-mutated patients with an abnormal and diffuse pattern concentrated on the flagellum's proximal part (Fig. S3 A). Moreover, staining of DNAH5 and DNALI1 , two proteins located

in the outer dynein arms (ODAs) and inner dynein arms (IDAs) respectively, were absent in FSIP2_3 and FSIP2_4 patients demonstrating that the whole axonemal structure was strongly affected by the presence of the FSIP2 mutations (Fig. S3 B,C).

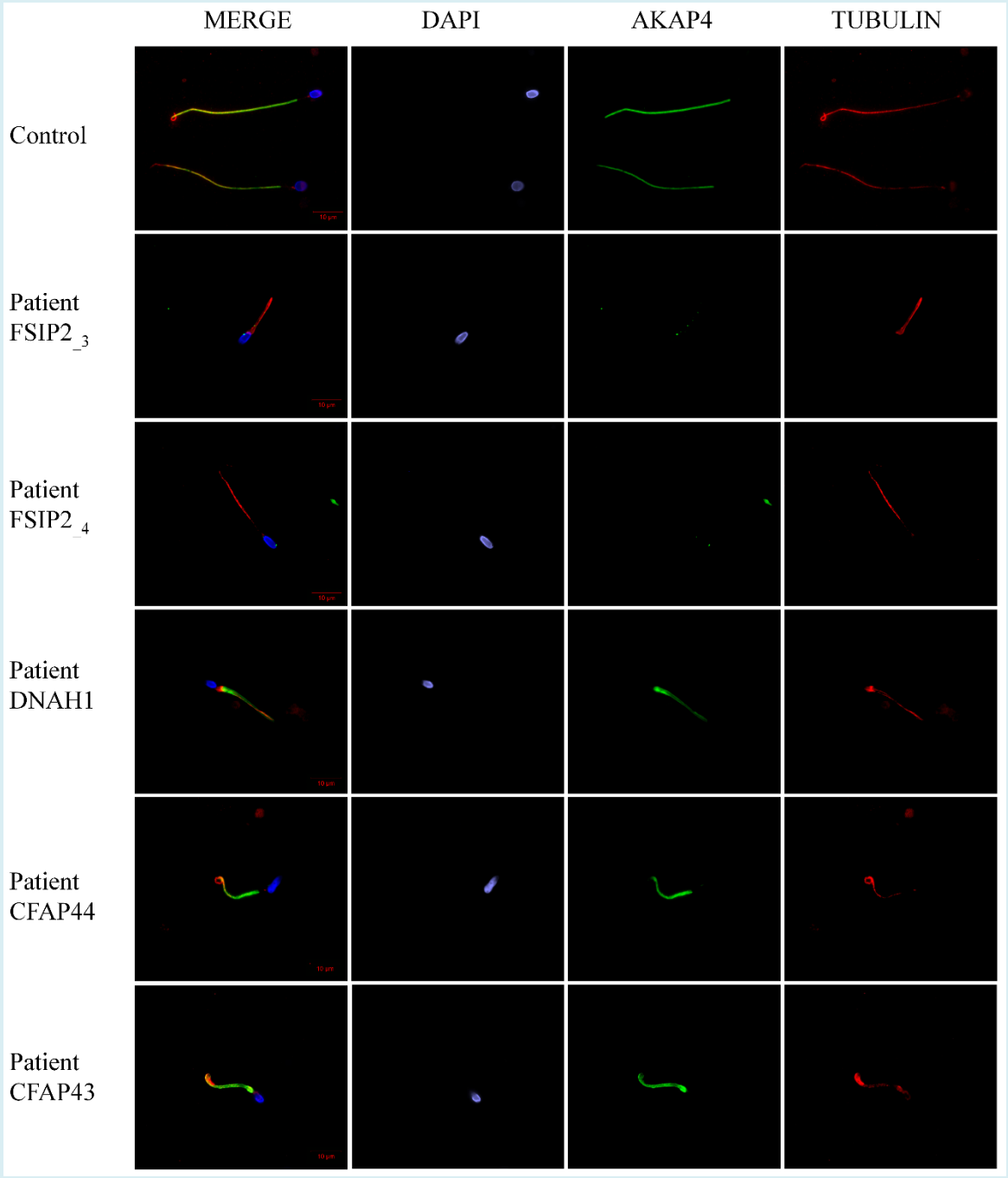


Figure 4. AKAP4 is absent from *FSIP2* patients' sperm cells.
Sperm cells from a fertile control and the two patients FSIP2_3 and FSIP2_4 were stained with anti-AKAP4 (green) and anti-acetylated tubulin (red) antibodies. DNA was counterstained with Hoechst 33342. The AKAP4 immunostaining is concentrated in the principal piece of the spermatozoa from the fertile control. AKAP4 staining is absent in sperm from the patients FSIP2_3 and FSIP2_4. Scale bar 10 µm.

Discussion

The study presented here permitted to demonstrate that the absence of a functional FSIP2, a protein of the flagella fibrous sheath, induced a MMAF phenotype. Logically, TEM analyses of sperm from our FSIP2 patients revealed a severe dysplasia of the FS which appeared extremely disorganized and hypertrophic. Such FS defects were previously reported and studied in patients with a MMAF-like phenotype initially described as DFS, for dysplasia of the fibrous sheath (Chemes *et al.*, 1987; Chemes and Rawe, 2003, 2010; Chemes and Alvarez Sedo, 2012). These FS defects do not seem to be as marked in sperm cells from other MMAF patients with mutations in the three axonemal genes *DNAH1*, *CFAP43* and *CFAP44* (Ben Khelifa *et al.*, 2014; Coutton *et al.*, 2015; Tang *et al.*, 2017). The FS is a cytoskeletal structure located in the principle piece of the sperm flagellum which seems to be unique to amniote sperm flagellum (Fawcett, 1970). This is consistent with our expression data showing a specific expression of *FSIP2* in the testis with no expression in the other ciliated tissues (Fig. S1). In the principal piece, the FS is formed by two longitudinal columns connected by circumferential ribs (Eddy *et al.*, 2003). The FS is believed to serve as a scaffold for numerous proteins including glycolytic enzymes or signaling protein and to play a role in the regulation of sperm motility but also in sperm maturation, capacitation and hyperactivation (Eddy *et al.*, 2003; Eddy, 2007). An important part of the enzymatic and energy metabolism of the principal piece is likely regulated by the cAMP-dependent protein kinase A (PKA) which is anchored in the fibrous sheath through different A-kinase anchoring protein (AKAPs) (Carr and Newell, 2007). AKAPs are a family composed of around 50 scaffold proteins which confine various enzymes including kinases, phosphatases and phosphodiesterases and their corresponding enzymatic activities to specific intracellular locations (Kumar *et al.*, 2017). In sperm, six AKAPs have been identified, three of which (AKAP3, AKAP4 and AKAP14) are located in the fibrous sheath (Fiedler *et al.*, 2008). FSIP2 was shown to bind to AKAP4, the most abundant AKAP of the FS which was

described to be important for completing fibrous sheath assembly (Brown *et al.*, 2003; Nipper *et al.*, 2006). Furthermore, the region where FSIP2 likely docks onto AKAP4 encompasses the RIa/RIIa-dual domain which is the docking domain of PKA onto the AKAPs' proteins, suggesting that FSIP2 could have a role in regulating PKA's anchoring to AKAP4 (Turner *et al.*, 1998; Brown *et al.*, 2003; Welch *et al.*, 2010). Here we showed that in FSIP2 deficient patients, AKAP4 was absent from the sperm flagellum contrary to what was observed in the other patients with a similar phenotype due to mutations in axonemal-associated MMAF genes (Fig. 4). Interestingly, it has been previously reported AKAP4 deficient mice also had shortened flagella with irregular diameter of the principal piece (Miki *et al.*, 2002) as observed in our FSIP2-mutated patients (Fig. 1). In addition, the only patient reported with a partial deletion of *AKAP4* (Baccetti *et al.*, 2005), also presented spermatozoa with extremely short tails and severe FS dysplasia.

In addition to the severe abnormalities of the peri-axonemal structures described before, sperm from FSIP2 patients also displayed some axonemal anomalies including the absence of the CP and of all inner/outer dynein arms (Fig. 3 and 4). Such axonemal defects were also described in sperm cells from patients described as DFS (Chemes and Rawe, 2003, 2010). Moreover, when discussing the phenotype of their *AKAP4* deleted patient, Baccetti et al. (2005) also reported that their patient's sperm had an altered axonemal structure with missing dynein arms and CP. They however did not present any sperm tail cross sections making reliable comparison with our patients difficult. Since the axoneme is assembled soon after the completion of meiosis and appears before the FS (Horowitz *et al.*, 2005), we can presume that these axonemal defects are secondary to FS misassembly or disorganization. This assumption is supported by studies on testicular biopsies from DFS patients showing that the phenotype

arises during late spermiogenesis (Ross *et al.*, 1973, 1973; Barthelemy *et al.*, 1990; Rawe *et al.*, 2002; Chemes and Rawe, 2010).

Absence of FSIP2 and consequently of AKAP4 could impede the correct localization of functionally important FS-associated proteins critical for the maintenance of sperm function and structure such as PKA or glycolytic enzymes (Miki *et al.*, 2002; de Boer *et al.*, 2014). Additionally, the important disorganization of the FS and the failure to form the longitudinal columns may also alter the direct interaction of DMTs 3 and 8 with the FS in the principal piece and weaken the stability of the axonemal structures leading to the total destabilization of the axoneme. We have just demonstrated that CFAP43 and CFAP44 likely link the axoneme with the extra-axonemal components (Coutton *et al.*, 2018). This suggests that most of the flagellar structures are tightly connected and defects in any of these structures may alter the whole process of flagellar genesis and produce a MMAF phenotype. Intriguingly, we observed that the sperm phenotype induced by the presence of *FSIP2* mutations is more severe than what was described for *Akap4* mutant mice (Miki *et al.*, 2002), where the FS only was described to be affected. This could indicate that FSIP2 is not only a structural protein of the FS but that it may be involved in the intraflagellar transport of several components, potentially including AKAP4. Further studies are now required to define the role of the different proteins found to be associated with a MMAF phenotype and to identify and characterize sub-phenotypes induced by the different molecular defects. Overall, this demonstrates the crucial importance of FSIP2 in maintaining the whole structure and function of the sperm flagellum.

WES analysis permitted to identify four FSIP2 loss of function variants. All were indels including nucleotide deletions, duplications or insertions. No obvious phenotype differences were observed among these four patients, concordant with the fact that all identified mutations

are expected induce the production of a truncated non-functional protein. Interestingly we noted that FSIP2 patients presented a significant lower rate of spermatozoa with absent flagella compared to patients with an identified mutation in other MMAF genes suggesting that *FSIP2* mutations might not alter the head-flagella junction as severely as mutations in axonemal-related genes. Moreover, although not significant, FSIP2 patients had a higher rate of flagella with an irregular diameter which is consistent with the overabundance and disorganization of the fibrous sheath (Chemes and Rawe, 2010). Several questions remain concerning the prognosis of ICSI using sperm cells from *FSIP2* patients. Successful application of ICSI in MMAF patients has been previously described, including patients carrying mutations in *DNAH1*, (Olmedo *et al.*, 2000; Chemes and Rawe, 2003; Chemes and Alvarez Sedo, 2012; Dávila Garza and Patrizio, 2013; Coutton *et al.*, 2015; Wambergue *et al.*, 2016). So far it however remains difficult to predict the prognosis of patients carrying FSIP mutations, as ICSI success rates has been described to be influenced by the type of ultrastructural flagellar defects (Mitchell *et al.*, 2006; Fauque *et al.*, 2009; Coutton *et al.*, 2015). Additional correlation studies should now be performed to take into account the individual genotypes in the counseling of MMAF patients.

The study of the cohort presented here permitted to identify 22 patients (28.2%) with mutations in *DNAH1*, *CFAP43* and *CFAP44* (Coutton *et al.,* 2018) and two patients with CFAP69 mutations (Dong et al. 2018). In the present study, we identified 4 additional patients with FSIP2 mutations and established FSIP2 as a new recurrent genetic cause of MMAF present in more than 5% of our cohort. The overall rate of a positive molecular diagnosis for this cohort is therefore 36% (28/78) which is close to the diagnosis rate found in other diseases with heterogeneous genetic etiologies (Farwell *et al.*, 2015; Fogel *et al.*, 2016). Our data thus demonstrate the clinical utility of WES for the identification of genetic causes responsible for

male infertility due to a MMAF phenotype. In-depth bioinformatic analyses of WES data from the remaining 50 undiagnosed individuals from this cohort, associated with the inclusion of additional MMAF patients, will likely improve the diagnosis yield and bring light to other rare genetic defects inducing a MMAF phenotype.

Acknowledgment

We thank all patients and control individuals for their participation.

Authors' roles

CC, PFR, CA, MB and AT analyzed the data and wrote the manuscript; TK, ZK, NTM, MB, AAY, PFR and CC performed and analyzed genetic data; GM, AS performed IF experiments. KPG, AB and GM performed EM experiments, sperm analyses, RZ, VS provided clinical samples and data; CC, PFR, CA and GM designed the study, supervised all molecular laboratory work, had full access to all of the data in the study and takes responsibility for the integrity of the data and its accuracy. All authors contributed to the report.

Funding

This work was supported by following grants: the 'MAS-Flagella' project financed by the French ANR and the DGOS for the program PRTS 2014 and the 'Whole genome sequencing of patients with Flagellar Growth Defects (FGD)' project financed by the Fondation Maladies Rares for the program Séquençage à haut débit 2012.

Conflict of interest

The authors have declared that no conflict of interest exists.

Tables

Table I Detailled semen parameters in the four MMAF patients harobring a *FSIP2* mutation.

***FSIP2* mutated patients**			**Semen parameters**									
Patient ID	Age (years)	*FSIP2* mutations	Sperm volume (ml)	Sperm concentration (10^6/ml)	Total motility (1h)	Vitality	Normal spermatozoa	Absent flagella	Short Flagella	Coiled Flagella	Bent Flagella	Flagella of irregular caliber
FSIP2$_{_1}$	44	c.[910delC]	2.5	6.7	1	47	0	2	32	26	-	57
FSIP2$_{_2}$	52	c.[2282dupA]	4.7	1.25	0	7	0	16	50	18	-	36
FSIP2$_{_3}$	45	c.[1606_1607insTGT; 1607_1616delAAAGATTGCA]	3.4	0.66	0	30	0	6	78	6	-	6
FSIP2$_{_4}$	43	c.[16389_16392delAATA]	1.7	49.5	5	52	0	6	70	14	-	80

Values are expressed in percent, unless specified otherwise.

Table II Average semen parameters in patients mutated in *FSIP2*, in axonemal-associated genes (*DNAH1*, *CFAP43* and *44*) and for the 78 included MMAF subjects in the study.

Semen parameters	MMAF with *FSIP2* mutations n=4	MMAF other mutations[$] n=22	Overall MMAF n=78	Reference limits[(%)]
Mean age (years)	46 ± 4 (n'=4)	39.8 ± 7 (n'=21)	41.6 ± 7.7 (n'=77)	NA
Sperm volume (ml)	3.1 ± 1.3 (n'=4)	3.4 ± 1.2 (n'=20)	3.5 ± 1.4 (n'=75)	1.5 (1.4-1.7)
Sperm concentration (10^6/ml)	14.5 ± 23.5 (n'=4)	20.1 ± 18.8 (n'=20)	25.6 ± 32.1 (n'=75)	15 (12-16)
Total motility (1 h)	1.5 ± 2.4 (n'=4)	0.7 ± 2.4 (n'=21)	3.9 ± 5.6 (n'=76)	40 (38-42)
Vitality	34 ± 20.3 (n'=4)	50.5 ± 22.7 (n'=19)	52.7 ± 20 (n'=72)	58 (55-63)
Normal spermatozoa	0 ± 0 (n'=4)	0.5 ± 2.3 (n'=20)	1.6 ± 2.7 (n'=61)	23 (20-26)
Absent flagella	7.5 ± 5.9* (n'=4)	28.1 ± 14.4 (n'=15)	20.7 ± 15.7 (n'=66)	5 (4-6)
Short Flagella	57.5 ± 20.6 (n'=4)	57.1 ± 27.9 (n'=19)	43.7 ± 27.3 (n'=72)	1 (0-2)
Coiled Flagella	16 ± 8.3 (n'=4)	10.4 ± 6.6 (n'=16)	12.8 ± 9.4 (n'=69)	17 (15-19)
Bent Flagella	NA	8.7 ± 5.8 (n'=6)	4.2 ± 8.4 (n'=26)	13 (11-15)
Flagella of irregular caliber	44.8 ± 31.5 (n'=4)	27.9 ± 19.4 (n'=15)	31.7 ± 25.1 (n'=67)	2 (1-3)

Values are expressed in percent, unless specified otherwise. Values are mean +/- SD; n= total number of patients in each group; n'= number of patients used to calculate the average based on available data; NA: not available.

[$]Other mutations correspond to patient mutated in *CFAP43*, *CFAP44* and *DNAH1* according to Coutton *et al.,* 2018.

[(%)] Reference limits (5th centiles and their 95% confidence intervals) according to the World Health Organization (WHO) standards (Cooper *et al.*, 2010) and the distribution range of morphologically normal spermatozoa observed in 926 fertile individuals (Auger *et al.*, 2016).

* Significant P <0.05 (t-test)

REFERENCES

1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;**491**:56–65.

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;**7**:248–249.

Amiri-Yekta A, Coutton C, Kherraf Z-E, Karaouzène T, Le Tanno P, Sanati MH, Sabbaghian M, Almadani N, Sadighi Gilani MA, Hosseini SH, *et al.* Whole-exome sequencing of familial cases of multiple morphological abnormalities of the sperm flagella (MMAF) reveals new DNAH1 mutations. *Hum Reprod Oxf Engl* 2016;**31**:2872–2880.

Auger J, Jouannet P, Eustache F. Another look at human sperm morphology. *Hum Reprod Oxf Engl* 2016;**31**:10–23.

Baccetti B, Collodel G, Estenoz M, Manca D, Moretti E, Piomboni P. Gene deletions in an infertile man with sperm fibrous sheath dysplasia. *Hum Reprod Oxf Engl* 2005;**20**:2790–2794.

Barthelemy C, Tharanne MJ, Lebos C, Lecomte P, Lansac J. Tail stump spermatozoa: morphogenesis of the defect. An ultrastructural study of sperm and testicular biopsy. *Andrologia* 1990;**22**:417–425.

Ben Khelifa M, Coutton C, Zouari R, Karaouzène T, Rendu J, Bidart M, Yassine S, Pierre V, Delaroche J, Hennebicq S, *et al.* 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* 2014;**94**:95–104.

Boer P de, Vries M de, Ramos L. A mutation study of sperm head shape and motility in the mouse: lessons for the clinic. *Andrology* 2014;

Brown PR, Miki K, Harper DB, Eddy EM. A-kinase anchoring protein 4 binding proteins in the fibrous sheath of the sperm flagellum. *Biol Reprod* 2003;**68**:2241–2248.

Carr DW, Newell AEH. The role of A-kinase anchoring proteins (AKaps) in regulating sperm function. *Soc Reprod Fertil Suppl* 2007;**63**:135–141.

Chemes EH, Rawe YV. Sperm pathology: a step beyond descriptive morphology. Origin, characterization and fertility potential of abnormal sperm phenotypes in infertile men. *Hum Reprod Update* 2003;**9**:405–428.

Chemes HE, Alvarez Sedo C. Tales of the tail and sperm head aches: changing concepts on the prognostic significance of sperm pathologies affecting the head, neck and tail. *Asian J Androl* 2012;**14**:14–23.

Chemes HE, Brugo S, Zanchetti F, Carrere C, Lavieri JC. Dysplasia of the fibrous sheath: an ultrastructural defect of human spermatozoa associated with sperm immotility and primary sterility. *Fertil Steril* 1987;**48**:664–669.

Chemes HE, Olmedo SB, Carrere C, Oses R, Carizza C, Leisner M, Blaquier J. Ultrastructural pathology of the sperm flagellum: association between flagellar pathology and fertility prognosis in severely asthenozoospermic men. *Hum Reprod Oxf Engl* 1998;**13**:2521–2526.

Chemes HE, Rawe VY. The making of abnormal spermatozoa: cellular and molecular mechanisms underlying pathological spermiogenesis. *Cell Tissue Res* 2010;**341**:349–357.

Cooper TG, Noonan E, Eckardstein S von, Auger J, Baker HWG, Behre HM, Haugen TB, Kruger T, Wang C, Mbizvo MT, *et al.* World Health Organization reference values for human semen characteristics. *Hum Reprod Update* 2010;**16**:231–245.

Coutton C, Escoffier J, Martinez G, Arnoult C, Ray PF. Teratozoospermia: spotlight on the main genetic actors in the human. *Hum Reprod Update* 2015;**21**:455–485.

Coutton C, Vargas AS, Amiri-Yekta A, Kherraf Z-E, Fourati Ben Mustapha, Le Tanno P, Wambergue C, Karaouzene T, Martinez G, Crouzy S, *et al.* Mutations in CFAP43 and CFAP44 cause male infertility and flagellum defects in Trypanosoma and human. *Nat Commun* 2018; 2018 15;9(1):686. .

Dávila Garza SA, Patrizio P. Reproductive outcomes in patients with male infertility because of Klinefelter's syndrome, Kartagener's syndrome, round-head sperm, dysplasia fibrous sheath, and "stump" tail sperm: an updated literature review. *Curr Opin Obstet Gynecol* 2013;**25**:229–246.

Dong FN, Amiri-Yekta A, Martinez G, Saut A, Tek J, Stouvenel L, Lorès P, Karaouzène T, Thierry-Mieg N, Satre V, Brouillet S, Daneshipour A, Hosseini SH, Bonhivers M, Gourabi H, Dulioust E, Arnoult C, Touré A, Ray PF, Zhao H, Coutton C. (2018) Absence of CFAP69 Causes Male Infertility due to Multiple Morphological Abnormalities of the Flagella in Human and Mouse. Am J Hum Genet. **5**;102(4):636-648.

Eddy EM. The scaffold role of the fibrous sheath. *Soc Reprod Fertil Suppl* 2007;**65**:45–62.

Eddy EM, Toshimori K, O'Brien DA. Fibrous sheath of mammalian spermatozoa. *Microsc Res Tech* 2003;**61**:103–115.

Escalier D, David G. Pathology of the cytoskeleton of the human sperm flagellum: axonemal and peri-axonemal anomalies. *Biol Cell Auspices Eur Cell Biol Organ* 1984;**50**:37–52.

Farwell KD, Shahmirzadi L, El-Khechen D, Powis Z, Chao EC, Tippin Davis B, Baxter RM, Zeng W, Mroske C, Parra MC, *et al.* Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. *Genet Med Off J Am Coll Med Genet* 2015;**17**:578–586.

Fauque P, Albert M, Serres C, Viallon V, Davy C, Epelboin S, Chalas C, Jouannet P, Patrat C. From ultrastructural flagellar sperm defects to the health of babies conceived by ICSI. *Reprod Biomed Online* 2009;**19**:326–336.

Fawcett DW. A comparative view of sperm ultrastructure. *Biol Reprod Suppl* 1970;**2**:90–127.

Fiedler SE, Bajpai M, Carr DW. Identification and characterization of RHOA-interacting proteins in bovine spermatozoa. *Biol Reprod* 2008;**78**:184–192.

Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang H-Y, Dosztányi Z, El-Gebali S, Fraser M, *et al.* InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res* 2017;**45**:D190–D199.

Fogel BL, Lee H, Strom SP, Deignan JL, Nelson SF. Clinical exome sequencing in neurogenetic and neuropsychiatric disorders. *Ann N Y Acad Sci* 2016;**1366**:49–60.

Horowitz E, Zhang Z, Jones BH, Moss SB, Ho C, Wood JR, Wang X, Sammel MD, Strauss JF. Patterns of expression of sperm flagellar genes: early expression of genes encoding axonemal proteins during the spermatogenic cycle and shared features of promoters of genes encoding central apparatus proteins. *Mol Hum Reprod* 2005;**11**:307–317.

Inaba K. Molecular architecture of the sperm flagella: molecules for motility and signaling. *Zoolog Sci* 2003;**20**:1043–1056.

Inaba K. Molecular basis of sperm flagellar axonemes: structural and evolutionary aspects. *Ann N Y Acad Sci* 2007;**1101**:506–526.

Inaba K. Sperm flagella: comparative and phylogenetic perspectives of protein components. *Mol Hum Reprod* 2011;**17**:524–538.

Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009;**4**:1073–1081.

Kumar V, Jagadish N, Suri A. Role of A-Kinase anchor protein (AKAP4) in growth and survival of ovarian cancer cells. *Oncotarget* 2017;**8**:53124–53136.

Lehti MS, Sironen A. Formation and function of sperm tail structures in association with sperm motility defects. *Biol Reprod* 2017;**97**:522–536.

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;**536**:285–291.

Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods San Diego Calif* 2001;**25**:402–408.

Magnoni R, Palmfeldt J, Hansen J, Christensen JH, Corydon TJ, Bross P. The Hsp60 folding machinery is crucial for manganese superoxide dismutase folding and function. *Free Radic Res* 2014;**48**:168–179.

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F. The Ensembl Variant Effect Predictor. *Genome Biol* 2016;**17**:122.

Miki K, Willis WD, Brown PR, Goulding EH, Fulcher KD, Eddy EM. Targeted disruption of the Akap4 gene causes defects in sperm flagellum and motility. *Dev Biol* 2002;**248**:331–342.

Mitchell V, Rives N, Albert M, Peers M-C, Selva J, Clavier B, Escudier E, Escalier D. Outcome of ICSI with ejaculated spermatozoa in a series of men with distinct ultrastructural flagellar abnormalities. *Hum Reprod Oxf Engl* 2006;**21**:2065–2074.

Nipper RW, Jones BH, Gerton GL, Moss SB. Protein domains govern the intracellular distribution of mouse sperm AKAP4. *Biol Reprod* 2006;**75**:189–196.

Olmedo SB, Rawe VY, Nodar FN, Galaverna GD, Acosta AA, Chemes HE. Pregnancies established through intracytoplasmic sperm injection (ICSI) using spermatozoa with dysplasia of fibrous sheath. *Asian J Androl* 2000;**2**:125–130.

Rawe VY, Terada Y, Nakamura S, Chillik CF, Olmedo SB, Chemes HE. A pathology of the sperm centriole responsible for defective sperm aster formation, syngamy and cleavage. *Hum Reprod Oxf Engl* 2002;**17**:2344–2349.

Ray PF, Toure A, Metzler-Guillemain C, Mitchell MJ, Arnoult C, Coutton C. Genetic abnormalities leading to qualitative defects of sperm morphology or function. *Clin Genet* 2017;**91**:217–232.

Ross A, Christie S, Edmond P. Ultrastructural tail defects in the spermatozoa from two men attending a subfertility clinic. *J Reprod Fertil* 1973;**32**:243–251.

Satir P, Christensen ST. Structure and function of mammalian cilia. *Histochem Cell Biol* 2008;**129**:687–693.

SEQC/MAQC-III Consortium. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat Biotechnol* 2014;**32**:903–914.

Sha Y, Yang X, Mei L, Ji Z, Wang X, Ding L, Li P, Yang S. DNAH1 gene mutations and their potential association with dysplasia of the sperm fibrous sheath and infertility in the Han Chinese population. *Fertil Steril* 2017;**107**:1312–1318.e2.

Tang S, Wang X, Li W, Yang X, Li Z, Liu W, Li C, Zhu Z, Wang L, Wang J, *et al.* Biallelic Mutations in CFAP43 and CFAP44 Cause Male Infertility with Multiple Morphological Abnormalities of the Sperm Flagella. *Am J Hum Genet* 2017;**100**:854–864.

Turner RM, Johnson LR, Haig-Ladewig L, Gerton GL, Moss SB. An X-linked gene encodes a major human sperm fibrous sheath protein, hAKAP82. Genomic organization, protein kinase A-RII binding, and distribution of the precursor in the sperm tail. *J Biol Chem* 1998;**273**:32135–32141.

Turner RM, Musse MP, Mandal A, Klotz K, Jayes FC, Herr JC, Gerton GL, Moss SB, Chemes HE. Molecular genetic analysis of two human sperm fibrous sheath proteins, AKAP4 and AKAP3, in men with dysplasia of the fibrous sheath. *J Androl* 2001;**22**:302–315.

Wambergue C, Zouari R, Fourati Ben Mustapha S, Martinez G, Devillard F, Hennebicq S, Satre V, Brouillet S, Halouani L, Marrakchi O, *et al.* Patients with multiple morphological abnormalities of the sperm flagella due to DNAH1 mutations have a good prognosis following intracytoplasmic sperm injection. *Hum Reprod Oxf Engl* 2016;**31**:1164–1172.

Wang Y, Yang J, Jia Y, Xiong C, Meng T, Guan H, Xia W, Ding M, Yuchi M. Variability in the morphologic assessment of human sperm: use of the strict criteria recommended by the World Health Organization in 2010. *Fertil Steril* 2014;

Welch EJ, Jones BW, Scott JD. Networking with AKAPs: context-dependent regulation of anchored enzymes. *Mol Interv* 2010;**10**:86–97.

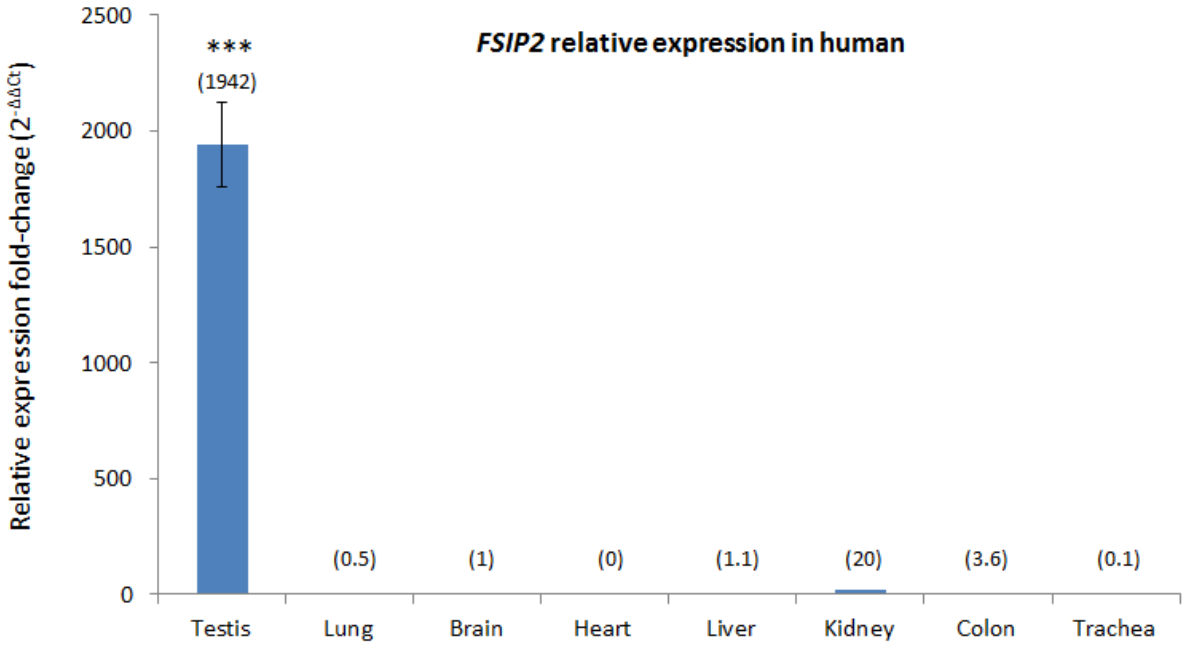
Supplementary data

Supplementary figures

Figure S1. Relative mRNA Expression of human and mouse *FSIP2* transcripts.

FSIP2 mRNA levels in a panel of human **(A)** and mouse **(B)** normal tissues. Results are presented as the mean of triplicates (ratio target gene/ACTB) ± Standard Deviation (SD). RT-qPCR data were normalized using the reference gene ACTB with the -ΔΔCt method. Brain expression is arbitrary set to 1. Relative expression fold-change are indicated into brackets. Similar expression pattern was shown for human and mouse. In human and mouse, *FSIP2* has the strongest expression in testis compared to other organs. Unpaired t-test, ***P< 0.001.

A


FSIP2 relative expression in human
Relative expression fold-change (2-ΔΔCt)
2500
2000
1500
1000
500
0
***
(1942)
(0.5)
(1)
(0)
(1.1)
(20)
(3.6)
(0.1)
Testis
Lung
Brain
Heart
Liver
Kidney
Colon
Trachea

B

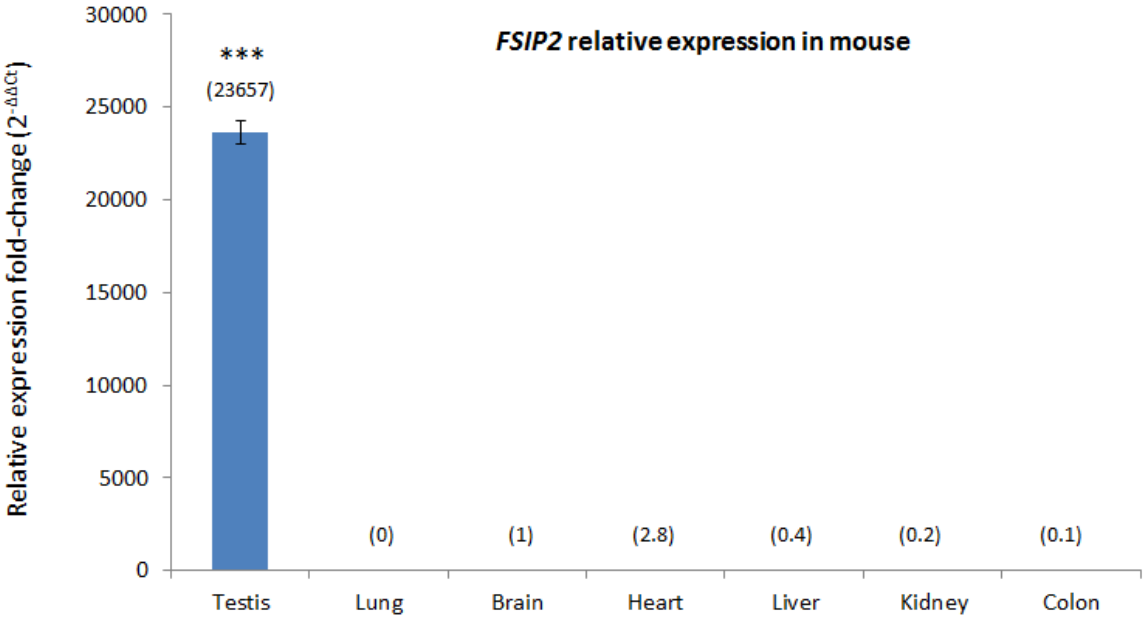

FSIP2 relative expression in mouse
Relative expression fold-change (2-ΔΔCt)
30000
25000
20000
15000
10000
5000
0
***
(23657)
(0)
(1)
(2.8)
(0.4)
(0.2)
(0.1)
Testis
Lung
Brain
Heart
Liver
Kidney
Colon

Figure S2. HSP60 immunostaining in human spermatozoa from controls and FSIP2 patient. Sperm cell from a fertile control and the patient FSIP2_4 were stained with anti-HSP60 (red), which detects a mitochondrial protein located in the midpiece, and anti-acetylated tubulin (green) antibodies. DNA was counterstained with Hoechst 33342. The HSP60 immunostaining is concentrated in the midpiece of the spermatozoa from the fertile control whereas it is found with an abnormal and diffuse pattern all along of the sperm flagellum in FSIP2_4 patient. Scale bar 10 µm.

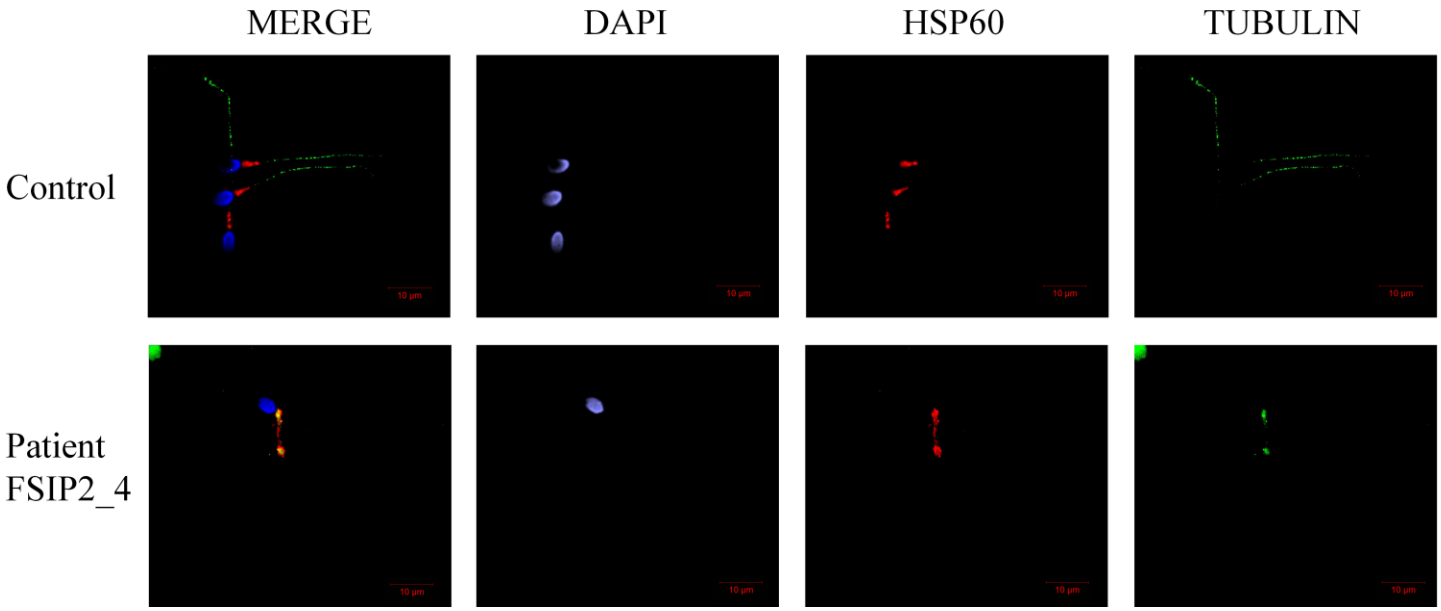


Figure S3. SPAG6 immunostainings of sperm cells from *FSIP2* patients. Sperm cells from a fertile control and the two patients FSIP2_3 and FSIP2_4 were stained with anti-SPAG6 (green), a protein located in the CP complex, and anti-acetylated tubulin (red) antibodies. DNA was counterstained with Hoechst 33342. The SPAG6 immunostaining is concentrated in the principal piece of the spermatozoa from the fertile control. SPAG6 staining is absent or strongly reduced with an abnormal and diffuse pattern concentrated in the midpiece in sperm from the patients FSIP2_3 and FSIP2_4. Scale bar 10 µm

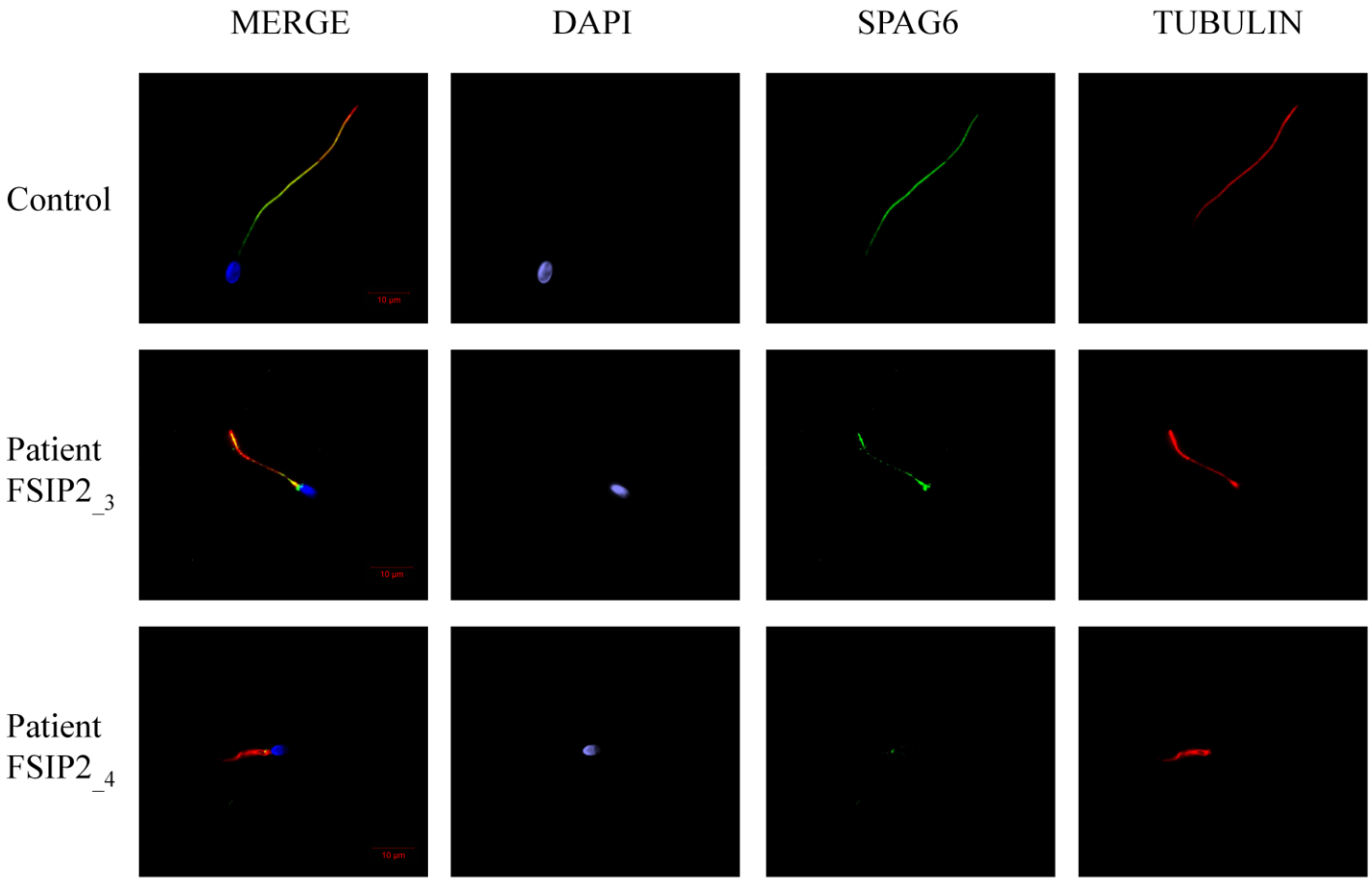
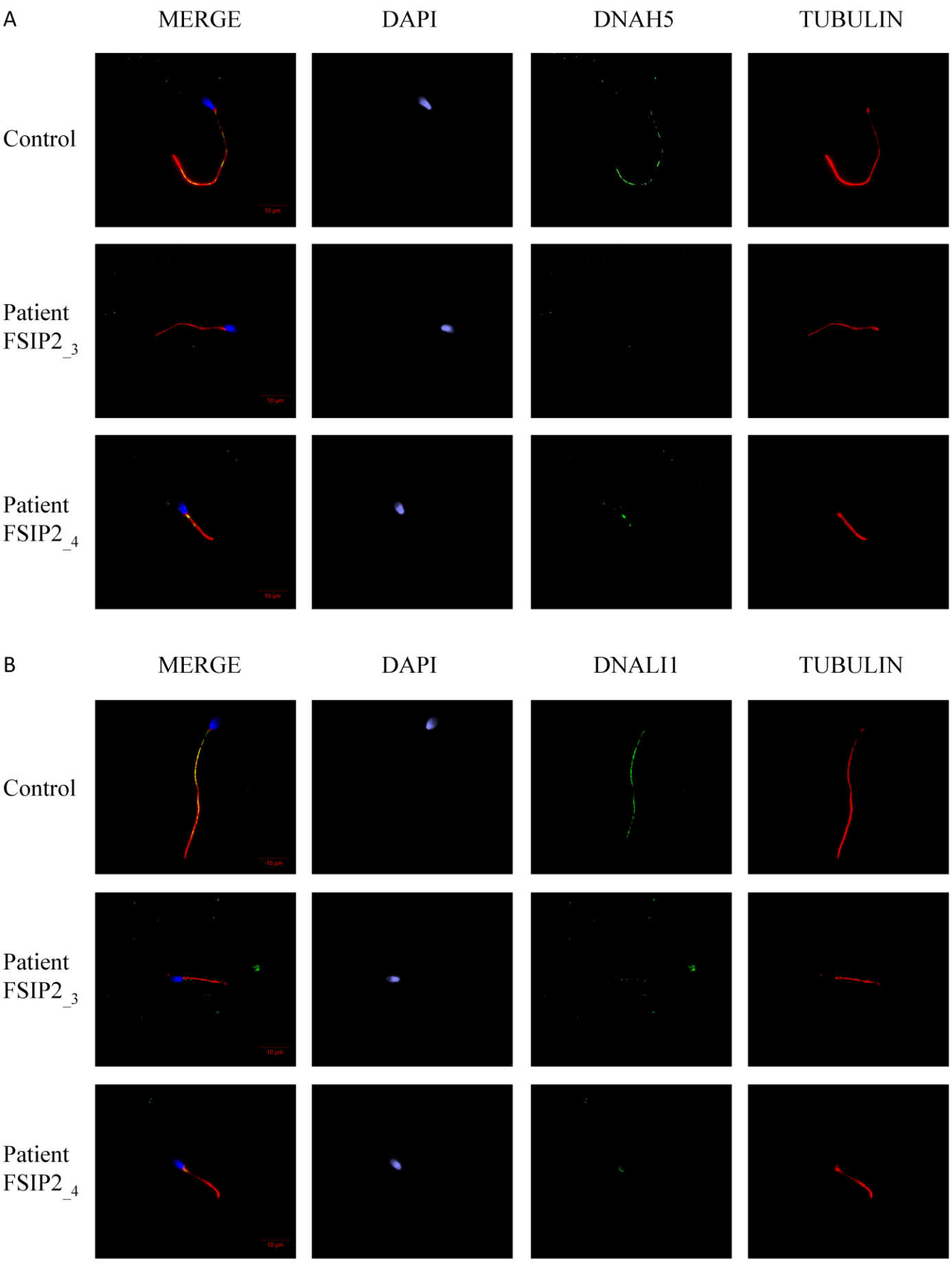


Figure S4. DNAH5 and DNAL1 immunostainings of sperm cells from *FSIP2* patients. (**A**) Sperm cells from a fertile control and the two patients FSIP2_3 and FSIP2_4 were stained with anti-DNAH5 (green), a protein located in the outer dynein arm, and anti-acetylated tubulin (red) antibodies. DNA was counterstained with Hoechst 33342. The DNAH5 immunostaining is totally absent in sperm from the patients FSIP2_3 and FSIP2_4 compared to the fertile control. Scale bar 10 µm. (**B**) Sperm cells from a fertile control and the two patients FSIP2_3 and FSIP2_4 were stained with anti-DNALI1 (green), a protein located in the inner dynein arm and anti-acetylated tubulin (red) antibodies. DNA was counterstained with Hoechst 33342. Contrary to the control, the DNALI1 immunostaining is not detectable in the principle piece in sperm cells from both FSIP2 patients. Scale bar 10 µm.


A
MERGE
DAPI
DNAH5
TUBULIN
Control
Patient FSIP2_3
Patient FSIP2_4
B
MERGE
DAPI
DNALI1
TUBULIN
Control
Patient FSIP2_3
Patient FSIP2_4

Supplementary tables

Table S1 Primer sequences used for Sanger sequencing verification of *FSIP2* variations and respective melting temperatures (Tm).

Primer names	Primer sequences (5'-3')	Tm
FSIP2-Ex8F	TGGAAATTTGCCCCTTACTAC	58°C
FSIP2-Ex8R	AATGTCTGGCACAAAGGGAG	
FSIP2-Ex16F	GGGGAAAAAGAAACTTCTGAGAG	58°C
FSIP2-Ex16R	TGTTTTCCCCACAACTGTTT	
FSIP2-Ex17F	TGCTGCTCTAACCCAGAAGG	60°C
FSIP2-Ex17R	CACAGTGGTAGTTGAAGTAACTTTTGA	

Table S2 Primers used for RT-qPCR of *FSIP2* in human and mouse.

Primer name	Primer sequence (5'-3')	Tm
FSIP2h_qPCR_Ex5F	TGATGTTGCACAAGTCCAAAACT	60°C
FSIP2h_qPCR_Ex6R	TCCATTAGGAATAGGCGTTGTTC	
Fsip2m_qPCR_Ex20F	CCAGATGCAAAGGCTAATGC	60°C
Fsip2m_qPCR_Ex22R	TTGAAAGACGTGCACTACCG	
ACTB-h_qPCR_F	CCAACCGCGAGAAGATGA	60°C
ACTB-h_qPCR_R	CCAGAGGCGTACAGGGATAG	
Actb-m_qPCR_F	ACCAGAGGCATACAGGGACA	60°C
Actb-m_qPCR_R	CTAAGGCCAACCGTGAAAAG	

Table S3 Coordinates of all FSIP2 variants.

Genomic Coordinates are based on the UCSC GRCh37/hg19 assembly and intragenic position are given for transcript ENST00000424728. All variants are homozygous.

Patients	Origin	Variant coordinates*	cDNA Variations	Amino acid variations
FSIP2_1	Tunisian	chr2:186618487	c.[910delC]	p.Gln304LysfsTer13
FSIP2_2	Tunisian	chr2:186654145	c.[2282dupA]	p.Asn761LysfsTer4
FSIP2_3	Tunisian	chr2:186653470 _186653479	c.[1606_1607insTGT; 1607_1616delAAAGATTGCA]	p.Lys536MetfsTer1
FSIP2_4	Tunisian	chr2:186670422 _186670425	c.[16389_16392delAATA]	p.Glu5463GlufsTer7

4- Article 6 : Whole-exome sequencing of familial cases of multiple morphological abnormalities of the sperm flagella (MMAF) reveals new *DNAH1* mutations

Amiri-Yekta A, Coutton C, **Kherraf ZE**, Karaouzène T, Le Tanno P, Sanati MH, Sabbaghian M, Almadani N, Sadighi Gilani MA, Seyedeh Hanieh Hosseini, Bahrami S, Daneshipour A, Bini M, Arnoult C, Colombo R, Gourabi H & Ray PF.

Human Reproduction, Octobre 2016.

Contexte et principaux résultats de l'étude :

Dans l'objectif de l'exploration génétique du phénotype MMAF, notre équipe a réalisé une première étude par cartographie d'homozygotie appliquée à une cohorte de 20 sujets infertiles présentant un phénotype MMAF isolé. A l'issue de cette étude, des mutations homozygotes délétères ont été identifiées dans le gène *DNAH1*, qui code pour une chaine lourde de dynéine axonémale. Ces mutations homozygotes ont été identifiées chez environ 30% des sujets testés (Ben Khelifa et al., 2014). Avec l'avènement du séquençage exomique, notre équipe a poursuivi l'exploration génétique du phénotype MMAF.

Dans cette nouvelle étude, le séquençage exomique a été appliqué chez 8 familles (12 sujets) d'origines iranienne ou italienne dans l'objectif d'identifier des mutations dans le gène *DNAH1*. Cette analyse nous a permis d'identifier des mutations homozygotes délétères chez deux familles parmi les 8 recrutées. Au total, 5 sujets étaient mutés dont deux cas index : un sujet iranien avec une mutation homozygote d'épissage, c.8626-1G>A et un sujet italien avec une mutation homozygote faux-sens, c.3860T>G (p.Val1287Gly) prédite comme délétère. L'analyse du sperme d'un sujet iranien porteur de la mutation homozygote d'épissage, c.8626-1G>A nous a permis de montrer que l'ARNm de *DNAH1* était dégradé par le système NMD (nonsense-mediated mRNA decay). L'analyse en immunofluorescence et microscopie confocale de ce sperme a confirmé l'absence de la protéine codée par ce gène. Afin d'estimer la fréquence du variant c.8626-1G>A chez les patients MMAF d'origine iranienne, nous avons séquencé l'exon en question chez une cohorte de 38 sujets de cette région et avons identifié un nouveau sujet homozygote pour cette mutation.

Dans leur ensemble, ces résultats montrent la forte implication des mutations du gène *DNAH1* dans le phénotype MMAF et suggèrent que ce dernier est génétiquement hétérogène et que *DNAH1* n'est probablement pas le seul gène impliqué dans sa pathogénie.

Human Reproduction, Vol.31, No.12 pp. 2872–2880, 2016
Advanced Access publication on October 26, 2016 doi:10.1093/humrep/dew262

human reproduction

ORIGINAL ARTICLE ***Reproductive genetics***

Whole-exome sequencing of familial cases of multiple morphological abnormalities of the sperm flagella (MMAF) reveals new *DNAH1* mutations

Amir Amiri-Yekta[1,2,†], Charles Coutton[2,3,†], Zine-Eddine Kherraf[2,4], Thomas Karaouzène[2,4], Pauline Le Tanno[2,3], Mohammad Hossein Sanati[1,5], Marjan Sabbaghian[6], Navid Almadani[1], Mohammad Ali Sadighi Gilani[6], Seyedeh Hanieh Hosseini[7], Salahadin Bahrami[1], Abbas Daneshipour[1], Maurizio Bini[7], Christophe Arnoult[2], Roberto Colombo[8,9], Hamid Gourabi[1,†#], and Pierre F. Ray[2,4,*#]

[1]Department of Genetics, Reproductive Biomedicine Research Center, Royan Institute for Reproductive Biomedicine, ACECR, PO Box 16635-148, Tehran, Iran [2]Genetic Epigenetic and Therapies of Infertility, Institute for Advanced Biosciences, INSERM 1209, CNRS UMR 5309, Université Grenoble Alpes, Grenoble F38000, France [3]CHU de Grenoble, UF de Génétique Chromosomique, Grenoble F-38000, France [4]CHU de Grenoble, UF de Biochimie Génétique et Moléculaire, Grenoble F-38000, France [5]Department of Medical Genetics, Institute of Medical Biotechnology, National Institute of Genetic Engineering and Biotechnology, Tehran, Iran [6]Department of Andrology, Reproductive Biomedicine Research Center, Royan Institute for Reproductive Biomedicine, ACECR, PO Box 16635-148, Tehran, Iran [7]Center for the Study and Treatment of Fertility Disorders, Niguarda Ca' Granda Metropolitan Hospital, Milan, Italy [8]Faculty of Medicine, Institute of Clinical Biochemistry, Catholic University, Rome, Italy [9]Center for the Study of Rare Hereditary Diseases, Niguarda Ca' Granda Metropolitan Hospital, Milan, Italy

*Correspondence address. UF de Biochimie et Génétique Moléculaire, CHU de Grenoble, 38043 Grenoble cedex 9, France. Tel: +33-4-76-76-55-73; E-mail: pray@chu-grenoble.fr

Submitted on July 15, 2016; resubmitted on September 9, 2016; accepted on September 15, 2016

STUDY QUESTION: Can whole-exome sequencing (WES) of patients with multiple morphological abnormalities of the sperm flagella (MMAF) identify causal mutations in new genes or mutations in the previously identified dynein axonemal heavy chain 1 (*DNAH1*) gene?

SUMMARY ANSWER: WES for six families with men affected by MMAF syndrome allowed the identification of *DNAH1* mutations in four affected men distributed in two out of the six families but no new candidate genes were identified.

WHAT IS KNOWN ALREADY: Mutations in *DNAH1*, an axonemal inner dynein arm heavy chain gene, have been shown to be responsible for male infertility due to a characteristic form of asthenozoospermia called MMAF, defined by the presence in the ejaculate of spermatozoa with a mosaic of flagellar abnormalities including absent, coiled, bent, angulated, irregular and short flagella.

STUDY DESIGN, SIZE, DURATION: This was a retrospective genetics study of patients presenting a MMAF phenotype. Patients were recruited in Iran and Italy between 2008 and 2015.

PARTICIPANTS/MATERIALS, SETTING, METHODS: WES was performed for a total of 10 subjects. All identified variants were confirmed by Sanger sequencing. Two additional affected family members were analyzed by direct Sanger sequencing. To establish the

[†]These authors are equally contributing first authors.
[#]These authors have co-leadership.

prevalence of the *DNAH1* mutation identified in an Iranian family, we carried out targeted sequencing on 38 additional MMAF patients of the same geographical origin. RT-PCR and immunochemistry were performed on sperm samples to assess the effect of the identified mutation on RNA and protein.

MAIN RESULTS AND THE ROLE OF CHANCE: WES in six families identified a causal mutations in two families. Two additional affected family members were confirmed to hold the same homozygous mutation as their sibling. In total, *DNAH1* mutations were identified in 5 out of 12 analyzed subjects (41.7%). If we only include index cases, we detected two mutated subjects out of six (33%) tested MMAF individuals. Furthermore we sequenced one *DNAH1* exon found to be mutated (c.8626-1G > A) in an Iranian family in an additional 38 MMAF patients from Iran. One of these patients carried the variant confirming that this variant is relatively frequent in the Iranian population. The effect of the c.8626-1G > A variant was confirmed by RT-PCR and immunochemistry as no RNA or protein could be observed in sperm from the affected men.

LARGE SCALE DATA: N/A.

LIMITATIONS, REASONS FOR CAUTION: WES allows the amplification of 80–90% of all coding exons. It is possible that some *DNAH1* exons may not have been sequenced and that we may have missed some additional mutations. Also, WES cannot identify deep intronic mutations and it is not efficient for detection of large genomic events (deletions, insertions, inversions). We did not identify any causal mutations in *DNAH1* or in other candidate genes in four out of the six tested families. This indicates that the technique and/or the analysis of our data can be improved to increase the diagnosis efficiency.

WIDER IMPLICATIONS OF THE FINDINGS: Our findings confirm that *DNAH1* is one of the main genes involved in MMAF syndrome. It is a large gene with 78 exons making it challenging and expensive to sequence using the traditional Sanger sequencing methods. We show that WES sequencing is good alternative to Sanger sequencing to reach a genetic diagnosis in patients with severe male infertility phenotypes.

STUDY FUNDING/COMPETING INTEREST(S): This work was supported by following grants: the 'MAS-Flagella' project financed by the French ANR and the DGOS for the program PRTS 2014 and the 'Whole genome sequencing of patients with Flagellar Growth Defects (FGD)' project financed by the Fondation Maladies Rares for the program Séquençage à haut débit 2012. The authors have no conflict of interest.

Key words: male infertility / genetic diagnosis / gene mutations / exome sequencing / teratozoospermia / flagellum / MMAF / DNAH1

Introduction

Men with primary ciliary dyskinesia (PCD), a pathology grouping different phenotypic entities due to molecular defects altering cilia function and affecting mainly the pulmonary function, are often infertile with asthenozoospermia and abnormal flagellar morphology (Storm van's and Omran, 2005). Some patients however present with similar sperm abnormalities with no other associated syndromes, suggesting analogous molecular mechanisms affecting only the sperm flagellum. As this phenotype is restricted to fertility, it could be considered as a new clinical entity and such sperm defects have indeed been described on many occasions since 1984 (Escalier and David, 1984). Patients consistently have had astheno-teratozoospermia characterized by a mosaic of flagellar abnormalities including absent, coiled, bent, angulated, irregular or short flagella mainly due to numerous ultrastructural defects of the axoneme. This phenotype has been reported as dysplasia of the fibrous sheath, short/stump tails or non-specific flagellar anomalies (Chemes *et al.*, 1987; Stalf *et al.*, 1995; Chemes and Rawe, 2003). Until recently, genetic causes of flagellar abnormalities have remained largely unexplained. In 2005, deletions in *AKAP3* and *AKAP4*, two genes encoding proteins of the fibrous sheath, were reported in one patient (Baccetti *et al.*, 2005); these findings however remain to be confirmed. In 2014, our team carried out a large genetics study and we proposed to use the term multiple morphological abnormalities of the sperm flagella (MMAF) which seemed to clearly define the phenotype of the included patients (Ben Khelifa *et al.*, 2014). In this study, homozygosity mapping using a single nucleotide polymorphism (SNP) array allowed the identification of pathogenic mutations in the dynein axonemal heavy chain 1 (*DNAH1*) gene in 28% of the analyzed patients. The identification of *DNAH1* mutations indicated that *DNAH1* is a major gene involved in the MMAF phenotype and is expected to account for up to one-third of cases (Ben Khelifa *et al.*, 2014). This indicates that MMAF is genetically heterogeneous and that other genes are likely involved in this syndrome (Coutton *et al.*, 2015).

It is currently estimated that 1500–2000 genes are involved in the control of spermatogenesis, including 300–600 specifically expressed in male germ cells (Matzuk and Lamb, 2008). The abundance of potential candidate genes makes the identification of pathogenic mutations difficult and complex. Gene identification is however the key to improving our understanding of the pathophysiology of infertility and could open new perspectives for the diagnosis and treatment of infertile patients. In recent years, different promising genomic approaches have catalyzed the identification of new genes involved in male infertility (El *et al.*, 2012). Whole-exome sequencing (WES), the sequencing of the coding sequence located in the exons of the translated genes, now appears as the best strategy to detect disease-causing variations in individuals affected with Mendelian disorders (Bamshad *et al.*, 2011). Analysis of WES data however remains challenging as 20 000–30 000 variants differing from the genomic reference sequence are usually found in any given individual (Gilissen *et al.*, 2012). It is then extremely arduous to identify the causal variant(s) from this large number of variants of usually unknown significance. The number of potentially pathogenic variants can be reduced by analyzing cohorts of affected individuals and looking for variants or defects in the same gene present in several affected

individuals. Alternatively, consanguineous kindreds can be analyzed and this strategy can be particularly successful for genetically heterogeneous pathology such as infertility (Boycott et al., 2013). Indeed, WES of small families has been successfully used for the detection of causal genes in phenotypes such as non-obstructive azoospermia (Ayhan et al., 2014) or sperm fertilization defects (Escoffier et al., 2015). In this study, we wanted to evaluate the efficiency of family-based WES to identify new *DNAH1* mutations or new genetic causes of MMAF syndrome.

Materials and Methods

Patient and control individuals

Five Iranian families (1–5) and one Italian family (6) were included in this study (Fig. 1). Among these 6 families, we included 12 subjects (P1–P12) presenting with asthenozoospermia due to a combination of 5 morphological defects of the sperm flagella including: absent, short, bent, coiled flagella and of irregular width (Table I) without any of the additional symptoms associated with PCD. About 10 individuals originated from Middle East (Iranians) and were treated in Tehran at the Royan Institute, (Reproductive Biomedicine Research Center) for primary infertility from 2008 to 2015. Two brothers of European origin (Southern Italia) consulted for primary infertility at the Center for the Study of Rare Inherited Diseases in Milan, Italia. About 10 of the 12 subjects were born from related parents, usually first cousins. WES was performed for a total of 10 patients and *DNAH1* targeted Sanger sequencing was performed for two additional brothers (P7 and P12) who had not been included in time for WES (Fig. 1). Except for Families 5 and 6, at least two infertile brothers from each family were included and were analyzed by WES (Fig. 1). All subjects had normal somatic karyotypes.

Sperm analysis was carried out in the source laboratories during the course of the routine biological examination of the patient, according to World Health Organization (WHO) guidelines (World Health Organization, 2010). The morphology of patients' sperm was assessed with Papanicolaou staining (Fig. 2). Small variations in protocol might occur between the different laboratories. Subjects were recruited on the basis of the identification of >5% of at least four of the aforementioned flagellar morphological abnormalities (absent, short, coiled, bent and irregular flagella) (Table I). Unfortunately, sperm parameters from patient P2 were not available. All subjects presented with severe asthenozoospermia: seven patients had no motility (≤5%), two had sperm motility < 15% and two (P11 and P12) had approximately 30% motility. These latter two patients were considered to have a milder form of MMAF syndrome.

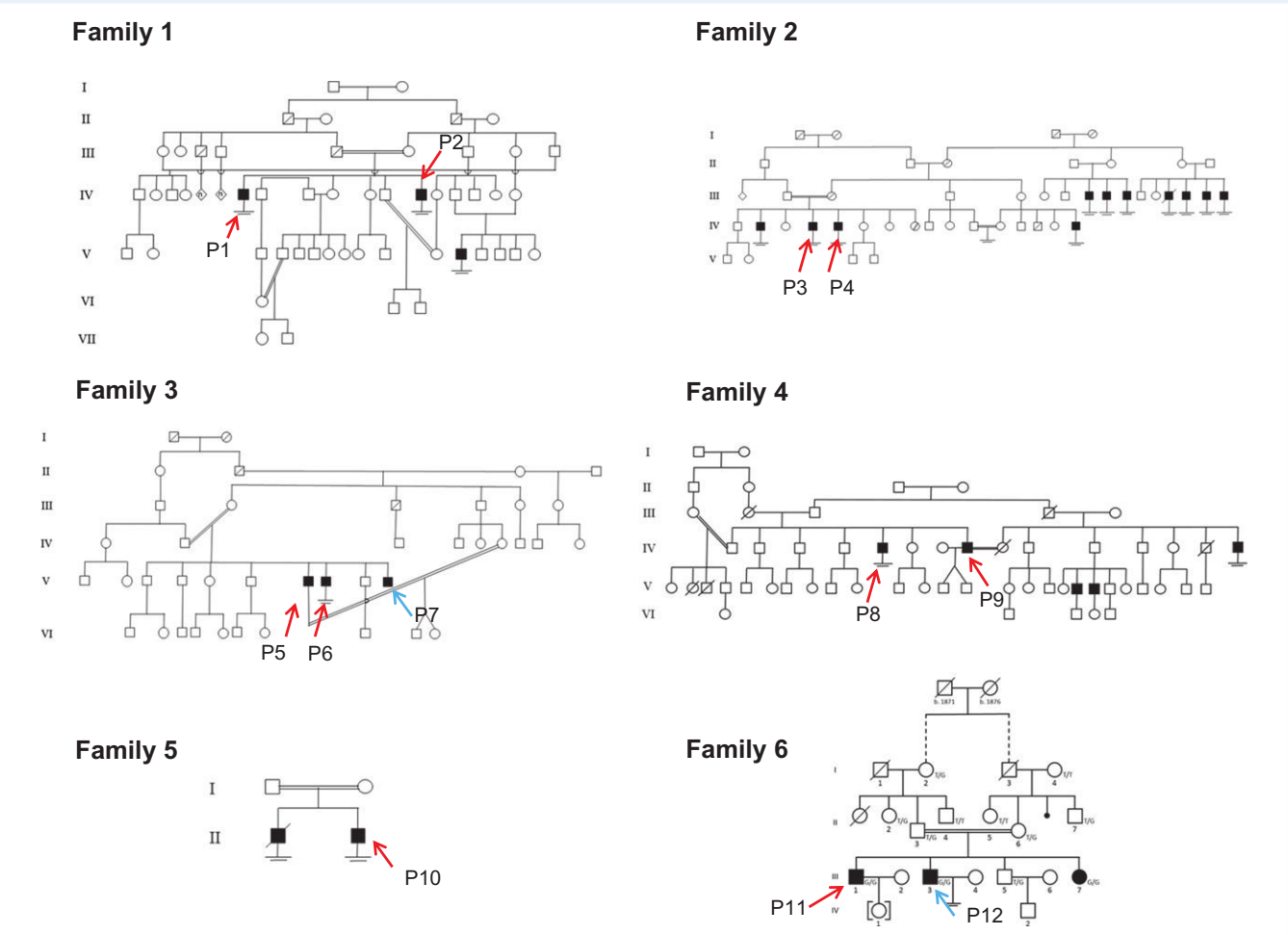


Figure 1 Pedigrees of five Iranian (1–5) and one Italian family (6). Black squares indicate infertile subjects in the family. Red arrows indicate patients for whom WES was performed. Blue arrow indicates patients (P7 and P12) who did not have WES but subsequent exon targeted *DNAH1* Sanger sequencing.

Table I Semen parameters of the 12 patients among whom five subjects (P5, P6, P7, P11 and P12) carried *DNAH1* homozygous variants. Semen parameters from the additional sporadic patient SP4 with a *DNAH1* mutation are also reported. *DNAH1* mutations were identified by WES in patients P5, P6 and P11 and by Sanger sequencing in patients P7, P12 and SP4. Values are the average of two separate analyses.

Semen parameters	Family 1		Family 2		Family 3			Family 4		Family 5	Family 6		SP4
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	SP4
DNAH1 mutation	None	None	None	None	c.8626-1G > A	c.8626-1G > A	c.8626-1G > A	None	None	None	c.3860.T > G	c.3860.T > G	c.8626-1G > A
Sperm volume (ml)	4	NA	3.2	3.5	4	4	4	1	0.9	1	3.5	3	4.5
Sperm concentration (10^6/ml)	20	NA	6	16	29	18	37	52	50	21	24	31	13
Motility (A + B) 1 h (%)	0	NA	15	14.3	5	0	0	0	0	0	27	33	0
Vitality (%)	88	NA	NA	NA	97	50	82	92	0	89	74	85	88
Normal spermatozoa (%)	0	NA	0	1	0	0	0	0	0	3	8	11	0
Anomalies of the head (%)	1	NA	6	34	12	33	22	2	5	14	24	18	13
Absent flagella (%)	0	NA	5	1	0	0	0	0	2	15	5	8	0
Short Flagella (%)	90	NA	53	6	85	65	80	70	70	35	41	48	86
Coiled Flagella (%)	+	NA	+	6	8	1	0	5	1	5	6	12	1

NA = not available. + : anomalies reported (>5%) but not accurately quantified.

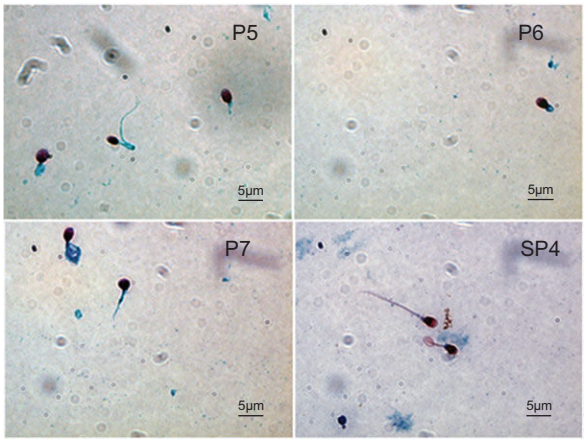


Figure 2 Light microscopy analysis of spermatozoa from three c.8626-1G > A homozygous Iranian brothers (P5, P6, P7) and one sporadic Iranian patient (SP4). Sperm samples were spread over a slide and dried at room temperature for Papanicolaou staining. These images are given as examples of typical spermatozoa observed in mutated patients.

In order to evaluate the incidence of the detected mutation in the general Iranian population, we recruited 38 additional patients (SP1 to SP38) presenting with MMAF phenotype with at less 80% of sperm with flagellar abnormalities (Supplementary Table S1).

Saliva and/or peripheral blood was obtained for all participants. Sperm samples were obtained, following informed consent, from patients P5, P6, P7 and SP4. All subjects answered a health questionnaire focused on PCD manifestations during their medical consultation for infertility. Informed consent was obtained from all the subjects participating in the study according to the local Institutional Review Board protocols and the principles of the Declaration of Helsinki. In addition, the study was approved by local ethic committees.

Molecular analysis

DNA extraction

DNA was extracted from blood or saliva. Blood DNA extraction was carried out from 5 to 10 ml of frozen EDTA blood using the quick guanidium chloride extraction procedure. Saliva was collected with Oragene DNA Self-Collection Kits (DNA genotek Inc., Canada) and DNA extraction was performed following the manufacturer's recommendations.

Exome sequencing and bioinformatics analysis

WES was carried out on DNA extracted from the 10 studied subjects. Coding regions and intron/exon boundaries were enriched using the 'all Exon V5 kit' (Agilent Technologies, Wokingham, UK). DNA sequencing was undertaken at the Plateforme Biopuces et Séquençage IGBMC, Illkirch, France, on the HiSeq 2000 from Illumina®. All steps from sequence mapping to variant selection were performed using the ExSQLibur pipeline (https://github.com/tkaraouzene/ExSQLibur). Sequence reads were aligned to the reference genome (hg19) using MAGIC (SEQC/MAQC-III Consortium, 2014). Duplicate reads and reads that mapped to multiple locations in the exome were excluded from further analysis. Positions whose sequence coverage was below 10 on either the forward or reverse strand were excluded. Single nucleotide variations (SNV) and small insertions/deletions (indels) were identified and quality-filtered using in-house

scripts. The most promising candidate variants were identified using an in-house bioinformatics pipeline. Variants with a minor allele frequency greater than 5% in the NHLBI ESP6500 or in 1000 Genomes Project Phase 1 data sets, or greater than 1% in ExAC were discarded. We also compared these variants to an in-house database of 56 control exomes. All variants present in a homozygous state in this database were excluded. We used Variant Effect Predictor to predict the impact of the selected variants. We only retained variants impacting splice donor/acceptor sites or causing frameshift, inframe insertions/deletions, stop gain, stop loss or missense mutations, except those scored as 'tolerated' by SIFT (sift.jcvi.org) and as 'benign' by Polyphen-2 (genetics.bwh.harvard.edu/pph2).

Sanger sequencing

The two variations identified in *DNAH1* using WES were verified by Sanger sequencing. The coding exon 23 and *intron-exon boundary* adjacent to intron54/*exon55 were amplified as indicated in the* Supplementary Table SII. Sequencing reactions were carried out with BigDye Terminator v3.1 (Applied Biosystems). Sequence analysis were carried out on ABI 3130XL (Applied Biosystems).

RNA extraction

Nucleated cells were isolated from whole blood using Ficoll® 400 (Sigma-Aldrich Corp., St. Louis, MO, USA) following the manufacturer's protocol. RNA extraction was carried out on the isolated white blood cells using Macherey Nagel NucleoSpin® RNA II columns (Macherey Nagel, Hoerdt, France) using the manufacturer's protocol.

RT-PCR

Reverse transcription was carried out in three patients P5, P6, P7 and two healthy controls (C1 and C2) with 5 µl of extracted RNA (approximately 500 ng). Hybridization of the oligo-dT was performed by incubating for 5 min at 65°C and quenching on ice with the following mix: 5 µl of RNA, 3 µl of poly T oligo primers (dT)12–18 (10 mM, Pharmacia), 3 µl of the four dNTPs (0.5 mM, Roche diagnostics) and 2.2 µl of H_2O. Reverse Transcription was then carried out for 30 min at 55°C after the addition of 4 µl of 5× buffer, 0.5 µl RNase inhibitor and 0.5 µl of Transcriptor Reverse transcriptase (Roche Diagnostics). Then 2 µl of the obtained cDNA mix was used for the subsequent PCR. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used as a housekeeping gene (internal control). Primers sequences and RT-PCR conditions are indicated in the Supplementary Table SIII.

Immunostaining

Sperm were fixed in PBS/4% paraformaldehyde for 1 min at room temperature. After washing in 1 ml PBS, the sperm suspension was spotted onto 0.1% poly L-lysine pre-coated slides (Thermo Scientific). After attachment, sperm were permeabilized with 0.1% (v/v) Triton X-100—DPBS (Triton X-100; Sigma-Aldrich) for 5 min at room temperature. Slides were then blocked in 5% corresponding normal serum—DPBS (normal goat or donkey serum; GIBCO, Invitrogen) and incubated overnight at 4°C with primary antibodies. Polyclonal DNAH1 antibodies were purchased from Prestige Antibodies® (Sigma-Aldrich, France) (1:50). Monoclonal mouse anti-acetylated-α-tubulin antibodies were purchased from Sigma-Aldrich (1:2500). Washes were performed with PBS containing 0.1% of Tween 20, followed by 1 h incubation at room temperature with Alexa Fluor 555-labeled goat anti-rabbit or Dylight 488-labeled goat anti-rabbit (1:400) secondary antibodies. Appropriate controls were performed, omitting the primary antibodies. Samples were counterstained with 5 mg/ml Hoechst 33342 (Sigma-Aldrich) and mounted with DAKO mounting media (Life Technology). Whole images were reconstructed and projected from Z-stack images using ZEN software.

Results

WES identified two new *DNAH1* mutations

Given the notion of consanguinity in the families we studied, we postulated that the infertility has been likely transmitted by autosomal recessive inheritance and was thus caused by a homozygous mutation. We proceeded to WES to identify a possible genetic defect that could explain the observed MMAF phenotype. After exclusion of frequent variant and applying stringent filters, a limited list of homozygous variants was identified in each proband. Apart from these three variants in *DNAH1*, no variants were present in genes described to be strongly expressed in the testis nor in any gene described to be connected with cilia, the axoneme or the flagellum. Only the *DNAH1* variants were retained as likely causal (Supplementary Table SIV). In cases where the brothers were simultaneously analyzed by WES, only common variants shared by both brothers were retained. Two different pathogenic mutations in the *DNAH1* gene were identified in Families 3 and 6.

In Family 3, the c.8626-1G > A variant was identified in the two MMAF brothers analyzed by WES. This mutation affects the final G nucleotide of *DNAH1* intron 54, one of the consensus splice acceptor bases (Fig. 3). The alternate splicing is predicted to cause a frameshift in the new transcript and to induce a premature stop codon. The c.8626-1G > A variant was absent from over 60 000 individuals described in the ExAC database which confirms that it is not a polymorphism and that a splicing mutation occurring at this localization would be negatively selected throughout evolution. No other variants were identified in the *DNAH1* coding sequence and UTR regions. Sanger sequencing confirmed the splicing mutation for both infertile brothers (Fig. 3) and showed that the third infertile brother has the same mutation while their father and a non-affected brother were heterozygous (data not shown).

In Family 6, a second mutation was found in *DNAH1*. The identified mutation is c.3860 T > G (p.Val1287Gly). This was confirmed in the same subject by bidirectional Sanger sequencing of *DNAH1* exon 23. No other unreported *DNAH1* variant was identified by exome sequencing. No rare variants were present in other genes reported to be associated with male infertility. The presence of the c.3860 T > G mutation was tested in the other family members available for genotyping by restriction analysis and confirmed by Sanger sequencing. In this large Italian family, the other infertile brother and one sister were homozygous for the *DNAH1* mutation (Fig. 3). Six other related fertile subjects of both sexes were heterozygous for the mutation. The sister, also homozygous for p.Val1287Gly mutation, was healthy by general evaluation and had not yet attempted to have children.

Targeted PCR-Sanger sequencing in 38 MMAF Iranian patients

We next assessed 38 Iranian patients (SP1 to SP38) presenting with morphological abnormality of flagella (Supplementary Table S1) for the c.8626-1G > A splicing mutation identified in the Iranian Family 3. Sequencing results showed that one of these affected men (SP4) is homozygous for this new splice site mutation indicating that this mutation segregates in the Iranian population.

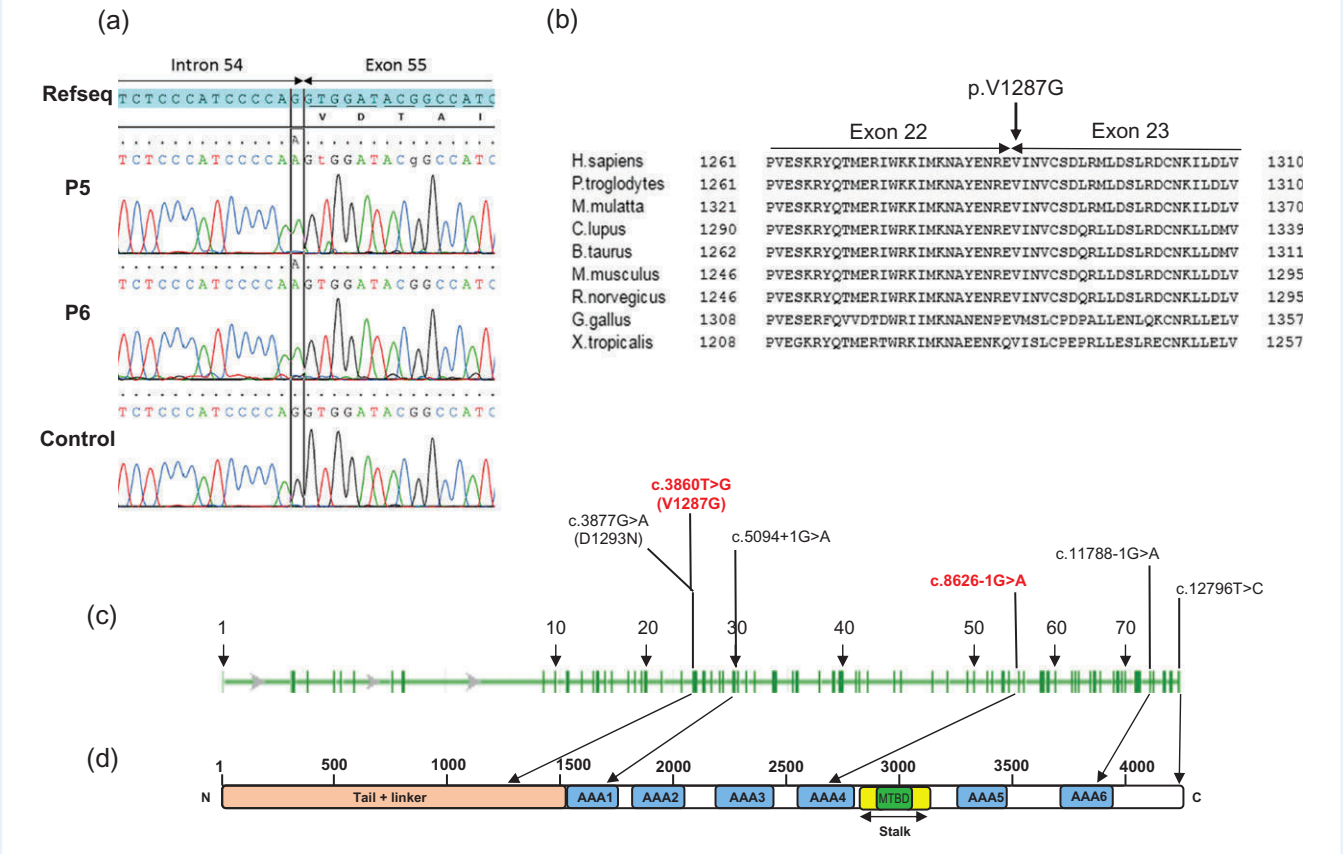


Figure 3 *DNAH1* novel point mutations. (**a**) Electropherogramms of *DNAH1* showing the homozygous splice acceptor site mutation c.8626-1G > A for P5, P6 and a heathy control individual. (**b**) Interspecies alignment. The two Italian brothers carry a homozygous missense mutation c.3860.T > G in *DNAH1* exon 23, leading to the change of the valine in position 1287 into a Glycine. The valine in position 1287 is well conserved between species. Amino acid sequence alignment was realized with homologene (http://www.ncbi.nlm.nih.gov.gate1.inist.fr/homologene/?term=dnah1). (**c**) Localization of the *DNAH1* mutations in the intron-exon structure showing the position of the two novel identified mutations (in red) and the mutations identified in our previous study (in black) (Ben Khelifa *et al.*, 2014). Green boxes indicate coding exons, white boxes indicate noncoding exons. (**d**) Protein domain map of the axonemal dynein heavy chain 1 (DNAH1). The blue boxes indicate the six known AAA-ATPase domains (AAA 1 to 6) as detected by homology (Uniprot server). The microtubule-binding domain (MTBD) lies between AAA4 and AAA5. The N-terminal part of the protein binds to the intermediate, light-intermediate dynein chains. The position of the stalk and the MTBD are indicated.

Detrimental effects of the two identified mutations

To assess the functional impact of the *DNAH1* splice acceptor site mutation c.8626-1G > A, we studied mRNA products isolated from two control and from P5, P6 and P7. RT-PCR of patients' samples yielded no product despite repeated attempts, contrary to what was observed from the controls yielding bands of the expected size (Fig. 4). RT-PCR targeting *GAPDH* confirmed the integrity of patients' RNA (Fig. 4). This suggests a specific degradation of the mutant *DNAH1* transcripts by nonsense mediated mRNA decay (NMD). To further validate the pathogenicity of this variant, we analyzed DNAH1 localization by immunofluorescence on patients' sperm. In control individuals, DNAH1 antisera decorated the full length of the sperm flagellum (Fig. 5). In contrast, in sperm from the three brothers carrying the c.8626-1G > A mutation as well as in sperm from the sporadic case SP4, DNAH1 immunostaining was absent, indicating that the splicing defect induces the degradation of the transcripts by NMD thus precluding protein production (Fig. 5).

Unfortunately, no mRNA analysis or immunostaining could be performed on sperm cells from the Italian patients, P11 and P12. We however found Val1287 to be very well conserved throughout evolution (Fig. 3) and this missense change is also predicted to be likely damaging by SIFT and PolyPhen-2, two prediction software for nonsynonymous SNPs. This variant was also absent from all the control sequence databases (dbSNP v137, 1000 Genomes Project, NHLBI Exome Variant Server).

Discussion

To date, SNP array-based homozygosity mapping has permitted the identification of mutations in three main genes leading to teratozoospermia (Dieterich *et al.*, 2007; Harbuz *et al.*, 2011; Ben Khelifa *et al.*, 2014). This strategy is however time-consuming and requires that several patients

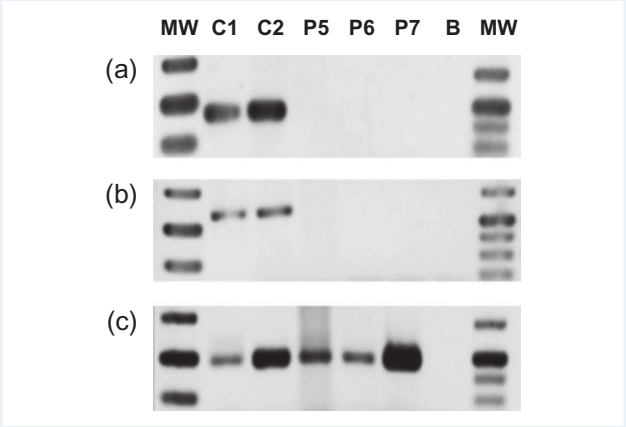


Figure 4 RT-PCR analyses of subjects P5, P6 and P7 (c.8626-1G > A homozygotes) and control individuals from the general population. Electrophoresis showing the RT-PCR amplification of (**a**) *DNAH1* exons 53–56 (**b**) DNAH1 exons 53–57 and (**c**) *GAPDH* amplification used as housekeeping gene. Controls DNA C1 and C2 show good amplification for all tested loci whereas the three patients do not amplify *DNAH1* but only the *GAPDH* control. There is no amplification from the RT-negative blank control (column B).

share a common genetic cause responsible for the phenotype. Due to the high genetic heterogeneity of infertility, SNPs array has now reached its limits and has been supplanted by next generation sequencing and in particular WES. Unfortunately in this study, this strategy has not led to the identification of new reliable candidate genes in MMAF patients. Some deleterious variants have been identified but they do not concern genes that have been described to have a strong connection with spermatogenesis. What appears as a negative result can be explained by: (i) a lack of information regarding a mutated gene explaining why it was not selected; (ii) a false starting assumption based on the hypothesis that the infertility has been likely transmitted by autosomal recessive inheritance; (iii) exclusion criteria being too stringent, eliminating the pathogenic variant (e.g. silent variant modifying splicing sites), (iv) a variant undetectable by the technique used or our current bio-informatic pipeline, e.g., genomic rearrangements (large deletions and duplications), unsequenced deep intronic variants or some exonic variants (as only about 90% of coding nucleotides were covered). These results therefore highlight the fact that WES cannot be expected to provide 100% positive diagnoses. A diagnosis was however reached for 2 out of 6 of the analyzed families (33%). Two new *DNAH1* mutations were identified. These results thus reinforce the fact that *DNAH1* remains the main gene associated with MMAF and corroborate our previous study indicating that approximately 30% of subjects with MMAF are expected to harbor *DNAH1* mutations. These findings raise

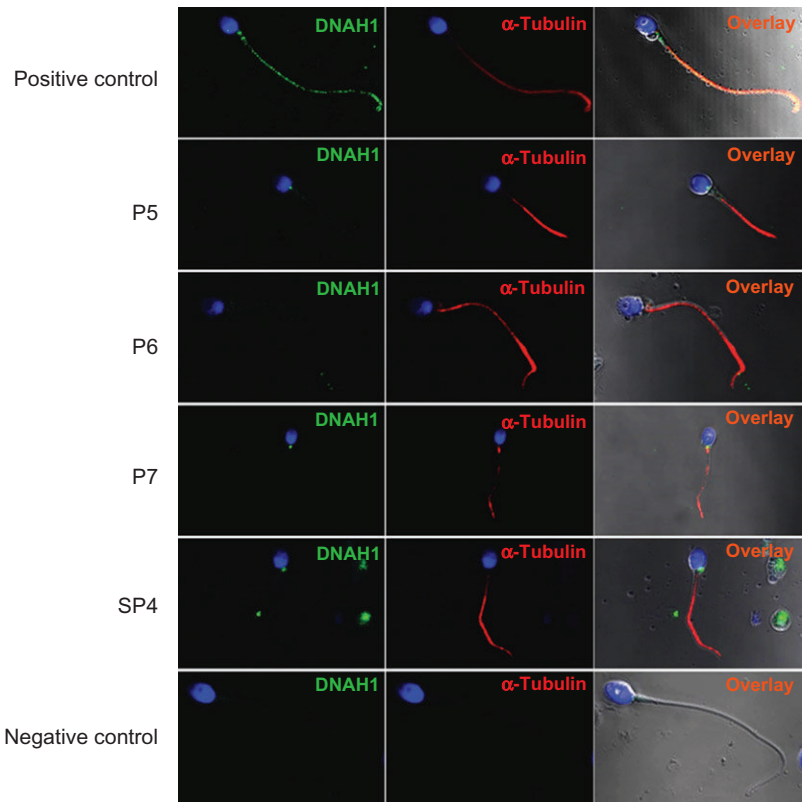


Figure 5 Immunofluorescence images of human spermatozoa from the four patients with the c.8626-1G > A mutation with DNAH1 antibodies and α-Tubulin. DNAH1 staining (green) was observed throughout the flagellum in positive control sperm, whereas it was absent in all sperm samples from patients with the c.8626-1G > A variant. The α-tubulin signal (red) was observed in controls and patients. Sperm were counterstained with Hoechst 33342 (blue) as nuclei marker. These images are given as examples of typical stainings observed in patients.

the question of whether *DNAH1* routine diagnosis should be proposed to all MMAF patients. *DNAH1* is a very large gene (84 Kb and 78 exons) making conventional Sanger sequencing difficult, laborious and costly. In view of our results, we recommend to perform WES in MMAF patients as a first approach. WES in MMAF patients provides a fast way to sequence all *DNAH1* exons while giving the opportunity to identify new gene defects. This strategy, followed by the assessment of ICSI results obtained with patients with different gene defects, could allow improvements in the prediction of ICSI success rates and thus to provide better counseling to MMAF patients. We recently reported that ICSI with spermatozoa from MMAF patients with *DNAH1* mutations had a high pregnancy rate following ICSI (Wambergue *et al.*, 2016). *DNAH1* mutation positive patients identified in this study can thus be encouraged to initiate an IVF/ICSI procedure.

One of the main difficulties associated with WES is the confirmation of the deleterious effect of the identified variant. The effect of the first *DNAH1* variant (c.8626-1G > A) cannot be questioned as: (i) it was identified in four individuals (P5, P6, P7 and SP4); (ii) it affects a consensus splice site known to be essential to the mRNA splicing machinery; and (iii) the mRNA and the protein were shown to be absent from mutated patients. The effect of the second variant, Val1287Gly, is not as easy to predict as it concerns only one amino acid. We however have several arguments in favor of its pathogenicity: (i1) the variant was described as deleterious by two prediction tools; (ii) it was not found in any database now including in excess of 60 000 individuals; and (iii) it affects a conserved residue located in the N-terminal part of the protein known to be important for the structure of the dynein arms (Habura *et al.*, 1999) and it is positioned close to the first missense *DNAH1* mutation (Asp1293Asn) identified previously. The four patients with the severe variant present 0% of morphologically normal spermatozoa with a motility <5% in contrast with the two brothers with the Val1287Gly variant who present a milder phenotype with approximately 30% motility and 10% of morphologically normal spermatozoa (Table I). This suggests that the Val1287Gly mutated protein is likely preserve a residual activity as was previously observed in the patient with the nearby missense mutation. This supports the hypothesis of a phenotype continuum depending on the severity of *DNAH1* mutations (Ben Khelifa *et al.*, 2014). We can therefore expect that individuals harboring homozygous or compound heterozygous *DNAH1* mutations of moderate severity could present with intermediate asthenozoospermia and low levels of morphological anomalies.

When all patients with MMAF analyzed in our previous and present work are pooled, we have a total of 32 patients including 24 index cases. Seven unrelated individuals out of 24 carried a homozygous variant in *DNAH1* (29.2%). Although a large majority of patients are of North African origin, we have now identified *DNAH1* mutations in Middle East and European patients indicating that *DNHA1* diagnosis should not be ruled out for non-Maghrebian individuals. The c.8626-1G > A variant was found in 1/38 sporadic MMAF Iranian cases resulting in a prevalence estimated at 2.6% in this Iranian population. It would be interesting to compare the haplotype of these affected Iranian patients in order to highlight a possible founder effect.

As reported in our previous work, all individuals carrying mutations in *DNAH1* only presented with male infertility and did not report any other symptoms associated with PCD such as an impairment of the respiratory functions (Ben Khelifa *et al.*, 2014). The data reported here is consistent with the hypothesis suggesting that DNAH1 function in cilia may be compensated by other dyneins. The absence of clinical signs does not formally exclude that *DNAH1* mutations could lead to a slight functional impairment of motile cilia function in respiratory epithelium but the impairment would not necessarily lead to lung dysfunction. Unfortunately, family members in which a *DNAH1* mutation was identified did not consent to the investigation of their respiratory function.

The abundance of potential candidate genes makes identification of pathogenic mutations difficult and complex. However, gene identification is the key to improving knowledge of the pathophysiology of MMAF and opens new perspectives for diagnosis and treatment of infertile patients. Further genetic studies are therefore warranted to identify other genes involved in MMAF to better characterize the genetic etiology of the MMAF phenotype and to improve the management of patients diagnosed with flagellar defects.

Supplementary data

Supplementary data are available at http://humrep.oxfordjournals.org/.

Acknowledgments

The authors thank the patients for their interest and cooperation.

Authors' roles

P.F.R., C.C., A.A.Y. and H.G. analyzed the data, wrote the manuscript and had full access to all of the data and take responsibility for the integrity of the data and its accuracy. A.A.Y., Z.E.K., P.L.T. performed the experimental analyses. T.K. performed data mining for the genetic data. M.H.S., M.S., N.A., M.A.S.G., S.H.H., S.B., A.D., C.A., R.C., M.B. and H.G. provided clinical samples and data and supplied biological materials. All authors contributed to the report.

Funding

This work was supported by following grants: the 'MAS-Flagella' project financed by the French ANR and the DGOS for the program PRTS 2014 and the 'Whole genome sequencing of patients with Flagellar Growth Defects (FGD)' project financed by the Fondation Maladies Rares for the program Séquençage à haut débit 2012.

Conflict of interest

None declared.

References

Ayhan O, Balkan M, Guven A, Hazan R, Atar M, Tok A, Tolun A. Truncating mutations in TAF4B and ZMYND15 causing recessive azoospermia. *J Med Genet* 2014;**51**:239–244.

Baccetti B, Collodel G, Estenoz M, Manca D, Moretti E, Piomboni P. Gene deletions in an infertile man with sperm fibrous sheath dysplasia. *Hum Reprod* 2005;**20**:2790–2794.

Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, Shendure J. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 2011;**12**:745–755.

Ben Khelifa M, Coutton C, Zouari R, Karaouzene T, Rendu J, Bidart M, Yassine S, Pierre V, Delaroche J, Hennebicq S *et al.* 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* 2014;**94**:95–104.

Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet* 2013;**14**:681–691.

Chemes EH, Rawe YV. Sperm pathology: a step beyond descriptive morphology. Origin, characterization and fertility potential of abnormal sperm phenotypes in infertile men. *Hum Reprod Update* 2003;**9**:405–428.

Chemes HE, Brugo S, Zanchetti F, Carrere C, Lavieri JC. Dysplasia of the fibrous sheath: an ultrastructural defect of human spermatozoa associated with sperm immotility and primary sterility. *Fertil Steril* 1987;**48**: 664–669.

Coutton C, Escoffier J, Martinez G, Arnoult C, Ray PF. Teratozoospermia: spotlight on the main genetic actors in the human. *Hum Reprod Update* 2015;**21**:455–485.

Dieterich K, Soto RR, Faure AK, Hennebicq S, Ben Amar B, Zahi M, Perrin J, Martinez D, Sele B, Jouk PS *et al*. Homozygous mutation of AURKC yields large-headed polyploid spermatozoa and causes male infertility. *Nat Genet* 2007;**39**:661–665.

El IE, Muller J, Viville S. Autosomal mutations and human spermatogenic failure. *Biochim Biophys Acta* 2012;**1822**:1873–1879.

Escalier D, David G. Pathology of the cytoskeleton of the human sperm flagellum: axonemal and peri-axonemal anomalies. *Biol Cell* 1984;**50**: 37–52.

Escoffier J, Lee HC, Yassine S, Zouari R, Martinez G, Karaouzene T, Coutton C, Kherraf ZE, Halouani L, Triki C *et al*. Homozygous mutation of PLCZ1 leads to defective human oocyte activation and infertility that is not rescued by the WW-binding protein PAWP. *Hum Mol Genet* 2016;**25**:878–891.

Gilissen C, Hoischen A, Brunner HG, Veltman JA. Disease gene identification strategies for exome sequencing. *Eur J Hum Genet* 2012;**20**: 490–497.

Habura A, Tikhonenko I, Chisholm RL, Koonce MP. Interaction mapping of a dynein heavy chain. Identification of dimerization and intermediate-chain binding domains. *J Biol Chem* 1999;**274**:15447–15453.

Harbuz R, Zouari R, Pierre V, Ben Khelifa M, Kharouf M, Coutton C, Merdassi G, Abada F, Escoffier J, Nikas Y *et al*. A recurrent deletion of DPY19L2 causes infertility in man by blocking sperm head elongation and acrosome formation. *Am J Hum Genet* 2011;**88**:351–361.

Matzuk MM, Lamb DJ. The biology of infertility: research advances and clinical challenges. *Nat Med* 2008;**14**:1197–1213.

Stalf T, Sanchez R, Kohn FM, Schalles U, Kleinstein J, Hinz V, Tielsch J, Khanaga O, Turley H, Gips H *et al*. Pregnancy and birth after intracytoplasmic sperm injection with spermatozoa from a patient with tail stump syndrome. *Hum Reprod* 1995;**10**:2112–2114.

Storm van's GK, Omran H. Primary ciliary dyskinesia: clinical presentation, diagnosis and genetics. *Ann Med* 2005;**37**:439–449.

Wambergue C, Zouari R, Fourati Ben MS, Martinez G, Devillard F, Hennebicq S, Satre V, Brouillet S, Halouani L, Marrakchi O *et al*. Patients with multiple morphological abnormalities of the sperm flagella due to DNAH1 mutations have a good prognosis following intracytoplasmic sperm injection. *Hum Reprod* 2016;**31**:1164–1172.

World Health Organization. *WHO Laboratory Manual for the Examination and Processing of Human Semen*, Geneva, 5th edn. World Health Organization, 2010.

Human Reproduction, 2016
doi:10.1093/humrep/dew262

human reproduction

SUPPLEMENTARY DATA

Supplementary Table S1 **Semen parameters of 38 Iranian sporadic patients.**

Patients	Semen volume (ml)	Sperm conc. (10^6/ml)	Total motility 1 h (%)	Vitality (%)	Normal spermatozoa (%)	Anomalies of the flagella (%)			Anomalies of the head (%)			Amorph (%)
						Short	Coiled	Other	Round	Pin	Giant	
SP1	4.50	0.15	0		0	98						2
SP2	5	7	0	60	0	86	2	2		1		9
SP3	0.3	6	0	60	1	92				2		5
SP4*	4.5	13	Few spz	88	0	86	1					13
SP5	2.5	30	0	70	0	98						2
SP6	4	11	Few spz	88	0	89			3	1	2	5
SP7	3	12	0	88	0	92		2				8
SP8	1	5	0	55	1	81				2	6	10
SP9	2.4	6	0	35	0	95				3	2	
SP10	2.1	3	0	82	0	84	1		4			11
SP11	5	4	Few spz	13	0	79	1		1			19
SP12	3.2	0.2	Few spz		0	86	5	2				5
SP13	3	20	0	92	0	96				2		2
SP14	2	20	0	80	0	100						
SP15	2.5	14	Few spz	88	0	86	4	2	2	2		4
SP16	3	1	0	85	0	88	2		4		2	4
SP17	5.5	7	0	88	0	92			3			5
SP18	4	3	0	88	0	90				5		5
SP19	3	16	0	85	0	90	4		2	1		3
SP20	6.5	6	0	80	0	90		3				7
SP21	3.2	5	0	95	0	100						
SP22	4.6	0.5	0	80	0	89	1		7			3
SP23	3.4	25	0	80	0	84	1	8				7
SP24	6.5	4	Few spz	70	0	92						8
SP25	4.2	2	0	35	2	83		2		2		11
SP26	4	5	Few spz	84	1	78		4		2		15
SP27	4.6	25	0	80	0	93	3		2			2
SP28	3	34	0	56	0	71	8	1	1	3		16
SP29	3.4	2.5	Few spz	48	0	79	3	2		1	1	14
SP30	3.5	30	5	75	1	60			5	2		32
SP31	2	5	0	3	0	88	1		1			10
SP32	3	0.25	0		0	86			8	4		2
SP33	5	0.5	Few spz		1	72		2		2		23
SP34	1	8	0	95	0	95						5
SP35	3	1	0		0	80	2		7	1		10
SP36	2.5	7	0	97	0	84	9		2	2		5
SP37	4.5	12	0	84	0	98						2
SP38	2.5	32	0	92	0	83			4	3		10

*Mutated patient.

5- Discussion

5.1- Exploration génétique du phénotype MMAF

Le phénotype MMAF est caractérisé par des anomalies morphologiques multiples du flagelle spermatique responsables asthénozoospermie sévère et d'infertilité. Ce phénotype a souvent été rapporté dans les publications portant sur l'exploration génétique des dyskinésies ciliaires primitives (PCD). Cependant, dans le cadre de ces pathologies principalement pulmonaires, la caractérisation phénotypique des anomalies spermatiques chez ces patients a rarement été détaillée. Inversement, l'exploration génétique des patients ne présentant pas de PCD mais présentant uniquement des anomalies flagellaires n'était jusqu'à récemment pas réalisée et notre équipe a été la première à initier des études génétiques ambitieuses pour le phénotype MMAF pur. En effet, bien que l'ultrastructure du cil mobile et du flagelle spermatique soit très similaire car structurée autour d'un axonème de type « 9+2 », il existe des formes de PCD et de MMAF dissociées. Parmi les hypothèses pouvant expliquer cette dissociation, la présence de structures para-axonémales spécifiques du flagelle comme la gaine mitochondriale, la gaine fibreuse et les fibres denses. Une autre hypothèse est justifiée par la nature des isoformes exprimés différentiellement dans le cil et le flagelle. En 2005, Baccetti et al. ont identifié chez un sujet MMAF une délétion partielle du gène *AKAP4* qui code pour une protéine d'ancrage des kinases A dans la gaine fibreuse du flagelle spermatique (Baccetti et al., 2005). La protéine AKAP4 permet en phosphorylant ces protéines cibles d'initier la transduction du signal de régulation du mouvement flagellaire. Il a été observé auparavant que l'inactivation de ce gène chez la souris induit une altération de la morphologie et de la mobilité progressive des spermatozoïdes associée à une infertilité. Par ailleurs, l'étude de l'ultrastructure des spermatozoïdes de ces animaux a montré une dysplasie de la gaine fibreuse (Miki et al., 2002). En 2014, l'équipe GETI, a identifié des mutations homozygotes dans le gène *DNAH1* en utilisant une stratégie de cartographie d'homozygotie chez 5 sujets parmi les 18 testés (28%) (Ben Khelifa et al., 2014). Le gène *DNAH1* code pour une chaine lourde de dynéine axonémale. L'étude de l'ultrastructure des spermatozoïdes provenant de ces patients a révélé une dysplasie de la gaine fibreuse associée à une désorganisation profonde de l'axonème et des autres structures para-axonémales. Avec l'avènement du séquençage exomique, notre équipe a poursuivi l'exploration génétique du phénotype MMAF en utilisant cette technologie. Nous avons réalisé le séquençage exomique pour 8 familles (12 sujets au total) d'origine iranienne ou italienne et avons identifié des mutations homozygotes délétères du gène *DNAH1* chez deux familles. Au total, 5 sujets étaient mutés dont deux cas index : un

sujet iranien avec une mutation homozygote d'épissage, c.8626-1G>A et un sujet italien avec une mutation homozygote faux-sens, c.3860T>G (p.Val1287Gly) prédite comme délétère. Ces résultat démontrent d'une part la fréquence des mutations du gène DNAH1 associées au phénotype MMAF et d'autre part, suggère que ce phénotype est liée à une assise génétique hétérogène et que *DNAH1* n'est probablement pas le seul gène impliqué dans sa pathogénie.

Dans ce contexte, nous avons poursuivi notre exploration génétique de ce phénotype en réalisant le séquençage exomique (WES) chez une cohorte de 78 individus MMAF non apparentés et avons identifié au total 11 gènes candidats dont 5 ont été présentés dans ce manuscrit. Ces résultats montrent d'une part les performances du WES dans l'exploration génétique des maladies héréditaires autosomiques récessives et d'autre part, la forte hétérogénéité génétique du MMAF. Le rendement diagnostique du WES dans ce phénotype est de l'ordre de 42% en tenant compte uniquement des 5 gènes présentés dans cette thèse (*DNAH1*, *CFAP43*, *CFAP44*, *WDR66* et *FSIP2*). Si l'on rajoute le nombre de sujets MMAF de la même cohorte chez qui nous avons identifié d'autres très bons gènes candidats, le rendement diagnostique atteint alors 63% **(Tableau 4)**. Ces nouveaux gènes, sont à l'heure de la rédaction de cette thèse, tous validés et un gène, *CFAP69* a été récemment publié (Dong et al., 2018).

Gènes identifiés et validés	**N° sujets diagnostiqués**
DNAH1	6 (8%)
CFAP43	10 (13%)
CFAP44	6 (8%)
WDR66	7 (9%)
FSIP2	4 (5%)
CFAP69	2 (2.5%)
FLAG-a	3 (4%)
FLAG-b	3 (4%)
FLAG-c	3 (4%)
FLAG-d	3 (4%)
FLAG-e	3 (4%)
Total	**49/78 (63%)**

Tableau 4. Liste des gènes identifiés par WES chez une cohorte de 78 sujets MMAF.

Depuis notre première étude du phénotype MMAF par WES qui nous a permis d'identifier des mutations causales sur le gène *DNAH1*, d'autres équipes dans le monde notamment en Chine ont commencé à s'intéresser fortement à ce phénotype. En 2017, une équipe chinoise a publié une étude dans laquelle les auteurs ont investigué 9 sujets MMAF par WES. Parmi ces sujets 4 avaient la même mutation frameshift c.11726_177delCT dans l'exon 73/78 du gène *DNAH1* (Wang et al., 2017). Une autre équipe chinoise a investigué les mutations du gène *DNAH1* par WES et a rapporté un total de 17 mutations dont une mutation homozygote et 17 mutations hétérozygotes chez 12 sujets parmi les 21 testés suggérant la possible association de ces mutations avec le phénotype MMAF (Sha et al., 2017a). Tang et al. ont investigué 12 cas de MMAF sporadiques idiopathiques par WES et CGH array et ont identifié des mutations homozygotes et hétérozygotes composites dans trois gènes qui codent pour des protéines associées aux cils et flagelles (*CFAP43*, *CFAP44* et *CFAP65*) chez 5 sujets (41%) (Tang et al., 2017). Cette équipe a également validé ces gènes chez des souris knock-out produites par la technique CRISPR/Cas9. Les souris mâles homozygotes étaient infertiles et présentaient un phénotype spermatique MMAF. Dans une autre étude Sha et al. ont séquencé un sujet MMAF appartenant à une famille consanguine et ont identifié une mutation homozygote faux-sens prédite comme délétère dans le gène CEP135 (c.A1364T :p.D455V) qui code pour une protéine nécessaire pour la biogenèse du centriole et l'assemblage de la paire centrale de l'axonème (Sha et al., 2017b). L'analyse par immunofluorescence et microscopie confocale de l'expression subcellulaire de la protéine mutée au sein des spermatozoïdes de ce patient a permis de mettre en évidence des agrégats protéiques accumulés dans la partie proximale de la pièce intermédiaire du flagelle à proximité du centriole. Par ailleurs, les auteurs ont observé une diminution significative du marquage CEP135 dans les centrioles spermatiques de ce patient. On note cependant qu'il est toujours difficile de prédire l'effet des mutations faux-sens telles que celles observées chez le patient décrit. On ne peut pas exclure la possibilité que les anomalies protéiques observées pour CEP135 peuvent être la conséquence et non à l'origine de l'atteinte. Le niveau de preuve pour l'implication de CEP135 est donc faible.

Une des principales questions à se poser actuellement concerne l'impact des mutations génétiques impliquées dans le phénotype MMAF sur le développement embryonnaire précoce. L'évaluation de la qualité nucléaire des spermatozoïdes tant génomique (fragmentation de l'ADN) qu'épigénétique ainsi que le pronostic du développement embryonnaire suite à une FIV-ICSI pour chaque génotype est une seconde

priorité dans le processus d'exploration de ce phénotype. Plusieurs études ont déjà été réalisées dans ce sens. Notre équipe a montré que les patients porteurs de mutations dans le gène *DNAH1*, considéré comme le premier gène associé au MMAF obtenaient des taux de fécondation et de grossesse comparables à ceux des autres patients MMAF, tous génotypes confondus et légèrement supérieurs à la moyenne des patients réalisant une FIV-ICSI (Wambergue et al., 2016). Sha et al ont également analysé les résultats de FIV-ICSI obtenus à partir des spermatozoïdes du patient muté pour *CEP135*. Cette procédure n'a pas permis d'obtenir une grossesse après la transplantation d'embryons malgré une augmentation initiale de l'HCG jusqu'à 57 mUI/mL (Sha et al., 2017b). Ces auteurs expliquent cet échec par le potentiel impact de la mutation du gène *CEP135* sur la fonction du centriole notamment lors du développement embryonnaire précoce.

5.2- Intérêt des modèles animaux dans l'étude du phénotype MMAF

La validation des variants candidats et des gènes associés aux phénotypes explorés représente actuellement un défi majeur. La complexité de la mise en place du système reproductif male pendant la vie intra-utérine et sa régulation en période post-natale notamment à partir de la puberté rend impossible l'exploration entière de la fertilité *in vitro*. Le recours aux modèles animaux est donc critique pour confirmer le lien entre les causes génétiques identifiées et le phénotype observé et caractériser la physiopathologie associée. En raison de sa forte homologie génétique avec l'Homme, la souris représente le modèle le plus utilisé dans le domaine de la recherche biomédicale y compris en biologie de la reproduction. En effet, environ 85% des gènes de la souris possèdent un orthologue humain (Uhl and Warner, 2015). L'approche la plus répandue pour étudier la fonction du gène candidat chez la souris est la création d'une lignée transgénique knock-out pour ce gène. Traditionnellement, ces souris étaient produites par mutagenèse insertionnelle ou par des stratégies de transgénèse ciblée utilisant des cellules souches embryonnaires (Wang et al., 2013). Ce processus est cependant laborieux et coûteux et c'est pour ces raisons que des technologies alternatives se sont développées. Les nouvelles techniques reposent principalement sur des systèmes de nucléases artificielles cible-dépendantes comme les ZFNs et les TALENs (Barman et al., 2017). Cependant, les difficultés pratiques associées au design et à la synthèse de ces nucléases ont limité la diffusion de leur utilisation. Plus récemment, le développement du système CRISPR/Cas9 a permis de simplifier la manipulation et l'édition du génome. En effet, son utilisation ne nécessite pas la synthèse d'une protéine spécifique pour chaque cible génomique (Sapranauskas et al., 2011). C'est pourquoi son utilisation s'est rapidement et

largement diffusée en recherche fondamentale et translationnelle. Pour la production des souris transgéniques knock-out ou knock-in, la Cas9, le sgRNA et le ssDNA (si nécessaire) sont injectés simultanément dans le pronucléus mâle du zygote (voir article 1 de ce manuscrit). Les souris adultes KO sont obtenues en moins d'un an. L'analyse phénotypique de ces souris permet de caractériser leur fertilité et les phénotypes spermatiques et testiculaires associées. L'autre intérêt de créer ces souris est d'étudier la physiopathologie qui découle de l'inactivation du gène candidat. Dans le domaine de l'infertilité masculine, en raison de la grande similarité physiologique que présente les système reproductifs humain et murin, plusieurs aspects peuvent être explorés comme la spermatogenèse, la maturation spermatique, la capacitation et la fécondation (Jamsai and O'Bryan, 2011). Enfin, ces propriétés génétiques et physiologiques ainsi que les avancées techniques dans la génération de lignées transgéniques ont fait de la souris un excellent modèle d'étude dans l'infertilité masculine comme le montrent d'ailleurs les résultats des travaux de cette thèse.

Nous avons pu obtenir des souris KO pour le gène Spink2 qui étaient disponibles dans le consortium international de phénotypage de la souris (IMPC). L'étude phénotypique des souris mâles KO a permis de montrer qu'elles étaient infertiles et que leurs phénotypes spermatiques et testiculaires étaient en parfaite corrélation avec ceux observés chez les patients mutés pour le gène orthologue. De plus, la localisation acrosomale de la protéine à partir du stade de spermatide ronde était conservée entre les deux espèces. Nous avons également décidé de produire des lignées de souris transgéniques pour confirmer l'implication de certains gènes candidats identifiés par séquençage exomique chez des patients présentant des anomalies morphologiques multiples du flagelle spermatique. Nous avons utilisé le système CRISPR/Cas9 et avons réussi à obtenir pour 4 gènes, plusieurs lignées KO et une lignée KI. L'analyse phénotypique de ces souris mâles montrait qu'elles présentent un phénotype MMAF plus ou moins marqué selon le gène étudié et surtout qu'elles étaient infertiles. Finalement, le recours à ces souris transgéniques nous a permis de valider l'implication de ces gènes de manière efficace et dans des délais raisonnables.

Si l'étude de la spermatogenèse in vivo requiert un modèle animal complexe, les études restreintes au processus de biogenèse du flagelle et l'étude de sa fonction peuvent être réalisées sur des modèles eucaryotes plus simples comme les organismes unicellulaires ciliés ou flagellés. En effet, en termes de composition moléculaire et mode d'assemblage, ces organites sont remarquablement conservés au cours de l'évolution des eucaryotes (Vincensini et al., 2011). Parmi les organismes les plus utilisés en recherche biomédicale, on retrouve

l'algue verte *Chlamydomonas* et les protistes tels que *Tetrahymena*, *Trypanosoma* et *Leishmania*. Pour valider certains gènes candidats dans le phénotype MMAF nous avons fait recours au modèle *Trypanosoma brucei*. Ce parasite cause chez l'Homme la trypanosomiase africaine communément appelée maladie du sommeil. L'implication de ce parasite dans cette maladie tropicale a favorisé son étude approfondie et actuellement, il apparait comme un excellent modèle pour l'étude de la biogenèse et la fonction du flagelle. Le flagelle de *Trypanosoma* comporte des similarités structurales et biologiques avec le flagelle spermatique de l'homme comme la présence d'un axonème de type 9+2 et un assemblage contrôlé par le transport intra-flagellaire. Cependant des différences existent comme la présence d'une structure paraflagellaire unique, le PFR (paraflagellar rod) et la connexion du flagelle avec le corps cellulaire dans presque toute sa longueur. Nous avons réalisé des études fonctionnelles chez *Trypanosoma* pour vérifier l'implication des gènes *CFAP43*, *CFAP44* et WDR66 dans le phénotype MMAF. Nous avons réprimé l'activité des orthologues de ces trois gènes par une méthode d'ARN-interférence (RNAi) et avons observé une altération de la mobilité flagellaire. Nous avons également localisé ces protéines chez ce protiste par STED (Immuno-electron and stimulated-emission-depletin) après l'insertion d'un tag immunogène. Cette étude nous a permis de confirmer la localisation axonémale de WDR66 et de définir la localisation des orthologues de CFAP43 et CFAP44 entre les microtubules périphériques de l'axonème et le PFR (paraflagellar rod), une structure similaire aux structures para-axonémales du flagelle spermatique comme les fibres denses et la gaine fibreuse. Ce modèle unicellulaire nous a montré qu'il pourrait présenter une excellente alternative pour valider les gènes candidats dans les phénotypes d'infertilité masculine liés à des anomalies du flagelle spermatique.

5.3- Conclusion

Pour conclure ce chapitre, nous avons investigué par WES le phénotype MMAF pur chez des sujets non apparentés ayant consulté pour une infertilité primaire isolée et avons montré l'implication de 11 gènes au total dans ce phénotype. Parmi ces gènes, j'ai présenté dans la partie expérimentale *DNHA1*, *CFAP 43*, *CFAP44*, *WDR66* et *FSIP2* et notre équipe a également démontré l'implication du gène *CFAP69* dans ce phénotype. Nous avons observé que les protéines codées par ces gènes sont localisées dans le flagelle spermatique. Néanmoins, leur localisation au sein de cet organite est très différente **(Figure 43)**. DNAH1 et WDR66 sont des protéines de l'axonème. DNAH1 est localisée dans les bras internes de dynéines et WDR66 est décrite pour être localisée à proximité de la base du pont radiaire 3 (RS3, radial spoke 3) qui fait partie du CSC (calmoduline and spoke associated complex), une structure axonémale nécessaire à la régulation de l'activité motrice des dynéines. CFAP43 et CFAP44 sont probablement localisées en périphérie de l'axonème entre les doublets de microtubules périphériques et les structures para-axonémales du flagelle spermatique par extrapolation des données obtenues par l'étude du modèle *Trypanosoma*. FSIP2 est localisée très probablement dans la pièce principale du flagelle au niveau de la gaine fibreuse et interagit avec AKAP4, une protéine d'ancrage des kinases dans cette gaine. Cette protéine peut être impliquée également dans le transport intra-flagellaire au cours de l'assemblage et l'élongation de cet organite. Finalement, des travaux réalisés en collaboration avec une équipe américaine ont permis à notre équipe de montrer que la protéine CFAP69 est présente dans la pièce intermédiaire au niveau de la gaine mitochondriale. On voit donc que l'absence de protéines localisées dans différentes structures internes du flagelle entraine un phénotype MMAF sans que nous n'ayons pu mettre en évidence chez l'Homme de réelle différence phénotypique. Ceci suggère qu'un défaut fonctionnel au niveau de toute protéine jouant un rôle critique au niveau de la formation ou de la structure du flagelle est susceptible d'entrainer un phénotype MMAF. Bien que nous ayons validé l'implication de ces gènes dans le phénotype MMAF, d'autres études doivent être menées dans le futur pour mieux caractériser le rôle de ces protéines au cours de la spermatogenèse et mieux comprendre pourquoi certaines mutations dans des gènes codant pour des protéines axonémales causent un phénotype MMAF dissocié des dyskinésies ciliaires primitives et vice versa pourquoi certaines ciliopathies ne sont pas associées à une infertilité masculine et un phénotype MMAF.

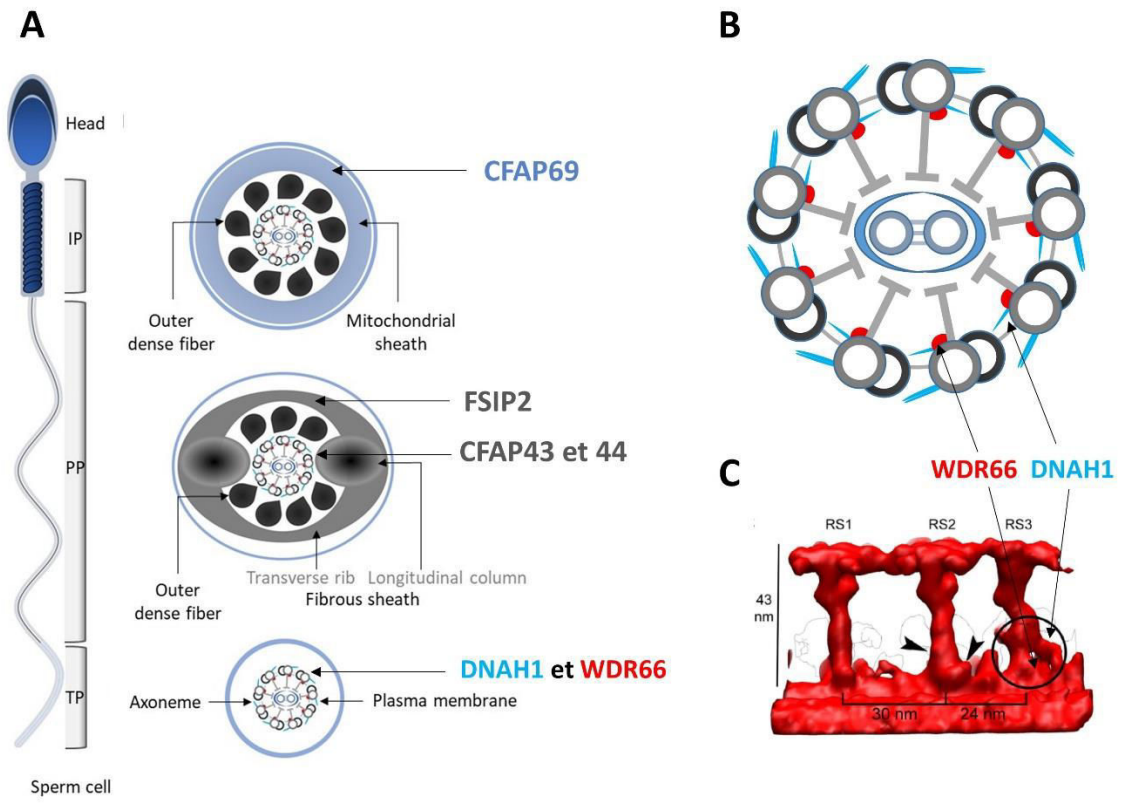


Figure 43. Synthèse sur la localisation intra-flagellaire de certaines protéines impliquées dans le phénotype MMAF chez nos patients.
(A) Présentation schématique du flagelle spermatique chez l'Homme et son ultrastructure interne. (B) Vue transversale d'un axonème flagellaire. (C) Vue longitudinale d'une unité axonémale fondamentale (d'après *Z-E KHERRAF, GETI*).

PARTIE 3 : Discussion générale

1- Exploration génétique de l'infertilité par séquençage haut débit

L'infertilité masculine est caractérisée par une étiologie multifactorielle souvent associée à une forte composante génétique notamment lorsqu'il s'agit de phénotypes sévères comme l'azoospermie et les tératozoospermies monomorphes. Dans ces cas, la pathogénie peut être monofactorielle et monogénique. Cependant, l'abondance de gènes potentiellement candidats rend l'identification des mutations responsables du phénotype difficile et complexe. Malgré les explorations génétiques réalisées depuis plusieurs années, la très grande majorité des cas d'infertilité masculine reste classée comme idiopathique. Néanmoins, le séquençage exomique est actuellement en train de révolutionner ce domaine.

Contrairement au séquençage conventionnel par la technique de Sanger qui permet de réaliser le séquençage progressif des fragments, le séquençage haut débit permet de réaliser un séquençage de masse en parallèle avec un coût raisonnable. Le séquençage exomique, une variante du séquençage haut débit permet de limiter le séquençage aux régions codantes qui représentent 1% du génome et de réduire ainsi le coût de la procédure en se concentrant sur les régions les plus susceptibles de contenir des variations de séquence pathogènes. Actuellement, les kits de séquençage exomique permettent aussi la capture des régions non codantes flanquantes comme les jonctions d'épissage et les régions 5'- et 3'-UTRs (untranslated regions). Depuis son avènement, le séquençage exomique a connu un grand succès en clinique et en recherche translationnelle notamment dans l'exploration génétique des maladies mendéliennes. Ces maladies sont souvent causées par une mutation génique délétère qui apparait de novo chez un individu puis se transmet vers sa descendance. Il s'agit principalement de maladies rares. Dans ces formes, les variants responsables sont aussi rares dans la population générale et ne sont pas présents chez les sujets contrôles permettant de faciliter leur détection et mise en évidence.

Cette technique sera probablement supplantée dans les années à venir par le séquençage du génome complet. Cette méthode de séquençage, est exhaustive (ou presque) et permet d'explorer les régions non codantes et de mieux détecter les variants de nombre de copies (CNVs) ce qui permettra d'améliorer le rendement du diagnostic génétique de l'infertilité. Une des questions majeures à se poser actuellement concerne l'intérêt clinique de l'exploration génétique chez les patients infertiles.

En effet, la connaissance du diagnostic précis de la cause de l'infertilité permet au médecin de mieux orienter le patient infertile dans le parcours de soin et éviter des explorations invasives et des traitements inutiles. Dans le cas de l'azoospermie non-obstructive, la pratique de la biopsie testiculaire à visée thérapeutique n'est pas dénuée de risques. La connaissance préalable de la corrélation génotype/phénotype permet de prédire le pronostic lié à la réussite de cette procédure et guider ainsi le médecin dans sa décision concernant l'indication de ce geste. La connaissance de cette corrélation a également un impact sur l'indication de certains traitements ou procédures. A titre d'exemple, un diagnostic positif du gène *AURKC* chez un patient présentant une macrozoospermie indique qu'une prise en charge par FIV-ICSI ne permettra pas d'obtenir de grossesse et que le couple doit envisager une solution alternative telle que l'adoption. Le diagnostic génétique du phénotype MMAF est pour l'instant moins impactant. En effet, les premières études ont montré que la prise en charge par FIV-ICSI était de bon pronostic chez ces patients. Tous les gènes n'ont cependant pas été identifiés et il est possible que certains gènes, en particulier s'ils sont associés au centriole, soient associés à un pronostic plus défavorable. Cela a été suggéré pour *CEP135*. Un diagnostic génétique dans un contexte de MMAF peut également potentiellement permettre de diagnostiquer une dyskinésie ciliaire primitive à bas bruit qui n'aurait pas été explorée correctement précédemment. Enfin, la connaissance des causes génétiques permet d'envisager le développement de nouvelles stratégies thérapeutiques personnalisées pour restaurer une spermatogenèse fonctionnelle et rétablir une fertilité spontanée notamment dans le cadre de l'azoospermie non-obstructive.

2- Approches thérapeutiques ciblées dans l'infertilité masculine

L'approche thérapeutique innovante dans le cadre de l'infertilité masculine ne peut se concevoir sans prendre en considération la difficulté éthique et réglementaire soulevée par la correction des mutations génomiques dans les cellules germinales chez les patients affectés. En effet, la modification du génome de la lignée germinale humaine pour des raisons thérapeutiques représente actuellement une frontière infranchissable notamment en France. Dans ce contexte, la publication d'une équipe chinoise en 2015 portant sur la modification génétique par CRISPR/Cas9 des embryons humains dans l'objectif de corriger le gène de la β-globine responsable de la thalassémie avait rencontré de très vives réactions au point d'appeler la communauté scientifique internationale à un moratoire pour bannir la modification génétique des embryons humains (Liang et al., 2015). En réalité, cette indignation n'était probablement pas dirigée principalement contre la manipulation génique

dans le cadre thérapeutique mais essentiellement contre le risque de dérives au-delà de ce champ d'application. D'autre part, pour certains, le diagnostic préimplantatoire participe à limiter le risque de transmission des anomalies géniques à la descendance et rend alors inutile ces approches thérapeutiques dans la grande majorité des cas. Il apparait cependant important de conduire des recherches fondamentales et précliniques afin d'évaluer la faisabilité, l'efficacité et l'innocuité des techniques de correction des anomalies géniques au sein des cellules germinales pour une potentielle application clinique (Camporesi and Cavaliere, 2016). Dans le cas de l'infertilité masculine, la correction génétique des cellules germinales testiculaires permettrait de restaurer une spermatogenèse fonctionnelle, de rétablir la fertilité chez l'homme et de limiter la transmission des anomalies génétiques responsables d'infertilité à la descendance, des objectifs qui dans leur globalité semblent assez louables.

Plusieurs stratégies de thérapies géniques ciblant les cellules germinales testiculaires se sont développées ces dernières années notamment chez le modèle murin (Vij et al., 2018). Parmi ces stratégies innovantes on retrouve l'autogreffe des cellules souches spermatogoniales modifiées par génie-génétique. La procédure consiste à prélever des spermatogonies à partir d'une biopsie testiculaire du patient et de les mettre en culture *in vitro* et les modifier par le système CRISPR/Cas9 avant de les retransférer dans le testicule du donneur (Chapman et al., 2015). Ce transfert se fait par injection rétrograde dans le rete-testis ou le canal déférent. Les spermatogonies injectées vont rapidement coloniser l'épithélium séminifère et s'engager dans le processus de spermatogenèse pour produire des spermatozoïdes fonctionnels. Cependant, le risque d'instabilité épigénétique que pourrait induire l'étape de culture *in vitro* reste à clarifier. En effet, cette instabilité pourrait constituer un risque de carcinogenèse. Il est donc important de réaliser des études précliniques supplémentaires dans l'objectif d'évaluer le risque d'instabilité épigénétique et de carcinogenèse (Forbes et al., 2018). L'autre limite de cette approche concerne les risques encourus par la technique employée dans l'édition du génome de ces cellules. En effet, le système CRISPR/Cas9 pourrait induire un mosaïcisme embryonnaire et des effets *off-target* aléatoires et imprévisibles qui pourraient avoir des conséquences cliniques néfastes. Il est donc important d'optimiser ce système avant d'envisager son utilisation chez l'Homme (Harper et al., 2018). Actuellement, des versions modifiées de la protéine Cas9 ayant une meilleure spécificité vis-à-vis de la région cible sont en cours de développement. On espère que ces nouvelles nucléases pourront réduire voire supprimer le risque des effets *off-target* (Kadam et al., 2018).

Au vu des restrictions éthiques et législatives abordées précédemment, il semble également pertinent d'explorer d'autres pistes thérapeutiques ne modifiant pas le patrimoine génétique transmissible des cellules germinales comme l'utilisation des systèmes de vectorisation. Idéalement, ces systèmes doivent être faciles à produire et à utiliser et avoir une bonne efficacité pour permettre une expression suffisante de l'agent vectorisé au sein des cellules germinales avec la moindre toxicité intra- et extra-testiculaire. Dans l'objectif d'une extrapolation vers la thérapie humaine, ce système ne doit pas s'exprimer de manière permanente dans les cellules germinales et donc doit cibler soit les cellules somatiques de Sertoli de manière permanente soit les cellules germinales de façon transitoire (Darbey and Smith, 2017). Parmi les vecteurs répondant à ces critères, les liposomes constituent un bon moyen de vectorisation d'une multitude d'agents thérapeutiques de natures différentes comme les protéines, l'ADN, l'ARNm et les siRNA. Les liposomes sont des vésicules composées d'une enveloppe lipidique renfermant un volume aqueux. La fusion de cette vésicule avec la membrane cytoplasmique de la cellule cible permet de transférer le contenu du liposome à l'intérieur du cytoplasme et du noyau. Il existe actuellement différentes variétés de liposomes comme ceux intégrant dans leur surface des anticorps spécifiques pour une cellule particulière augmentant ainsi la spécificité de ces vecteurs et diminuant leur toxicité. Cependant, ce système ne devrait pas être utilisé pour délivrer des transgènes dans les cellules germinales en raison du risque élevé d'intégration génomique (Miao and Zhang, 2011). Dans cette approche, d'autres systèmes de vectorisation utilisant des lentivirus dits non intégratifs sont en cours de développement. L'injection de ces vecteurs cause une infection transitoire des cellules germinales permettant l'expression du gène administré pendant une phase de la spermatogenèse ou pendant la totalité de ce processus. En raison de l'inactivation du gène de l'intégrase lentivirale, le virus infectant est incapable d'intégrer le transgène dans la cellule hôte. Cependant, il persiste un risque, même faible, d'intégration aléatoire dans le génome qui est à prendre en considération. Néanmoins, cette stratégie semble être efficace et permet de restaurer une spermatogenèse fonctionnelle chez le modèle murin transgénique (Li et al., 2013). Finalement, je suis persuadé que les développements futurs ainsi que la maitrise de ces approches de thérapies ciblées vont permettre d'envisager la réalisation d'essais thérapeutiques cliniques personnalisés chez les patients infertiles dans l'objectif de rétablir transitoirement une spermatogenèse fonctionnelle et d'offrir à ces couples la possibilité d'obtenir un enfant en restant dans le cadre d'un projet intraconjugal.

3- Conclusion

La génétique de l'infertilité masculine est un domaine en plein essor. Ces dernières années, grâce à l'avènement et la disponibilité du séquençage exomique, de nouvelles équipes à travers le monde se sont lancées dans l'exploration génétique et physiopathologique des troubles du système reproductif masculin et plus particulièrement des troubles sévères de la spermatogenèse conduisent à une infertilité. L'apport de l'équipe GETI dans ce domaine est conséquent. Après l'identification et la caractérisation fonctionnelle des gènes *AURKC* dans la macrozoospermie, *DPY19L2* dans la globozoospermie, puis *DNAH1* dans le phénotype MMAF, nous nous sommes focalisés sur l'azoospermie non obstructive (ANO) et le phénotype MMAF. Notre travail récent nous a permis d'identifier le gène *SPINK2* dans l'ANO et l'oligotératozoospermie sévère et de détailler son rôle au cours de la spermatogenèse normale et la physiopathologie qui découle de son altération. Nous avons également amélioré significativement notre rendement de diagnostic génétique dans le phénotype MMAF en passant de 30 à 63%. Nous avons montré que ce phénotype est caractérisé par une forte hétérogénéité génétique. En plus du gène *DNAH1* précédemment décrit, nous avons identifié des mutations homozygotes délétères dans dix autres gènes flagellaires dont *CFAP43*, *CFAP44*, *WDR66*, *FSIP2* et *CFAP69*. Nous poursuivons actuellement notre investigation de ces deux phénotypes (MMAF et ANO) par le recrutement de plus larges cohortes et le renforcement de notre analyse bio-informatique des données du séquençage exomique. Je pense donc que ce travail ainsi que celui des autres équipes concurrentes va rapidement permettre de décrypter la majorité des causes génétiques dans l'infertilité masculine associée aux troubles sévères de la spermatogenèse. Ces avancées devraient nous permettre d'initier des essais précliniques de thérapies ciblées chez le petit animal. J'espère enfin que nos patients pourront ensuite profiter de ces avancées scientifiques et bénéficier de thérapies personnalisées permettant de restaurer, au moins de manière temporaire, une spermatogénèse fonctionnelle permettant une prise en charge efficace de leur infertilité.

Références

Agarwal, A., Mulgund, A., Hamada, A., and Chyatte, M.R. (2015). A unique view on male infertility around the globe. Reprod. Biol. Endocrinol. RBE *13*, 37.

Ahmadi, A., and Ng, S.-C. (1999). Destruction of Protamine in Human Sperm Inhibits Sperm Binding and Penetration in the Zona-Free Hamster Penetration Test but Increases Sperm Head Decondensation and Male Pronuclear Formation in the Hamster–ICSI Assay. J. Assist. Reprod. Genet. *16*, 128–132.

Amann, R.P. (2008). The cycle of the seminiferous epithelium in humans: a need to revisit? J. Androl. *29*, 469–487.

Annamalai, J., and Namasivayam, V. (2015). Endocrine disrupting chemicals in the atmosphere: Their effects on humans and wildlife. Environ. Int. *76*, 78–97.

Avenarius, M.R., Hildebrand, M.S., Zhang, Y., Meyer, N.C., Smith, L.L.H., Kahrizi, K., Najmabadi, H., and Smith, R.J.H. (2009). Human Male Infertility Caused by Mutations in the CATSPER1 Channel Protein. Am. J. Hum. Genet. *84*, 505–510.

Avidan, N., Tamary, H., Dgany, O., Cattan, D., Pariente, A., Thulliez, M., Borot, N., Moati, L., Barthelme, A., Shalmon, L., et al. (2003). CATSPER2, a human autosomal nonsyndromic male infertility gene. Eur. J. Hum. Genet. *11*, 497–502.

Ayhan, Ö., Balkan, M., Guven, A., Hazan, R., Atar, M., Tok, A., and Tolun, A. (2014). Truncating mutations in TAF4B and ZMYND15 causing recessive azoospermia. J. Med. Genet. *51*, 239–244.

Baazeem, A., Belzile, E., Ciampi, A., Dohle, G., Jarvi, K., Salonia, A., Weidner, W., and Zini, A. (2011). Varicocele and male factor infertility treatment: a new meta-analysis and review of the role of varicocele repair. Eur. Urol. *60*, 796–808.

Baccetti, B., Collodel, G., Estenoz, M., Manca, D., Moretti, E., and Piomboni, P. (2005). Gene deletions in an infertile man with sperm fibrous sheath dysplasia. Hum. Reprod. *20*, 2790–2794.

Balhorn, R. (2007). The protamine family of sperm nuclear proteins. Genome Biol. *8*, 227.

Baltz, J.M., Williams, P.O., and Cone, R.A. (1990). Dense fibers protect mammalian sperm against damage. Biol. Reprod. *43*, 485–491.

Bamshad, M.J., Ng, S.B., Bigham, A.W., Tabor, H.K., Emond, M.J., Nickerson, D.A., and Shendure, J. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. Nat. Rev. Genet. *12*, 745–755.

Bao, J., and Bedford, M.T. (2016). Epigenetic regulation of the histone-to-protamine transition during spermiogenesis. Reproduction *151*, R55–R70.

Barman, H.K., Rasal, K.D., Chakrapani, V., Ninawe, A.S., Vengayil, D.T., Asrafuzzaman, S., Sundaray, J.K., and Jayasankar, P. (2017). Gene editing tools: state-of-the-art and the road ahead for the model and non-model fishes. Transgenic Res.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709–1712.

Bashamboo, A., Eozenou, C., Rojo, S., and McElreavey, K. (2017). Anomalies in human sex determination provide unique insights into the complex genetic interactions of early gonad development. Clin. Genet. *91*, 143–156.

Belleannée, C., Thimon, V., and Sullivan, R. (2012). Region-specific gene expression in the epididymis. Cell Tissue Res. *349*, 717–731.

Ben Khelifa, M., Zouari, R., Harbuz, R., Halouani, L., Arnoult, C., Lunardi, J., and Ray, P.F. (2011). A new AURKC mutation causing macrozoospermia: implications for human spermatogenesis and clinical diagnosis. Mol. Hum. Reprod. *17*, 762–768.

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.

Berruti, G. (2016). Towards defining an 'origin'-The case for the mammalian acrosome. Semin. Cell Dev. Biol. *59*, 46–53.

Bittman, E.L. (2016). Timing in the Testis. J. Biol. Rhythms *31*, 12–36.

Boitani, C., Di Persio, S., Esposito, V., and Vicini, E. (2016). Spermatogonial cells: mouse, monkey and man comparison. Semin. Cell Dev. Biol.

Boivin, J., Bunting, L., Collins, J.A., and Nygren, K.G. (2007). International estimates of infertility prevalence and treatment-seeking: potential need and demand for infertility medical care. Hum. Reprod. Oxf. Engl. *22*, 1506–1512.

Bolcun-Filas, E., Hall, E., Speed, R., Taggart, M., Grey, C., de Massy, B., Benavente, R., and Cooke, H.J. (2009). Mutation of the mouse Syce1 gene disrupts synapsis and suggests a link between synaptonemal complex structural components and DNA repair. PLoS Genet. *5*, e1000393.

Bower, R., Tritschler, D., Mills, K.V., Heuser, T., Nicastro, D., and Porter, M.E. (2018). DRC2/CCDC65 is a central hub for assembly of the nexin–dynein regulatory complex and other regulators of ciliary and flagellar motility. Mol. Biol. Cell *29*, 137–153.

Burgoyne, P.S., Mahadevaiah, S.K., and Turner, J.M.A. (2009). The consequences of asynapsis for mammalian meiosis. Nat. Rev. Genet. *10*, 207–216.

Burton, A. (2013). Study suggests long-term decline in French sperm quality. Environ. Health Perspect. *121*, a46.

Busada, J.T., and Geyer, C.B. (2016). The Role of Retinoic Acid (RA) in Spermatogonial Differentiation. Biol. Reprod. *94*, 10.

Camporesi, S., and Cavaliere, G. (2016). Emerging ethical perspectives in the clustered regularly interspaced short palindromic repeats genome-editing debate. Pers. Med. *13*, 575–586.

Capecchi, M.R. (1989). Altering the genome by homologous recombination. Science *244*, 1288–1292.

Capecchi, M.R. (2005). Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. Nat. Rev. Genet. *6*, 507–512.

Capozzi, A., Scambia, G., Pontecorvi, A., and Lello, S. (2015). Hyperprolactinemia: pathophysiology and therapeutic approach. Gynecol. Endocrinol. Off. J. Int. Soc. Gynecol. Endocrinol. *31*, 506–510.

Carbajal-González, B.I., Heuser, T., Fu, X., Lin, J., Smith, B.W., Mitchell, D.R., and Nicastro, D. (2013). Conserved structural motifs in the central pair complex of eukaryotic flagella. Cytoskelet. Hoboken NJ *70*, 101–120.

Carter, A.P. (2013). Crystal clear insights into how the dynein motor moves. J Cell Sci *126*, 705–713.

Chapman, K.M., Medrano, G.A., Jaichander, P., Chaudhary, J., Waits, A.E., Nobrega, M.A., Hotaling, J.M., Ober, C., and Hamra, F.K. (2015). Targeted Germline Modifications in Rats using CRISPR/Cas9 and Spermatogonial Stem Cells. Cell Rep. *10*, 1828–1835.

Chen, J., Kim, J., and Dalton, J.T. (2005). Discovery AND Therapeutic Promise OF Selective Androgen Receptor Modulators. Mol. Interv. *5*, 173–188.

Chen, Q., Deng, T., and Han, D. (2016). Testicular immunoregulation and spermatogenesis. Semin. Cell Dev. Biol.

Chung, E., and Brock, G.B. (2011). Cryptorchidism and its impact on male fertility: a state of art review of current literature. Can. Urol. Assoc. J. *5*, 210–214.

Cooper, T.G., Noonan, E., Eckardstein, S. von, Auger, J., Baker, H.W.G., Behre, H.M., Haugen, T.B., Kruger, T., Wang, C., Mbizvo, M.T., et al. (2010). World Health Organization reference values for human semen characteristics. Hum. Reprod. Update *16*, 231–245.

Coutton, C., Zouari, R., Abada, F., Khelifa, M.B., Merdassi, G., Triki, C., Escalier, D., Hesters, L., Mitchell, V., Levy, R., et al. (2012). MLPA and sequence analysis of DPY19L2 reveals point mutations causing globozoospermia. Hum. Reprod. *27*, 2549–2558.

Coutton, C., Escoffier, J., Martinez, G., Arnoult, C., and Ray, P.F. (2015). Teratozoospermia: spotlight on the main genetic actors in the human. Hum. Reprod. Update *21*, 455–485.

Dabbous, Z., and Atkin, S.L. (2018). Hyperprolactinaemia in male infertility: Clinical case scenarios. Arab J. Urol. *16*, 44–52.

Dacheux, J.-L., and Dacheux, F. (2014). New insights into epididymal function in relation to sperm maturation. Reproduction *147*, R27–R42.

Dam, A.H.D.M., Koscinski, I., Kremer, J.A.M., Moutou, C., Jaeger, A.-S., Oudakker, A.R., Tournaye, H., Charlet, N., Lagier-Tourenne, C., Van Bokhoven, H., et al. (2007). Homozygous mutation in SPATA16 is associated with male infertility in human globozoospermia. Am. J. Hum. Genet. *81*, 813–820.

Darbey, A., and Smith, L.B. (2017). Deliverable transgenics & gene therapy possibilities for the testes. Mol. Cell. Endocrinol.

Davies, B., Baumann, C., Kirchhoff, C., Ivell, R., Nubbemeyer, R., Habenicht, U.-F., Theuring, F., and Gottwald, U. (2004). Targeted deletion of the epididymal receptor HE6 results in fluid dysregulation and male infertility. Mol. Cell. Biol. *24*, 8642–8648.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature *471*, 602–607.

Dequeker, E., Stuhrmann, M., Morris, M.A., Casals, T., Castellani, C., Claustres, M., Cuppens, H., des Georges, M., Ferec, C., Macek, M., et al. (2009). Best practice guidelines for molecular genetic diagnosis of cystic fibrosis and CFTR-related disorders – updated European recommendations. Eur. J. Hum. Genet. *17*, 51–65.

Dieterich, K., Soto Rifo, R., Karen Faure, A., Hennebicq, S., Amar, B.B., Zahi, M., Perrin, J., Martinez, D., Sèle, B., Jouk, P.-S., et al. (2007). Homozygous mutation of AURKC yields large-headed polyploid spermatozoa and causes male infertility. Nat. Genet. *39*, 661–665.

Dieterich, K., Zouari, R., Harbuz, R., Vialard, F., Martinez, D., Bellayou, H., Prisant, N., Zoghmar, A., Guichaoua, M.R., Koscinski, I., et al. (2009). The Aurora Kinase C c.144delC mutation causes meiosis I arrest in men and is frequent in the North African population. Hum. Mol. Genet. *18*, 1301–1309.

Dimitriadis, F., Tsiampali, C., Chaliasos, N., Tsounapi, P., Takenaka, A., and Sofikitis, N. (2015). The Sertoli cell as the orchestra conductor of spermatogenesis: spermatogenic cells dance to the tune of testosterone. Horm. Athens Greece *14*, 479–503.

Docampo, M.J., and Hadziselimovic, F. (2015). Molecular Pathology of Cryptorchidism-Induced Infertility. Sex. Dev. *9*, 269–278.

Dolci, S., Campolo, F., and De Felici, M. (2015). Gonadal development and germ cell tumors in mouse and humans. Semin. Cell Dev. Biol. *45*, 114–123.

Domenice, S., Arnhold, I.J.P., Costa, E.M.F., and Mendonca, B.B. (2017). 46,XY Disorders of Sexual Development (MDText.com, Inc.).

Dong, F.N., Amiri-Yekta, A., Martinez, G., Saut, A., Tek, J., Stouvenel, L., Lorès, P., Karaouzène, T., Thierry-Mieg, N., Satre, V., et al. (2018). Absence of CFAP69 Causes Male Infertility due to Multiple Morphological Abnormalities of the Flagella in Human and Mouse. Am. J. Hum. Genet. *102*, 636–648.

Doudna, J.A., and Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science *346*, 1258096.

Du Plessis, S.S., Gokul, S., and Agarwal, A. (2013). Semen hyperviscosity: causes, consequences, and cures. Front. Biosci. Elite Ed. *5*, 224–231.

Dua, A., and Vaidya, S.R. (1996). Sperm motility and morphology as changing parameters linked to sperm count variations. J. Postgrad. Med. *42*, 93–96.

Dymek, E.E., and Smith, E.F. (2007). A conserved CaM- and radial spoke associated complex mediates regulation of flagellar dynein activity. J. Cell Biol. *179*, 515–526.

Eggers, S., Ohnesorg, T., and Sinclair, A. (2014). Genetic regulation of mammalian gonad development. Nat. Rev. Endocrinol. *10*, 673–683.

ElInati, E., Kuentz, P., Redin, C., Jaber, S., Meerschaut, F.V., Makarian, J., Koscinski, I., Nasr-Esfahani, M.H., Demirol, A., Gurgan, T., et al. (2012). Globozoospermia is mainly due to DPY19L2 deletion via non-allelic homologous recombination involving two recombination hotspots. Hum. Mol. Genet. dds200.

ElInati, E., Fossard, C., Okutman, O., Ghédir, H., Ibala-Romdhane, S., Ray, P.F., Saad, A., Hennebicq, S., and Viville, S. (2016). A new mutation identified in SPATA16 in two globozoospermic patients. J. Assist. Reprod. Genet. *33*, 815–820.

Escalier, D., and Touré, A. (2012). Malformations de l'appareil flagellaire du spermatozoïde impliquées dans l'infertilité chez l'homme. médecine/sciences *28*, 503–511.

Escoffier, J., Lee, H.C., Yassine, S., Zouari, R., Martinez, G., Karaouzène, T., Coutton, C., Kherraf, Z.-E., Halouani, L., Triki, C., et al. (2016). Homozygous mutation of PLCZ1 leads to defective human oocyte activation and infertility that is not rescued by the WW-binding protein PAWP. Hum. Mol. Genet. *25*, 878–891.

Esteves, S.C. (2016). Novel concepts in male factor infertility: clinical and laboratory perspectives. J. Assist. Reprod. Genet. 1–17.

Esteves, S.C., Miyaoka, R., and Agarwal, A. (2011a). Surgical treatment of male infertility in the era of intracytoplasmic sperm injection – new insights. Clinics *66*, 1463–1477.

Esteves, S.C., Miyaoka, R., and Agarwal, A. (2011b). Sperm retrieval techniques for assisted reproduction. Int. Braz J Urol Off. J. Braz. Soc. Urol. *37*, 570–583.

Estivill, X., Bancells, C., and Ramos, C. (1997). Geographic distribution and regional origin of 272 cystic fibrosis mutations in European populations. The Biomed CF Mutation Analysis Consortium. Hum. Mutat. *10*, 135–154.

Forbes, C.M., Flannigan, R., and Schlegel, P.N. (2018). Spermatogonial stem cell transplantation and male infertility: Current status and future directions. Arab J. Urol. *16*, 171–180.

França, L.R., Hess, R.A., Dufour, J.M., Hofmann, M.C., and Griswold, M.D. (2016). The Sertoli cell: one hundred fifty years of beauty and plasticity. Andrology *4*, 189–212.

Friedman, K.J., Teichtahl, H., De Kretser, D.M., Temple-Smith, P., Southwick, G.J., Silverman, L.M., Highsmith, W.E., Boucher, R.C., and Knowles, M.R. (1995). Screening Young syndrome patients for CFTR mutations. Am. J. Respir. Crit. Care Med. *152*, 1353–1357.

Gao, Y., Xiao, X., Lui, W.-Y., Lee, W.M., Mruk, D., and Cheng, C.Y. (2016). Cell polarity proteins and spermatogenesis. Semin. Cell Dev. Biol.

Garneau, J.E., Dupuis, M.-È., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature *468*, 67–71.

Gervasi M. G., and Visconti P. E. (2017). Molecular changes and signaling events occurring in spermatozoa during epididymal maturation. Andrology *5*, 204–218.

Goel, P., Rawat, J.D., Wakhlu, A., and Kureel, S.N. (2015). Undescended testicle: An update on fertility in cryptorchid men. Indian J. Med. Res. *141*, 163–171.

Goh, G., and Choi, M. (2012). Application of Whole Exome Sequencing to Identify Disease-Causing Variants in Inherited Human Diseases. Genomics Inform. *10*, 214–219.

Gonzales, G.F. (2001). Function of seminal vesicles and their role on male fertility. Asian J. Androl. *3*, 251–258.

Gore, A.C., Chappell, V.A., Fenton, S.E., Flaws, J.A., Nadal, A., Prins, G.S., Toppari, J., and Zoeller, R.T. (2015). Executive Summary to EDC-2: The Endocrine Society's Second Scientific Statement on Endocrine-Disrupting Chemicals. Endocr. Rev. *36*, 593–602.

Griswold, M.D. (2016). Spermatogenesis: The Commitment to Meiosis. Physiol. Rev. *96*, 1–17.

Gubbay, J., Vivian, N., Economou, A., Jackson, D., Goodfellow, P., and Lovell-Badge, R. (1992). Inverted repeat structure of the Sry locus in mice. Proc. Natl. Acad. Sci. U. S. A. *89*, 7953–7957.

Hamada, A.J., Esteves, S.C., and Agarwal, A. (2013). A comprehensive review of genetics and genetic testing in azoospermia. Clin. São Paulo Braz. *68 Suppl 1*, 39–60.

Harbuz, R., Zouari, R., Pierre, V., Ben Khelifa, M., Kharouf, M., Coutton, C., Merdassi, G., Abada, F., Escoffier, J., Nikas, Y., et al. (2011). A recurrent deletion of DPY19L2 causes infertility in man by blocking sperm head elongation and acrosome formation. Am. J. Hum. Genet. *88*, 351–361.

Harper, J.C., Aittomäki, K., Borry, P., Cornel, M.C., de Wert, G., Dondorp, W., Geraedts, J., Gianaroli, L., Ketterson, K., Liebaers, I., et al. (2018). Recent developments in genetics and medically assisted reproduction: from research to clinical applications. Eur. J. Hum. Genet. EJHG *26*, 12–33.

Hassold, T., and Hunt, P. (2001). To err (meiotically) is human: the genesis of human aneuploidy. Nat. Rev. Genet. *2*, 280–291.

Hawkins, J.R., Taylor, A., Goodfellow, P.N., Migeon, C.J., Smith, K.D., and Berkovitz, G.D. (1992). Evidence for increased prevalence of SRY mutations in XY females with complete rather than partial gonadal dysgenesis. Am. J. Hum. Genet. *51*, 979–984.

Heather, J.M., and Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. Genomics *107*, 1–8.

Hefner, J., Csef, H., and Seufert, J. (2009). [Kallmann syndrome. Fundamentals and two medical histories]. Nervenarzt *80*, 1169–1170, 1172–1175.

Heuser, T., Raytchev, M., Krell, J., Porter, M.E., and Nicastro, D. (2009). The dynein regulatory complex is the nexin link and a major regulatory node in cilia and flagella. J. Cell Biol. *187*, 921–933.

Heuser, T., Dymek, E.E., Lin, J., Smith, E.F., and Nicastro, D. (2012). The CSC connects three major axonemal complexes involved in dynein regulation. Mol. Biol. Cell *23*, 3143–3155.

Hotaling, J., and Carrell, D.T. (2014). Clinical genetic testing for male factor infertility: current applications and future directions. Andrology *2*, 339–350.

Hsieh, J.-T., Kuo, Y.-C., Chang, H.-C., Liu, S.-P., Chen, J.-H., and Tsai, V.F.S. (2014). The role of sympathetic and parasympathetic nerve systems on the smooth muscle of rat seminal vesicles - experimental results and speculation for physiological implication on ejaculation. Andrology *2*, 59–64.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell *157*, 1262–1278.

Huhtaniemi, I. (2015). A short evolutionary history of FSH-stimulated spermatogenesis. Horm. Athens Greece *14*, 468–478.

Hutson, J.M., Li, R., Southwell, B.R., Petersen, B.L., Thorup, J., and Cortes, D. (2013). Germ cell development in the postnatal testis: the key to prevent malignancy in cryptorchidism? Front. Endocrinol. *3*.

Iliadou, P., Tsametis, C., Kaprara, A., Papadimas, I., and Goulis, D. (2015). The Sertoli cell: Novel clinical potentiality. HORMONES *14*, 504–514.

Inhorn, M.C., and Patrizio, P. (2015). Infertility around the globe: new thinking on gender, reproductive technologies and global movements in the 21st century. Hum. Reprod. Update *21*, 411–426.

Irie, N., Weinberger, L., Tang, W.W.C., Kobayashi, T., Viukov, S., Manor, Y.S., Dietmann, S., Hanna, J.H., and Surani, M.A. (2015). SOX17 is a critical specifier of human primordial germ cell fate. Cell *160*, 253–268.

Jamsai, D., and O'Bryan, M.K. (2011). Mouse models in male fertility research. Asian J. Androl. *13*, 139–151.

Jarow, J.P., Chen, H., Rosner, W., Trentacoste, S., and Zirkin, B.R. (2001). Assessment of the Androgen Environment Within the Human Testis: Minimally Invasive Method to Obtain Intratesticular Fluid. J. Androl. *22*, 640–645.

Jensen, C.F.S., Østergren, P., Dupree, J.M., Ohl, D.A., Sønksen, J., and Fode, M. (2017). Varicocele and male infertility. Nat. Rev. Urol. *14*, 523–533.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816–821.

Jungwirth, A., Giwercman, A., Tournaye, H., Diemer, T., Kopa, Z., Dohle, G., Krausz, C., and European Association of Urology Working Group on Male Infertility (2012). European Association of Urology guidelines on Male Infertility: the 2012 update. Eur. Urol. *62*, 324–332.

Juul, A., Almstrup, K., Andersson, A.-M., Jensen, T.K., Jørgensen, N., Main, K.M., Meyts, E.R.-D., Toppari, J., and Skakkebæk, N.E. (2014). Possible fetal determinants of male infertility. Nat. Rev. Endocrinol. *10*, 553–562.

Kadam, U.S., Shelake, R.M., Chavhan, R.L., and Suprasanna, P. (2018). Concerns regarding 'off-target' activity of genome editing endonucleases. Plant Physiol. Biochem.

Kahraman, S., Akarsu, C., Cengiz, G., Dirican, K., Sözen, E., Can, B., Güven, C., and Vanderzwalmen, P. (1999). Fertility of ejaculated and testicular megalohead spermatozoa with intracytoplasmic sperm injection. Hum. Reprod. *14*, 726–730.

Kambhampati, S., Park, W., and Habtezion, A. (2014). Pharmacologic therapy for acute pancreatitis. World J. Gastroenterol. *20*, 16868–16880.

Karaca, N., Yilmaz, R., Kanten, G.E., Kervancioglu, E., Solakoglu, S., and Kervancioglu, M.E. (2014). First successful pregnancy in a globozoospermic patient having homozygous mutation in SPATA16. Fertil. Steril. *102*, 103–107.

Kierszenbaum, A.L., Rivkin, E., and Tres, L.L. (2003). Acroplaxome, an F-actin-keratin-containing plate, anchors the acrosome to the nucleus during shaping of the spermatid head. Mol. Biol. Cell *14*, 4628–4640.

Kohn, T.P., Ohlander, S.J., Jacob, J.S., Griffin, T.M., Lipshultz, L.I., and Pastuszak, A.W. (2018). The Effect of Subclinical Varicocele on Pregnancy Rates and Semen Parameters: a Systematic Review and Meta-Analysis. Curr. Urol. Rep. *19*, 53.

Koopman, P., Münsterberg, A., Capel, B., Vivian, N., and Lovell-Badge, R. (1990). Expression of a candidate sex-determining gene during mouse testis differentiation. Nature *348*, 450–452.

von Kopylow, K., and Spiess, A.-N. (2017). Human spermatogonial markers. Stem Cell Res. *25*, 300–309.

Kuo, Y.-C., Lin, Y.-H., Chen, H.-I., Wang, Y.-Y., Chiou, Y.-W., Lin, H.-H., Pan, H.-A., Wu, C.-M., Su, S.-M., Hsu, C.-C., et al. (2012). SEPT12 mutations cause male infertility with defective sperm annulus. Hum. Mutat. *33*, 710–719.

Kusumi, N., Watanabe, M., Yamada, H., Li, S.-A., Kashiwakura, Y., Matsukawa, T., Nagai, A., Nasu, Y., Kumon, H., and Takei, K. (2007). Implication of Amphiphysin 1 and Dynamin 2 in Tubulobulbar Complex Formation and Spermatid Release. Cell Struct. Funct. *32*, 101–113.

Lam, I., and Keeney, S. (2014). Mechanism and regulation of meiotic recombination initiation. Cold Spring Harb. Perspect. Biol. *7*, a016634.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

Lanfranco, F., Kamischke, A., Zitzmann, M., and Nieschlag, E. (2004). Klinefelter's syndrome. Lancet Lond. Engl. *364*, 273–283.

Larney, C., Bailey, T.L., and Koopman, P. (2014). Switching on sex: transcriptional regulation of the testis-determining gene Sry. Dev. Camb. Engl. *141*, 2195–2205.

Lavelle, G.M., White, M.M., Browne, N., McElvaney, N.G., and Reeves, E.P. (2016). Animal Models of Cystic Fibrosis Pathology: Phenotypic Parallels and Divergences. BioMed Res. Int. *2016*.

Layman, L.C. (2013). Clinical genetic testing for Kallmann syndrome. J. Clin. Endocrinol. Metab. *98*, 1860–1862.

Lee, P.A., and Houk, C.P. (2013). Cryptorchidism: Curr. Opin. Endocrinol. Diabetes Obes. *20*, 210–216.

Lejeune, H., Brosse, A., and Plotton, I. (2014). Fertilité dans le syndrome de Klinefelter. Presse Médicale *43*, 162–170.

Levine, H., Jørgensen, N., Martino-Andrade, A., Mendiola, J., Weksler-Derri, D., Mindlis, I., Pinotti, R., and Swan, S.H. (2017). Temporal trends in sperm count: a systematic review and meta-regression analysis. Hum. Reprod. Update *23*, 646–659.

Li, X., Mao, Z., Wu, M., and Xia, J. (2013). Rescuing Infertility of Pick1 Knockout Mice by Generating Testis-specific Transgenic Mice via Testicular Infection. Sci. Rep. *3*.

Liang, P., Xu, Y., Zhang, X., Ding, C., Huang, R., Zhang, Z., Lv, J., Xie, X., Chen, Y., Li, Y., et al. (2015). CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. Protein Cell *6*, 363–372.

Lindemann, C.B., and Lesich, K.A. (2016). Functional anatomy of the mammalian sperm flagellum. Cytoskelet. Hoboken NJ *73*, 652–669.

Lofrano-Porto, A., Barra, G.B., Giacomini, L.A., Nascimento, P.P., Latronico, A.C., Casulari, L.A., and da Rocha Neves, F. de A. (2007). Luteinizing hormone beta mutation and hypogonadism in men and women. N. Engl. J. Med. *357*, 897–904.

Lucas-Herald, A.K., and Bashamboo, A. (2014). Gonadal development. Endocr. Dev. *27*, 1–16.

Lutzmann, M., Grey, C., Traver, S., Ganier, O., Maya-Mendoza, A., Ranisavljevic, N., Bernex, F., Nishiyama, A., Montel, N., Gavois, E., et al. (2012). MCM8- and MCM9-deficient mice reveal gametogenesis defects and genome instability due to impaired homologous recombination. Mol. Cell *47*, 523–534.

Manara, R., Salvalaggio, A., Favaro, A., Palumbo, V., Citton, V., Elefante, A., Brunetti, A., Salle, F.D., Bonanni, G., Sinisi, A.A., et al. (2014). Brain Changes in Kallmann Syndrome. Am. J. Neuroradiol. *35*, 1700–1706.

Mao, H.-T., and Yang, W.-X. (2013). Modes of acrosin functioning during fertilization. Gene *526*, 75–79.

Maor-Sagie, E., Cinnamon, Y., Yaacov, B., Shaag, A., Goldsmidt, H., Zenvirt, S., Laufer, N., Richler, C., and Frumkin, A. (2015). Deleterious mutation in SYCE1 is associated with non-obstructive azoospermia. J. Assist. Reprod. Genet. *32*, 887–891.

Mascarenhas, M.N., Flaxman, S.R., Boerma, T., Vanderpoel, S., and Stevens, G.A. (2012). National, regional, and global trends in infertility prevalence since 1990: a systematic analysis of 277 health surveys. PLoS Med. *9*, e1001356.

Mawhinney Michael, and Mariotti Angelo (2012). Physiology, pathology and pharmacology of the male reproductive system. Periodontol. 2000 *61*, 232–251.

McKnight, K., and McKenzie, L.J. (2016). Evaluation of Infertility, Ovulation Induction and Assisted Reproduction (MDText.com, Inc.).

Mechlin, C.W., and Kogan, B.A. (2014). What lessons can be learned from testicular histology in undescended testes? Transl. Androl. Urol. *3*, 365–369.

Mehta, A., Mielnik, A., Schlegel, P.N., and Paduch, D.A. (2014). Novel methylation specific real-time PCR test for the diagnosis of Klinefelter syndrome. Asian J. Androl. *16*, 684–688.

Miao, X.-Y., and Zhang, X. (2011). Production of transgenic mice carrying the Thanatin gene by intratesticular injection. Biochem. Biophys. Res. Commun. *415*, 429–433.

Michaut, M., Tomes, C.N., De Blas, G., Yunes, R., and Mayorga, L.S. (2000). Calcium-triggered acrosomal exocytosis in human spermatozoa requires the coordinated activation of Rab3A and N-ethylmaleimide-sensitive factor. Proc. Natl. Acad. Sci. U. S. A. *97*, 9996–10001.

Miki, K., Willis, W.D., Brown, P.R., Goulding, E.H., Fulcher, K.D., and Eddy, E.M. (2002). Targeted disruption of the Akap4 gene causes defects in sperm flagellum and motility. Dev. Biol. *248*, 331–342.

Mima, M., Greenwald, D., and Ohlander, S. (2018). Environmental Toxins and Male Fertility. Curr. Urol. Rep. *19*, 50.

Mruk, D.D., and Cheng, C.Y. (2015a). The Mammalian Blood-Testis Barrier: Its Biology and Regulation. Endocr. Rev. *36*, 564–591.

Mruk, D.D., and Cheng, C.Y. (2015b). The Mammalian Blood-Testis Barrier: Its Biology and Regulation. Endocr. Rev. *36*, 564–591.

Muratorio, C., Meunier, M., Sonigo, C., Massart, P., Boitrelle, F., and Hugues, J.-N. (2013). [Varicocele and infertility: where do we stand in 2013?]. Gynecol. Obstet. Fertil. *41*, 660–666.

Nakashima, S., Ohishi, A., Takada, F., Kawamura, H., Igarashi, M., Fukami, M., and Ogata, T. (2014). Clinical and molecular studies in four patients with SRY-positive 46,XX testicular disorders of sex development: implications for variable sex development and genomic rearrangements. J. Hum. Genet. *59*, 549–553.

Namwanje, M., and Brown, C.W. (2016). Activins and Inhibins: Roles in Development, Physiology, and Disease. Cold Spring Harb. Perspect. Biol. *8*.

Nicastro, D., Schwartz, C., Pierson, J., Gaudette, R., Porter, M.E., and McIntosh, J.R. (2006). The Molecular Architecture of Axonemes Revealed by Cryoelectron Tomography. Science *313*, 944–948.

Nicholls, P.K., Harrison, C.A., Walton, K.L., McLachlan, R.I., O'Donnell, L., and Stanton, P.G. (2011). Hormonal Regulation of Sertoli Cell Micro-RNAs at Spermiation. Endocrinology *152*, 1670–1683.

Nuti, F., and Krausz, C. (2008). Gene polymorphisms/mutations relevant to abnormal spermatogenesis. Reprod. Biomed. Online *16*, 504–513.

O'Donnell, L. (2014). Mechanisms of spermiogenesis and spermiation and how they are disturbed. Spermatogenesis *4*, e979623.

O'Donnell, L., Nicholls, P.K., O'Bryan, M.K., McLachlan, R.I., and Stanton, P.G. (2011). Spermiation: The process of sperm release. Spermatogenesis *1*, 14–35.

O'Donnell, L., Stanton, P., and Kretser, D.M. de (2017). Endocrinology of the Male Reproductive System and Spermatogenesis (MDText.com, Inc.).

Oduwole, O.O., Peltoketo, H., Poliandri, A., Vengadabady, L., Chrusciel, M., Doroszko, M., Samanta, L., Owen, L., Keevil, B., Rahman, N.A., et al. (2018). Constitutively active follicle-stimulating hormone receptor enables androgen-independent spermatogenesis. J. Clin. Invest. *128*, 1787–1792.

O'Flynn O'Brien, K.L., Varghese, A.C., and Agarwal, A. (2010). The genetic causes of male factor infertility: A review. Fertil. Steril. *93*, 1–12.

Ohkura, H. (2015). Meiosis: An Overview of Key Differences from Mitosis. Cold Spring Harb. Perspect. Biol. *7*, a015859.

Okutman, O., Muller, J., Baert, Y., Serdarogullari, M., Gultomruk, M., Piton, A., Rombaut, C., Benkhalifa, M., Teletin, M., Skory, V., et al. (2015). Exome sequencing reveals a nonsense mutation in TEX15 causing spermatogenic failure in a Turkish family. Hum. Mol. Genet. *24*, 5581–5588.

Oliva, R. (2006). Protamines and male infertility. Hum. Reprod. Update *12*, 417–435.

Ombelet, W., Cooke, I., Dyer, S., Serour, G., and Devroey, P. (2008). Infertility and the provision of infertility medical services in developing countries. Hum. Reprod. Update *14*, 605–621.

O'Shaughnessy, P.J. (2014). Hormonal control of germ cell development and spermatogenesis. Semin. Cell Dev. Biol. *29*, 55–65.

O'Shaughnessy, P.J., Monteiro, A., and Abel, M. (2012). Testicular Development in Mice Lacking Receptors for Follicle Stimulating Hormone and Androgen. PLoS ONE *7*.

Page, S.L., and Hawley, R.S. (2003). Chromosome Choreography: The Meiotic Ballet. Science *301*, 785–789.

Pakravan, N., Ghaffarinia, A., Jalili, C., Riazi-Rad, F., Tajedini, M., and Mostafaie, A. (2015). Seminal vesicle fluid ameliorates autoimmune response within central nervous system. Cell. Mol. Immunol. *12*, 116–118.

Palermo G, Joris H, Devroey P, and Van Steirteghem AC (1992). Pregnancies after intracytoplasmic injection of single spermatozoon into an oocyte. Lancet *340*, 17–18.

Patat, O., Pagin, A., Siegfried, A., Mitchell, V., Chassaing, N., Faguer, S., Monteil, L., Gaston, V., Bujan, L., Courtade-Saïdi, M., et al. (2016). Truncating Mutations in the Adhesion G Protein-Coupled Receptor G2 Gene ADGRG2 Cause an X-Linked Congenital Bilateral Absence of Vas Deferens. Am. J. Hum. Genet. *99*, 437–442.

Petersen, C., and Söder, O. (2006). The Sertoli Cell – A Hormonal Target and 'Super' Nurse for Germ Cells That Determines Testicular Size. Horm. Res. Paediatr. *66*, 153–161.

Petersen, B.-S., Fredrich, B., Hoeppner, M.P., Ellinghaus, D., and Franke, A. (2017). Opportunities and challenges of whole-genome and -exome sequencing. BMC Genet. *18*, 14.

Pierre, V., Martinez, G., Coutton, C., Delaroche, J., Yassine, S., Novella, C., Pernet-Gallay, K., Hennebicq, S., Ray, P.F., and Arnoult, C. (2012). Absence of Dpy19l2, a new inner nuclear membrane protein, causes globozoospermia in mice by preventing the anchoring of the acrosome to the nucleus. Dev. Camb. Engl. *139*, 2955–2965.

Pigino, G., and Ishikawa, T. (2012). Axonemal radial spokes. Bioarchitecture *2*, 50–58.

Pigino, G., Bui, K.H., Maheshwari, A., Lupetti, P., Diener, D., and Ishikawa, T. (2011). Cryoelectron tomography of radial spokes in cilia and flagella. J. Cell Biol. *195*, 673–687.

Radpour, R., Gourabi, H., Dizaj, A.V., Holzgreve, W., and Zhong, X.Y. (2008). Genetic Investigations of CFTR Mutations in Congenital Absence of Vas Deferens, Uterus, and Vagina as a Cause of Infertility. J. Androl. *29*, 506–513.

Ramalho-Santos, J., Schatten, G., and Moreno, R.D. (2002). Control of Membrane Fusion During Spermiogenesis and the Acrosome Reaction. Biol. Reprod. *67*, 1043–1051.

Rato, L., Alves, M.G., Socorro, S., Duarte, A.I., Cavaco, J.E., and Oliveira, P.F. (2012). Metabolic regulation is important for spermatogenesis. Nat. Rev. Urol. *9*, 330–338.

Ray, P.F., Toure, A., Metzler-Guillemain, C., Mitchell, M.J., Arnoult, C., and Coutton, C. (2017). Genetic abnormalities leading to qualitative defects of sperm morphology or function. Clin. Genet. *91*, 217–232.

Roberts, A.J., Kon, T., Knight, P.J., Sutoh, K., and Burgess, S.A. (2013). Functions and mechanics of dynein motor proteins. Nat. Rev. Mol. Cell Biol. *14*, 713–726.

Rossi, P., and Dolci, S. (2013). Paracrine mechanisms involved in the control of early stages of Mammalian spermatogenesis. Front. Endocrinol. *4*, 181.

Saez, F., and Sullivan, R. (2016). Prostasomes, post-testicular sperm maturation and fertility. Front. Biosci. Landmark Ed. *21*, 1464–1473.

Sakkas, D., Ramalingam, M., Garrido, N., and Barratt, C.L.R. (2015). Sperm selection in natural conception: what can we learn from Mother Nature to improve assisted reproduction outcomes? Hum. Reprod. Update *21*, 711–726.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U. S. A. *74*, 5463–5467.

Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. Nucleic Acids Res. *39*, 9275–9282.

Schagdarsurengin, U., Paradowska, A., and Steger, K. (2012). Analysing the sperm epigenome: roles in early embryogenesis and assisted reproduction. Nat. Rev. Urol. *9*, 609–619.

Schlatt, S., and Ehmcke, J. (2014). Regulation of spermatogenesis: an evolutionary biologist's perspective. Semin. Cell Dev. Biol. *29*, 2–16.

Sedó, C.A., Rawe, V.Y., and Chemes, H.E. (2012). Acrosomal biogenesis in human globozoospermia: immunocytochemical, ultrastructural and proteomic studies. Hum. Reprod. *27*, 1912–1921.

Sha, Y., Yang, X., Mei, L., Ji, Z., Wang, X., Ding, L., Li, P., and Yang, S. (2017a). DNAH1 gene mutations and their potential association with dysplasia of the sperm fibrous sheath and infertility in the Han Chinese population. Fertil. Steril. *107*, 1312–1318.e2.

Sha, Y.-W., Xu, X., Mei, L.-B., Li, P., Su, Z.-Y., He, X.-Q., and Li, L. (2017b). A homozygous CEP135 mutation is associated with multiple morphological abnormalities of the sperm flagella (MMAF). Gene *633*, 48–53.

Singh, P., Singh, M., Cugati, G., and Singh, A.K. (2011). Hyperprolactinemia: An often missed cause of male infertility. J. Hum. Reprod. Sci. *4*, 102.

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., et al. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature *423*, 825–837.

Skerget, S., Rosenow, M.A., Petritis, K., and Karr, T.L. (2015). Sperm Proteome Maturation in the Mouse Epididymis. PloS One *10*, e0140650.

Smith, M. (2017). DNA Sequence Analysis in Clinical Medicine, Proceeding Cautiously. Front. Mol. Biosci. *4*.

Sullivan, R., and Mieusset, R. (2016). The human epididymis: its function in sperm maturation. Hum. Reprod. Update *22*, 574–587.

Svingen, T., and Koopman, P. (2013). Building the mammalian testis: origins, differentiation, and assembly of the component cell populations. Genes Dev. *27*, 2409–2426.

Sweeney, H.L., and Holzbaur, E.L.F. (2018). Motor Proteins. Cold Spring Harb. Perspect. Biol. *10*, a021931.

Tahmasbpour, E., Balasubramanian, D., and Agarwal, A. (2014). A multi-faceted approach to understanding male infertility: gene mutations, molecular defects and assisted reproductive techniques (ART). J. Assist. Reprod. Genet. *31*, 1115–1137.

Tanaka, S.S., and Nishinakamura, R. (2014). Regulation of male sex determination: genital ridge formation and Sry activation in mice. Cell. Mol. Life Sci. CMLS *71*, 4781–4802.

Tang, S., Wang, X., Li, W., Yang, X., Li, Z., Liu, W., Li, C., Zhu, Z., Wang, L., Wang, J., et al. (2017). Biallelic Mutations in CFAP43 and CFAP44 Cause Male Infertility with Multiple Morphological Abnormalities of the Sperm Flagella. Am. J. Hum. Genet. *100*, 854–864.

Tapanainen, J.S., Aittomäki, K., Min, J., Vaskivuo, T., and Huhtaniemi, I.T. (1997). Men homozygous for an inactivating mutation of the follicle-stimulating hormone (FSH) receptor gene present variable suppression of spermatogenesis and fertility. Nat. Genet. *15*, 205–206.

Teletin, M., Vernet, N., Ghyselinck, N.B., and Mark, M. (2017). Chapter Seven - Roles of Retinoic Acid in Germ Cell Differentiation. In Current Topics in Developmental Biology, D. Forrest, and S. Tsai, eds. (Academic Press), pp. 191–225.

Tenenbaum-Rakover, Y., Weinberg-Shukron, A., Renbaum, P., Lobel, O., Eideh, H., Gulsuner, S., Dahary, D., Abu-Rayyan, A., Kanaan, M., Levy-Lahad, E., et al. (2015). Minichromosome maintenance complex component 8 (MCM8) gene mutations result in primary gonadal failure. J. Med. Genet. *52*, 391–399.

Thonneau, P., Marchand, S., Tallec, A., Ferial, M.L., Ducot, B., Lansac, J., Lopes, P., Tabaste, J.M., and Spira, A. (1991). Incidence and main causes of infertility in a resident population (1,850,000) of three French regions (1988-1989). Hum. Reprod. Oxf. Engl. *6*, 811–816.

Tüttelmann, F., Simoni, M., Kliesch, S., Ledig, S., Dworniczak, B., Wieacker, P., and Röpke, A. (2011). Copy number variants in patients with severe oligozoospermia and Sertoli-cell-only syndrome. PloS One *6*, e19426.

Tüttelmann, F., Ruckert, C., and Röpke, A. (2018). Disorders of spermatogenesis. Med. Genet. *30*, 12–20.

Uhl, E.W., and Warner, N.J. (2015). Mouse Models as Predictors of Human Responses: Evolutionary Medicine. Curr. Pathobiol. Rep. *3*, 219–223.

Urbanska, P., Song, K., Joachimiak, E., Krzemien-Ojak, L., Koprowski, P., Hennessey, T., Jerka-Dziadosz, M., Fabczak, H., Gaertig, J., Nicastro, D., et al. (2015). The CSC proteins FAP61 and FAP251 build the basal substructures of radial spoke 3 in cilia. Mol. Biol. Cell *26*, 1463–1475.

Vadnais, M.L., Aghajanian, H.K., Lin, A., and Gerton, G.L. (2013). Signaling in Sperm: Toward a Molecular Understanding of the Acquisition of Sperm Motility in the Mouse Epididymis. Biol. Reprod. *89*.

Valenti, D., La Vignera, S., Condorelli, R.A., Rago, R., Barone, N., Vicari, E., and Calogero, A.E. (2013). Follicle-stimulating hormone treatment in normogonadotropic infertile men. Nat. Rev. Urol. *10*, 55–62.

Van Saen, D., Gies, I., De Schepper, J., Tournaye, H., and Goossens, E. (2012). Can pubertal boys with Klinefelter syndrome benefit from spermatogonial stem cell banking? Hum. Reprod. Oxf. Engl. *27*, 323–330.

Vander Borght, M., and Wyns, C. (2018). Fertility and infertility: Definition and epidemiology. Clin. Biochem.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. (2001). The sequence of the human genome. Science *291*, 1304–1351.

Verze, P., Cai, T., and Lorenzetti, S. (2016). The role of the prostate in male fertility, health and disease. Nat. Rev. Urol. *13*, 379–386.

Vij, S.C., Jr, E.S., and Agarwal, A. (2018). Biological therapy for non-obstructive azoospermia. Expert Opin. Biol. Ther. *18*, 19–23.

Vincensini, L., Blisnick, T., and Bastin, P. (2011). [The importance of model organisms to study cilia and flagella biology]. Biol. Aujourdhui *205*, 5–28.

Viswanadha, R., Sale, W.S., and Porter, M.E. (2017). Ciliary Motility: Regulation of Axonemal Dynein Motors. Cold Spring Harb. Perspect. Biol. *9*, a018325.

Vogl, A.W., Young, J.S., and Du, M. (2013). Chapter Eight - New Insights into Roles of Tubulobulbar Complexes in Sperm Release and Turnover of Blood-Testis Barrier. In International Review of Cell and Molecular Biology, K.W. Jeon, ed. (Academic Press), pp. 319–355.

Vogl, A.W., Du, M., Wang, X.Y., and Young, J.S. (2014). Novel clathrin/actin-based endocytic machinery associated with junction turnover in the seminiferous epithelium. Semin. Cell Dev. Biol. *30*, 55–64.

Wambergue, C., Zouari, R., Mustapha, S.F.B., Martinez, G., Devillard, F., Hennebicq, S., Satre, V., Brouillet, S., Halouani, L., Marrakchi, O., et al. (2016). Patients with multiple morphological abnormalities of the sperm flagella due to DNAH1 mutations have a good prognosis following intracytoplasmic sperm injection. Hum. Reprod. dew083.

Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., and Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell *153*, 910–918.

Wang, X., Jin, H., Han, F., Cui, Y., Chen, J., Yang, C., Zhu, P., Wang, W., Jiao, G., Wang, W., et al. (2017). Homozygous DNAH1 frameshift mutation causes multiple morphological anomalies of the sperm flagella in Chinese. Clin. Genet. *91*, 313–321.

Wei, Y.-L., and Yang, W.-X. (2018). The acroframosome-acroplaxome-manchette axis may function in sperm head shaping and male fertility. Gene *660*, 28–40.

Wu, Q.-Y., Li, N., Li, W.-W., Li, T.-F., Zhang, C., Cui, Y.-X., Xia, X.-Y., and Zhai, J.-S. (2014). Clinical, molecular and cytogenetic analysis of 46, XX testicular disorder of sex development with SRY-positive. BMC Urol. *14*, 70.

Yang, F., Eckardt, S., Leu, N.A., McLaughlin, K.J., and Wang, P.J. (2008). Mouse TEX15 is essential for DNA double-strand break repair and chromosomal synapsis during male meiosis. J. Cell Biol. *180*, 673–679.

Yang, F., Silber, S., Leu, N.A., Oates, R.D., Marszalek, J.D., Skaletsky, H., Brown, L.G., Rozen, S., Page, D.C., and Wang, P.J. (2015). TEX11 is mutated in infertile men with azoospermia and regulates genome-wide recombination rates in mouse. EMBO Mol. Med. *7*, 1198–1210.

Yatsenko, A.N., Georgiadis, A.P., Röpke, A., Berman, A.J., 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.

Yoshida, K., Kawano, N., Yoshiike, M., Yoshida, M., Iwamoto, T., and Morisawa, M. (2008). Physiological roles of semenogelin I and zinc in sperm motility and semen coagulation on ejaculation in humans. MHR Basic Sci. Reprod. Med. *14*, 151–156.

Yu, J., Chen, Z., Ni, Y., and Li, Z. (2012). CFTR mutations in men with congenital bilateral absence of the vas deferens (CBAVD): a systemic review and meta-analysis. Hum. Reprod. *27*, 25–35.

Zakrzewski, P., Lenartowski, R., Rędowicz, M.J., Miller, K.G., and Lenartowska, M. (2017). Expression and localization of myosin VI in developing mouse spermatids. Histochem. Cell Biol. *148*, 445–462.

Zegers-Hochschild, F., Adamson, G.D., Dyer, S., Racowsky, C., de Mouzon, J., Sokol, R., Rienzi, L., Sunde, A., Schmidt, L., Cooke, I.D., et al. (2017). The International Glossary on Infertility and Fertility Care, 2017. Hum. Reprod. *32*, 1786–1801.

Zhang, X., Gabriel, M.S., and Zini, A. (2006). Sperm Nuclear Histone to Protamine Ratio in Fertile and Infertile Men: Evidence of Heterogeneous Subpopulations of Spermatozoa in the Ejaculate. J. Androl. *27*, 414–420.

Zhou, J., Du, Y.-R., Qin, W.-H., Hu, Y.-G., Huang, Y.-N., Bao, L., Han, D., Mansouri, A., and Xu, G.-L. (2009). RIM-BP3 is a manchette-associated protein essential for spermiogenesis. Dev. Camb. Engl. *136*, 373–382.

Zhu, F., Gong, F., Lin, G., and Lu, G. (2013). DPY19L2 gene mutations are a major cause of globozoospermia: identification of three novel point mutations. Mol. Hum. Reprod. *19*, 395–404.

ANNEXES

4- Article Annexe 1 : PATL2 is a key actor of oocyte maturation whose invalidation causes infertility in women and mice

Marie Christou-Kent, **Zine-Eddine Kherraf**, Amir Amiri-Yekta, Emilie Le Blévec, Thomas Karaouzène, Béatrice Conne, Jessica Escoffier, Said Assou, Audrey Guttin, Emeline Lambert, Guillaume Martinez, Magalie Boguenet, Selima Fourati Ben Mustapha, Isabelle Cedrin Durnerin, Lazhar Halouani, Ouafi Marrakchi, Mounir Makni, Habib Latrous, Mahmoud Kharouf, Charles Coutton, Nicolas Thierry-Mieg, Serge Nef, Serge P Bottari, Raoudha Zouari, Jean Paul Issartel, Pierre F Ray & Christophe Arnoult.

EMBO Molecular Medicine. Avril 2018.

Contexte et principaux résultats :

Dans les pays occidentaux, la majorité des couples infertiles ont recours aux techniques d'assistance médicale à la procréation (AMP) pour obtenir une grossesse. Malgré le succès de ces techniques, la moitié des couples qui ont recours à l'AMP sortent du parcours de soin sans enfant. Une partie de ces échecs est causée par l'altération de la gamétogenèse et la production de gamètes incompatibles avec la fécondation et/ou le développement embryonnaire. Nous nous sommes intéressés à un phénotype rare d'infertilité féminine associé à l'ovulation d'ovocytes immatures qui ne peuvent maturer in vitro ou être utilisés en FIV ou en FIV-ICSI.

Nous avons analysé par séquençage exomique, une cohorte de 23 femmes d'origine nord-africaine présentant ce phénotype de blocage de la maturation ovocytaire et avons identifié un variant homozygote délétère dans le gène *PATL2* dans 26 des cas (n=6). Ce gène code pour une protéine de liaison à l'ARN. Il a été montré auparavant que l'orthologue de *PATL2* chez le *Xenopus* joue un rôle important dans la maturation des ovocytes. Cependant la fonction de ce gène n'a pas encore été explorée chez les mammifères. Nous avons étudié un modèle murin knock-out (KO) pour le gène *Patl2* et avons montré que les femelles KO homozygotes présentent une hypofertilité sévère associée à un déficit de la maturation ovocytaire. Nous avons également montré par une analyse transcriptomique des ovocytes de ces souris que l'expression de nombreux gènes est dérégulée et que parmi ces gènes, certains sont impliqués dans l'ovogenèse et le développement embryonnaire précoce. Nous avons par la suite produit par CRISPR/Cas9, un modèle murin exprimant la protéine Patl2 taguée. L'étude de l'expression spatio-temporelle de Patl2 chez ce modèle nous a permis de découvrir que cette protéine présente un profil d'expression différent des autres protéines de liaison à l'ARN. Ce profil d'expression particulier suggère que Patl2 joue un rôle spécifique dans la régulation des ARNm pendant la maturation ovocytaire chez les mammifères.

Cette découverte est très encourageante pour continuer l'exploration moléculaire et physiopathologique de ce phénotype. Elle ouvre également des perspectives intéressantes pour le développement d'approches thérapeutiques ciblées dans l'objectif de rétablir la fertilité chez ces patientes.

Published online: April 16, 2018

Research Article

TRANSPARENT PROCESS OPEN ACCESS

EMBO Molecular Medicine

PATL2 is a key actor of oocyte maturation whose invalidation causes infertility in women and mice

Marie Christou-Kent[1], Zine-Eddine Kherraf[1], Amir Amiri-Yekta[1,2,3], Emilie Le Blévec[1], Thomas Karaouzène[1], Béatrice Conne[4], Jessica Escoffier[1], Said Assou[5], Audrey Guttin[6], Emeline Lambert[1], Guillaume Martinez[1,2,7], Magalie Boguenet[1], Selima Fourati Ben Mustapha[8], Isabelle Cedrin Durnerin[9], Lazhar Halouani[8], Ouafi Marrakchi[8], Mounir Makni[8], Habib Latrous[8], Mahmoud Kharouf[8], Charles Coutton[1,2,7], Nicolas Thierry-Mieg[10], Serge Nef[4], Serge P Bottari[1], Raoudha Zouari[8], Jean Paul Issartel[6], Pierre F Ray[1,2,†] & Christophe Arnoult[1,*,†]

Abstract

The genetic causes of oocyte meiotic deficiency (OMD), a form of primary infertility characterised by the production of immature oocytes, remain largely unexplored. Using whole exome sequencing, we found that 26% of a cohort of 23 subjects with OMD harboured the same homozygous nonsense pathogenic mutation in *PATL2*, a gene encoding a putative RNA-binding protein. Using *Patl2* knockout mice, we confirmed that PATL2 deficiency disturbs oocyte maturation, since oocytes and zygotes exhibit morphological and developmental defects, respectively. PATL2's amphibian orthologue is involved in the regulation of oocyte mRNA as a partner of CPEB. However, Patl2's expression profile throughout oocyte development in mice, alongside colocalisation experiments with Cpeb1, Msy2 and Ddx6 (three oocyte RNA regulators) suggest an original role for Patl2 in mammals. Accordingly, transcriptomic analysis of oocytes from WT and *Patl2*$^{-/-}$ animals demonstrated that in the absence of Patl2, expression levels of a select number of highly relevant genes involved in oocyte maturation and early embryonic development are deregulated. In conclusion, PATL2 is a novel actor of mammalian oocyte maturation whose invalidation causes OMD in humans.

Keywords female sterility; oocyte developmental competence; oocyte maturation arrest; oocyte maturation failure; Patl2

Subject Categories Genetics, Gene Therapy & Genetic Disease; Urogenital System

DOI 10.15252/emmm.201708515 | Received 21 September 2017 | Revised 12 March 2018 | Accepted 19 March 2018

EMBO Mol Med (2018) e8515

Introduction

In humans, oocyte production is a lengthy process that begins during embryonic development and is characterised by a long diapause lasting over a decade until resumption of maturation at puberty. The quiescent oocytes, contained within primordial follicles, are arrested in the prophase of meiosis I. Periodically, a group of primordial follicles are recruited to the pool of growing follicles. The germinal vesicle (GV) oocyte and surrounding follicular cells develop in tight coordination to produce a fully grown GV oocyte within an antral follicle. This process takes around 290 days Williams & Erickson, 2012; Li & Albertini, 2013). At this stage, the oocyte is sensitive to hormonal stimulation, which causes meiosis to resume, as revealed by GV breakdown (GVBD), and extrusion of the first polar body before arresting again at the metaphase 2 (MII) stage of meiosis II. The second meiosis is completed, with exclusion of the second polar body, upon fertilisation.

Several reports have been published describing cases of infertile women whose ovaries repeatedly produce mostly/only immature oocytes. This poorly defined syndrome is known as "oocyte factor infertility" or "bad eggs syndrome" (Hartshorne *et al*, 1999; Levran

1 Genetics, Epigenetics and Therapies of Infertility, Institute for Advanced Biosciences, Inserm U1209, CNRS UMR 5309, Université Grenoble Alpes, Grenoble, France
2 UM GI-DPI, CHU de Grenoble, Grenoble, France
3 Department of Genetics, Reproductive Biomedicine Research Center, Royan Institute for Reproductive Biomedicine, ACECR, Tehran, Iran
4 Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland
5 IRMB, INSERM U1183, CHRU Montpellier, Université Montpellier, Montpellier, France
6 Grenoble Neuroscience Institute, INSERM 1216, Université Grenoble Alpes, Grenoble, France
7 UM de Génétique Chromosomique, CHU de Grenoble, Grenoble, France
8 Polyclinique les Jasmins, Centre d'Aide Médicale à la Procréation, Centre Urbain Nord, Tunis, Tunisia
9 Service de Médecine de la Reproduction, Centre Hospitalier Universitaire Jean Verdier, Assistance Publique - Hôpitaux de Paris, Bondy, France
10 Univ. Grenoble Alpes/CNRS, TIMC-IMAG, CNRS UMR 5525, Grenoble, France
*Corresponding author. Tel: +33 476 637 408; E-mail: christophe.arnoult@univ-grenoble-alpes.fr
†These authors contributed equally to this work as senior authors

et al, 2002; Beall *et al*, 2010; Hourvitz *et al*, 2010). We studied a cohort of patients who had all had at least one *in vitro* fertilisation (IVF) cycle yielding only GV, MI or atretic oocytes, and named this phenotype oocyte meiotic deficiency (OMD).

Generation of knockout mouse models has allowed the identification of several genetic variants linked to oocyte meiotic arrest at various stages. For instance, mice deficient in Cdc25b, a gene involved in cyclic AMP control, show GV arrest (Lincoln *et al*, 2002; Vaccari *et al*, 2008). Similarly, deletion of H1foo, a transcription factor for Mei1 (required for normal meiotic chromosome synapsis) and Ubb (a ubiquitin controlling the destruction of key cell cycle regulators), resulted in MI arrest (Libby *et al*, 2002; Furuya *et al*, 2007; Ryu *et al*, 2008). Finally, invalidation of Smc1b, a meiosis-specific component of the cohesin complex, causes MII arrest (Takabayashi *et al*, 2009) while deletion of Mlh3, which maintains homologous chromosome pairing at meiosis, induces mixed arrests (Lipkin *et al*, 2002). While it is tempting to suggest that mutation of any of the above-mentioned genes could cause OMD in women, none has so far been associated with this disease. Recently, heterozygous mis-sense mutations in *TUBB8*, an oocyte-specific tubulin required to form the meiotic spindle, were identified in a cohort of Chinese patients with OMD (Feng *et al*, 2016). Thus, *TUBB8* was established as the first human gene linked to OMD.

Here, we analysed 23 unrelated OMD patients from North Africa and found that six (26%) had the same homozygous truncating mutation in the *PATL2* gene, encoding a putative oocyte-specific RNA-binding protein. The role of this protein has yet to be characterised in mammals. A *TUBB8* variant was only found in a single patient in our cohort, indicating that absence of PATL2 is the main cause of OMD in this region.

Results

A homozygous truncating mutation in *PATL2* identified by whole exome and Sanger sequencing in 26% of tested subjects

We analysed a cohort of 23 infertile women presenting with OMD (Table 1). These patients responded normally to ovarian stimulation, and the number of follicles and oocytes harvested was similar to numbers for control patients. However, examination of the oocytes revealed only either GV or MI-arrested or atretic cells (identified by an irregular shape with a dark ooplasm), and a complete absence of MII oocytes.

Given that most of the patients were Tunisian and that 20–30% of marriages are consanguineous in this country, we hypothesised that infertility could be transmitted through recessive inheritance and we therefore focused on homozygous mutations. Exome analysis was performed first on samples from 15 patients. After exclusion of common variants and application of technical and biological filters (Coutton *et al*, 2018), three genes were found to be homozygously mutated in at least two subjects. Only one gene carried a homozygous variant scored as "high" and was predicted to induce loss of function by the "Variant Effect Predictor" tool (Ensembl). Interestingly, the same variant, p.Arg160Ter, c.478C>T in *PATL2* transcript ENST00000434130, was detected in five different patients. Since the orthologue of PATL2 in *Xenopus* is described as an important factor in *Xenopus* oocyte maturation (Nakamura *et al*, 2010), it was possible that this variant could be the cause of these subjects' infertility. The variant identified is expected to lead to either the production of a truncated protein (Fig 1A) or a complete absence of expression due to possible nonsense-mediated mRNA decay. The truncated protein would contain less than one-third of the complete amino acid sequence, in particular lacking the topoisomerase II-associated protein PAT1 domain. This domain, common to all Pat1 proteins, has been shown to be necessary for its paralogue, PATL1, to function through interaction with its partners (Braun *et al*, 2010) (Fig 1A). Because we did not have access to the relevant biological material (patients' ovaries), it was impossible to assess RNA decay in the presence of this mutation.

The presence of the genetic variant was confirmed by Sanger sequencing for the five mutated patients (Fig 1B). This variant was also identified in a heterozygous state in five out of 148,732 alleles (rs548527219) in the Genome Aggregation Database (gnomAD). This rate corresponds to a very low frequency of 0.003362%, compatible with recessive transmission of a genetic disease. Sanger sequencing of *PATL2* coding sequences was then performed on another eight OMD subjects. An additional patient was identified with the same homozygous mutation, increasing the final number to six out of 23 subjects analysed (26%) carrying the *PATL2* p.Arg160Ter variant.

To complete the analysis of the cohort, WES was performed on the newly recruited patients ($n = 8$) except for the subject harbouring the PATL2 mutation. WES analysis was therefore performed on a total of 22 subjects. From these data, we also sought *TUBB8* heterozygous mutations, which have also been described to induce OMD (Feng *et al*, 2016). One deleterious heterozygous variant (ENST00000309812.4:c.363_366del, ENSP00000311042.4:p.Lys122ArgfsTer13) was identified in patient P16 (Table 1), which could be the reason for this patient's infertility.

Overall, in this cohort, six out of 23 subjects analysed (26%) were observed to carry the *PATL2* p.Arg160Ter mutation, and one patient presented a new *TUBB8* variant (4.5%, 1/22), the pathogenicity of which remains to be confirmed. In our cohort, we compared patient characteristics between subjects with a *PATL2* mutation or presenting no *PATL2* mutation (Fig EV1). Although both groups were of similar ages at the time of analysis, and the numbers of oocytes retrieved were comparable, the two groups were clearly distinct in terms of the type of oocyte arrest. Oocytes from *PATL2* patients were mainly arrested at the GV stage, whereas oocytes from non-*PATL2* patients were generally arrested at the MI stage (Fig EV1).

During evaluation of our data, PATL2 gene mutations were also reported to be associated with OMD in two Asian studies based on cohorts from China and Saudi Arabia (Chen *et al*, 2017; Maddirevula *et al*, 2017). These findings support the causality of our *PATL2* variant and indicate a wide global spread for *PATL2*-dependent OMD.

Patl2 is not expressed in the hypothalamic–pituitary–gonadal axis in mice

x-pat1a, the *Xenopus* orthologue of *PATL2*, has been reported to be specifically expressed in growing oocytes (Marnef *et al*, 2010;

Table 1. Medical history, laboratory investigations and oocyte collection outcomes for patients presenting with OMD.

	Origin	Age (years)	Number of oocytes collected					FSH U/L	LH U/L	TSH U/L	Prolactin µg/L	Menst.	Comments
			GV	MI	MII	At.	Tot.						
Patients with *PATL2* mutation													
P1	Tunisia	35	4	0	0	1	5			1.36			
		34					2						Medical records not available
		34					8						Medical records not available
P2	Tunisia	28	9	0	0	11	20					YES	
		28.9	15	0	0	4	19						
P3	Tunisia	24	11	0	0	5	16	10.31	3.54		22.28		1 GV maturated to M1 *in vitro*
P4	Tunisia	34.28	8	0	0	2	10			1.07	23	YES	
P5	Arab	41	2	0	0	2	4	9.39	6.1	2.3	12.9	YES	
		42	3	1	0	1	5						
P6	Mauritania	36	2	10	0	4	16	3.01	3.38	2.88	25	YES	Cytoplasmic vacuoles in MI oocyte
		36.8	0	0	0	5	5						
Patients without *PATL2* mutation													
P7	Algeria	37	2	0	0	2	4	10.1	9.15	2.3		YES	
P8	Algeria	32	0	4	0	0	4						
		32	0	2	0	0	2						
P9	Tunisia	32	0	0	0	8	8			3.73			First cousin couple
P10	Tunisia	37	0	2	0	3	5						
		37	2	3	0	3	8						
P11	Tunisia	38.9	0	3	2	2	7	8.49	3.42	1.17	17.05		
P12	Libya	28	2	15	0	0	17	1.8			9.6		
P13	Arab	33	0	3	2	7	12						
		37	3	1	0	11	15						
P14	Tunisia	26	0	0	0	7	7					YES	
P15	Arab	38	0	0	0	4	4						
		39	0	0	0	5	5						
P16	Arab	33	0	5	2	2	9	4.65	2.71		14.25		Heterozygous mutation in *TUBB8*
P17	Arab	34	0	0	0	0	0					YES	
P18	Arab-FR	27	0	8	0	4	12						
P19	Arab-FR	24	0	3	0	0	3						
P20		29	0	0	0	7	7						
P21	Tunisia	29	0	0	0	13	13					YES	
P22	Tunisia	39	0	10	0	0	10						
P23	Tunisia-FR	No data available											No fertilisation
	Mean P1-P23	33.42	2.17	2.50	0.21	3.90	8.45						
	Mean P1-P6	34.00	6.00	1.38	0	3.89	10.00						

Table 1 (continued)

	Origin	Age (years)	Number of oocytes collected					FSH U/L	LH U/L	TSH U/L	Prolactin µg/L	Menst.	Comments
			GV	MI	MII	At.	Tot.						
	Mean P7-P23	33.1	0.45	2.95	0.3	3.9	7.6						
	Control cohort values (n = 238)	34.4	2.2	1.8	6	2.3	9.1	< 10.2	< 16.9	0.5–5	2–20		

Normal values correspond to couples where the male suffers from azoospermia or teratozoospermia (n = 234). Arab-FR = French of Arab origin, At. = atretic, Tot. = total, Menst. = menstruation.

Nakamura *et al*, 2010), and analysis of publicly accessible data banks shows that *PATL2* is also expressed at high levels in both human and mouse oocytes (Appendix Fig S1), indicating an important role for PATL2 in female gametogenesis. It should be noted that PATL2 expression is very low in human follicular cells (Appendix Fig S1), suggesting that the maturation defect is of oocyte rather than follicular origin. It also appears that PATL2 is expressed at low levels in a number of other tissues (Appendix Fig S1). We therefore wondered whether an element of the infertility phenotype could be caused by alteration of the hypothalamic–pituitary–gonadal axis. To address this question, we performed comparative Western blots on extracts from GV oocytes, hypothalamus and pituitary glands from Patl2-HA-tagged mice created using CrispR-cas9 technology. Whereas a clear and specific signal is observed for oocyte extracts from PATL2-HA females, no signal was observed in extracts from the hypothalamus or pituitary gland indicating that the direct control of the hypothalamus/pituitary gland on oocyte maturation is not altered in mice (Appendix Fig S2). Since our *PATL2* patients exhibited normal hormone levels (when data were available, Table 1) and reported regular menstrual cycles, these results taken together suggest that the human infertility phenotype is purely due to an oocyte defect.

Absence of Patl2 modifies the number of the primordial follicles at 26 dpp but not at 12 dpp

To decipher the molecular pathogenesis of the phenotype observed in our *PATL2* patients, we assessed the reproductive phenotype of Patl2-deficient mice ($Patl2^{-/-}$). The gene was invalidated by insertion of a LacZ cassette and deletion of exon 7, inducing a downstream translational frameshift (Appendix Fig S3). The putative transcript produced from this construct would consist of the first 102 amino acids (out of 529) tethered to β-galactosidase. Even if a protein product was generated from the modified Patl2 gene, it would not contain the topoisomerase II-associated protein (PAT1) domain and would therefore not be functional.

Initially, we performed a comparative histological study of control and $Patl2^{-/-}$ ovaries at 12 and 26 days postpartum (dpp). At 12 dpp, ovary sections from control and $Patl2^{-/-}$ females revealed no differences in the mean number of primordial, primary and secondary follicles (Appendix Fig S4). These data indicate that Patl2 plays a marginal role during the development of embryonic ovaries. At 26 dpp, a similar number of primary and secondary follicles were also observed in ovary sections (Fig EV2A and B). However, there was an unexpected increase in the number of primordial oocytes per section in $Patl2^{-/-}$ ovaries (Fig EV2B). To take the range of secondary follicle sizes into account and to assess follicle growth, we compared histograms plotting the amplitude of follicle diameter for secondary follicles between $Patl2^{-/-}$ and control animals and found no difference (Fig EV2C).

Patl2 may play a major role in oocyte growth: it is expressed in oocytes from the primary follicle stage and is less abundant from the late GV stage

We used PATL2-HA mice to characterise the function of Patl2 in mouse oocytes. The HA tag was selected for its small size (nine amino acids) making it is less likely to induce tag-dependent relocalisation. Homozygous Patl2-HA females' fertility parameters are well within the normal range for this strain in natural mating (first litters 24 days after crossing with males and litter sizes of 7 and 8, n = 2), implying that the HA tag does not impair the function of Patl2. We performed IF and confocal microscopy to quantify the expression of Patl2-HA in the different stages of GV oocytes, in MII oocytes and in ovary sections from tagged mice. The specificity of the fluorescence signal was first validated by immunostaining WT ovary sections alongside Patl2-HA sections using the same anti-HA antibody. No signal was observed in any oocyte in control sections (Appendix Fig S5). In Patl2-HA ovary sections, primordial follicle oocytes produced no detectable fluorescence signal (Fig 2). Signal intensity increased, interpreted as an increase in protein concentration, in oocytes from primary to pre-antral secondary follicles, becoming weaker in oocytes contained in tertiary/antral follicles. It should be noted that fluorescence intensity is not a direct measure of total protein quantity, since oocyte volume increases as follicular stages progress. Because the volume of antral follicle oocytes is larger than that of pre-antral follicle oocytes (approx. 70 versus 50 µm), the total quantity of Patl2 in antral follicle oocytes remains greater, indicating that production of Patl2 is continuous during oocyte growth (Fig 2A–C).

We next compared Patl2-HA immunostained GV oocytes at various stages, and MII oocytes obtained after hormonal stimulation. GV oocytes from pre-antral and antral follicles were obtained by collagenase treatment and ovarian puncture, respectively. In agreement with our results on ovarian sections, the strongest fluorescence signal was observed in pre-antral (secondary) follicle oocytes. GV oocytes from antral follicles can be divided into two categories

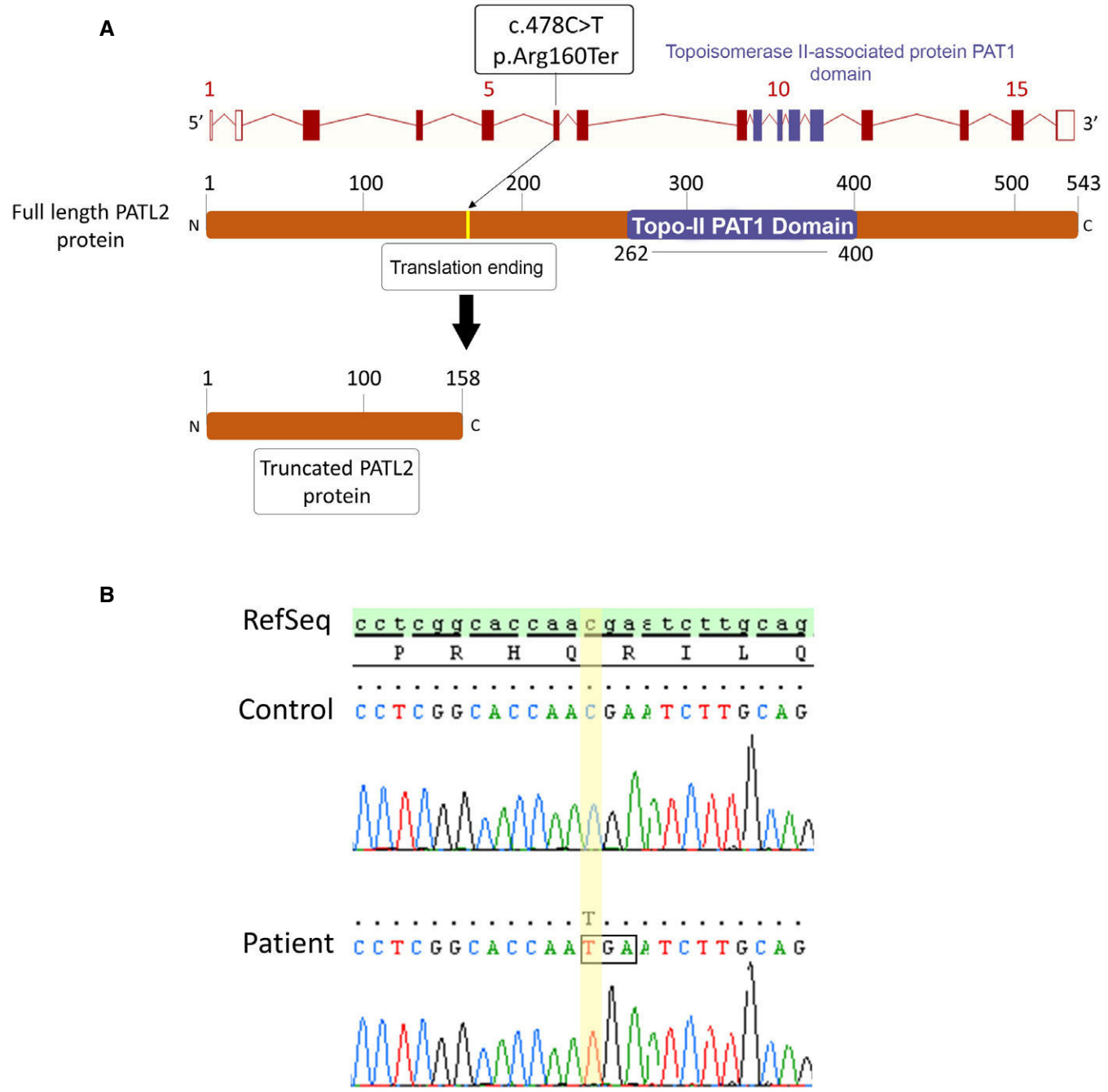


Figure 1. Identification of a truncating mutation in *PATL2*.

A Location of the *PATL2* mutation in the intron–exon structure and in a representation of the corresponding amino acid sequence. The variant identified, homozygous in the six patients, is located in exon 6 and creates a STOP codon, ending translation and producing a truncated 158-amino acid (aa) protein instead of the full-length 543 aa, and lacking the essential PAT1 (topoisomerase II-associated protein PAT1) domain.

B Electropherograms of Sanger sequencing for patients harbouring *PATL2* mutations compared to reference sequence.

based on the nuclear distribution of their chromatin: the non-surrounded nucleolus (NSN) to surrounded nucleolus (SN) conformational change occurs in the final stages of GV oocyte development and correlates with transcriptional arrest (De La Fuente, 2006). Stronger Patl2-HA fluorescence was observed in NSN than in SN GV oocytes (Fig 2D). PATL2-HA was also detected in MII oocytes at a level comparable to that in SN GV oocytes (Fig 2D).

During oocyte growth, a large quantity of stable mRNA necessary for growth and maturation accumulates within the oocyte. Up to 30% of this mRNA is translationally repressed until meiotic

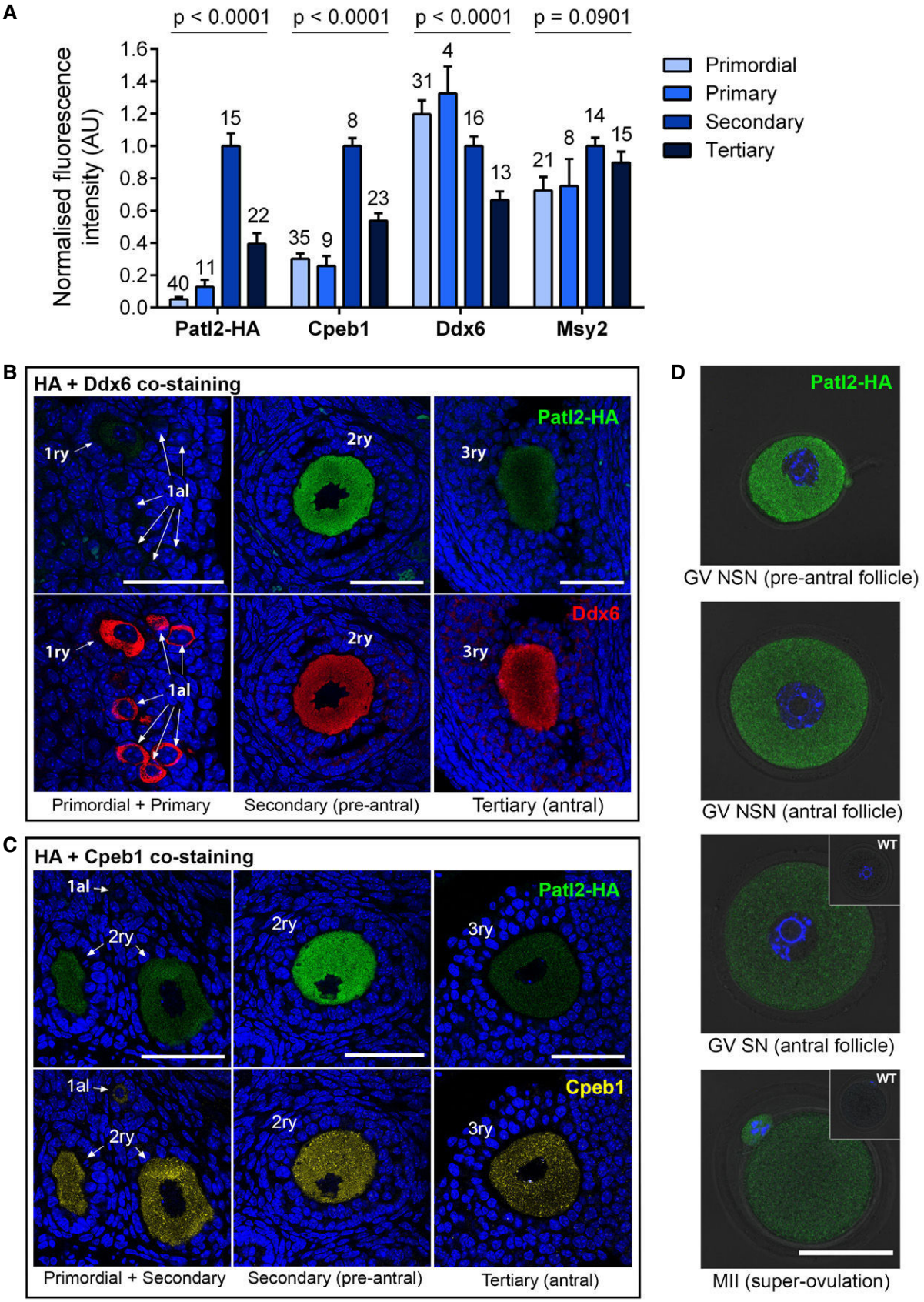


Figure 2.

Figure 2. Patl2, Cpeb1, Ddx6 and Msy2 expression profiles during oocyte growth and meiotic maturation.

A Ovary sections (3 μm thick) from Patl2-HA homozygous mice were co-stained with antibodies against the HA tag and either Cpeb1, Ddx6 or Msy2. Normalised mean fluorescence intensity at different follicle stages (as indicated) was measured using confocal microscopy. The mean fluorescence intensity of secondary follicle oocytes was used to normalise intensities for each protein to take into account variations in overall staining intensity between slides and obtain comparable values. Numbers above bars correspond to the size of the sample. Data are presented as mean ± SEM. Statistical differences were determined based on ANOVA test, *P*-value as indicated.

B Variation in fluorescence intensity for Patl2 and Ddx6 during oocyte growth. Confocal images obtained from the same section of an ovary from a Patl2-HA female, co-stained with antibodies against HA tag and Ddx6. Patl2 staining is not detectable in primordial oocytes, barely detectable in primary oocytes, and has a maximum intensity in secondary oocytes. Note that Ddx6 staining is strong in primordial oocytes. Primordial (1al), primary (1ry), secondary (2ry) follicles are indicated. Sections were counterstained with Hoechst to reveal the nucleus. Scale bar = 50 μm.

C Variation in fluorescence intensity for Patl2 and Ddx6 during oocyte growth. Confocal images obtained from the same section of an ovary from a Patl2-HA female, co-stained with antibodies against HA tag and Cpeb1. Note that Cpeb1 is detectable in primordial oocytes and that its staining is more punctiform than Patl2 staining, with numerous foci observable at all stages. Primordial (1al), primary (1ry), secondary (2ry) follicles are indicated. Sections were counterstained with Hoechst to reveal the nucleus. Scale bar = 50 μm.

D Comparative Patl2 staining of GV oocytes from pre-antral, antral with NSN (non-surrounded nucleolus) chromatin, antral with SN (surrounded nucleolus) chromatin and MII oocytes from Patl2-HA-tagged mice. GV oocytes were isolated by ovarian puncture, and MII oocytes were collected in the oviduct from stimulated Patl2-HA females. After fixation, oocytes were stained with anti-HA antibody and observed by confocal microscopy. In GV SN and MII oocytes, insets correspond to WT oocytes at the same developmental stage, showing no fluorescent staining. Scale bar = 50 μm.

maturation or after fertilisation. RNA processing involves very large complexes of ribonucleoproteins (RNPs), which are involved in the storage, processing, regulation and/or degradation of mRNA, and whose function often depends on the protein's phosphorylation status and the protein composition of the RNP. Clusters of RNP complexes are known as P-bodies. In *Xenopus* oocytes, x-Pat1a binds to the cytoplasmic polyadenylation element binding complex (CPEB), a central RNP involved in RNA translation/storage (Marnef *et al*, 2010; Nakamura *et al*, 2010). We therefore wondered whether other proteins known to be RNP components expressed in mouse oocytes, such as Cpeb1, Msy2 (thought to be x-Pat1a partners) and Ddx6 (a crucial P-body component) presented similar patterns of expression/abundance to Patl2 during oocyte growth (Flemr *et al*, 2010; Medvedev *et al*, 2011). To answer this question, we quantified these proteins in oocytes at the different follicular stages by IF and confocal microscopy (Fig 2A). Cpeb1 showed a similar, but not identical, pattern of intensities to that observed for Patl2. The main difference was that Cpeb1 is expressed in primordial follicle oocytes (in which Patl2 is undetectable). As for Patl2, the intensity was highest in secondary follicle oocytes and weakened in tertiary follicle oocytes (Fig 2B). Unlike Patl2, Msy2 and Ddx6 were abundantly expressed in primordial follicle oocytes, and their fluorescence intensity varied little throughout oocyte growth (Fig 2C and also Fig EV2 for Msy2).

We next assessed possible colocalisation between Patl2 and Cpeb1, Msy2 and Ddx6 (Fig 3). These experiments were performed on ovarian sections of secondary/pre-antral follicles, where the strongest Patl2 signal was recorded. Cpeb1 and ddx6 present a clear punctiform signal (Fig 3: A4, B4), unlike Msy2, for which the signal is more homogenous (Fig 3: C4). The Patl2-HA signal can be described as a homogenous scattering of small dots (Fig 3: A3, B3, C3). The dots are clearly smaller in size than those corresponding to Cpeb1 and Ddx6 staining, and no obvious colocalisation between Patl2 and any of the three proteins was observed (Fig 3: A2, B2, C2).

Subfertility in *Patl2*$^{-/-}$ female mice is due to compromised oocyte maturation and poor developmental competence of oocytes and embryos

We next assessed fertility in *Patl2*$^{-/-}$ animals by crossing them with WT animals and counting the number of live pups per litter, the total number of live pups born and the number of litters per month over a 6-month period. *Patl2*$^{-/-}$ females exhibited severe subfertility: the number of pups per litter dropped from 7.3 ± 0.8 (n = 14 litters for three females) for WT to 2.3 ± 0.4 (n = 7 litters for three females) for *Patl2*$^{-/-}$ mice (Fig 4A and B), and both the total number of pups and of litters per month per female were reduced (Fig 4C and D). Conversely, *Patl2*$^{-/-}$ males showed normal fertility: no difference in litter size was observed compared to WT (7.6 ± 0.7, n = 17 litters for five WT males and 7.6 ± 0.5, n = 28 litters for five *Patl2*$^{-/-}$ males) (Fig 4E).

Next, ovarian stimulation was performed and oocytes were collected for morphological and IVF studies. Stimulation with pregnant mare serum gonadotropin (PMSG) was used to obtain GV oocytes with well-defined nucleoli from both WT and *Patl2*$^{-/-}$ females. However, GV oocytes from *Patl2*$^{-/-}$ mice were smaller in diameter than those from WT mice, suggesting that the absence of Patl2 affected oocyte growth (Fig EV3A and B). Full stimulation (PMSG + hCG) produced a comparable number of oocytes in *Patl2*$^{-/-}$ and WT mice (Fig EV3C). This result concurs with observations that patients harbouring *PATL2* mutations produce a comparable number of oocytes to control patients (Table 1). *Patl2*$^{-/-}$ mice produced MII stage oocytes (Fig EV3D), identified by the presence of the first polar body (PB1), indicating that the phenotype is not as severe in mice as in humans, where no MII oocytes were produced (Table 1). As for GV oocytes, MII *Patl2*$^{-/-}$ oocytes were smaller in diameter than control MII oocytes (Fig EV3E), indicating that oocyte meiotic maturation as well as oocyte growth was impaired in the absence of Patl2. This finding was corroborated by the increased percentage of oocytes released at stages before MII: 26% for WT and 45% for *Patl2*$^{-/-}$ (Fig 5A). These oocytes were probably blocked in metaphase I (MI), as indicated by the absence of PB1 (Fig 5B). Notably, many apparent MI-arrested oocytes presented misaligned chromosomes and abundant cytoplasmic asters (Fig 5B). A significant increase in morphological defects such as abnormal spindle morphology, misalignment of chromosomes on the spindle and numerous cytoplasmic asters was also observed in *Patl2*$^{-/-}$ MII oocytes (Fig 5C and D). Thus, both women carrying a *PATL2* mutation and *Patl2*$^{-/-}$ mice exhibit oocyte maturation defects.

Next, the developmental competence of the oocytes collected was challenged in IVF experiments. We chose not to denude oocytes for these experiments since removal of cumulus cells has a negative impact on fertilisation. For WT females, the percentage of eggs

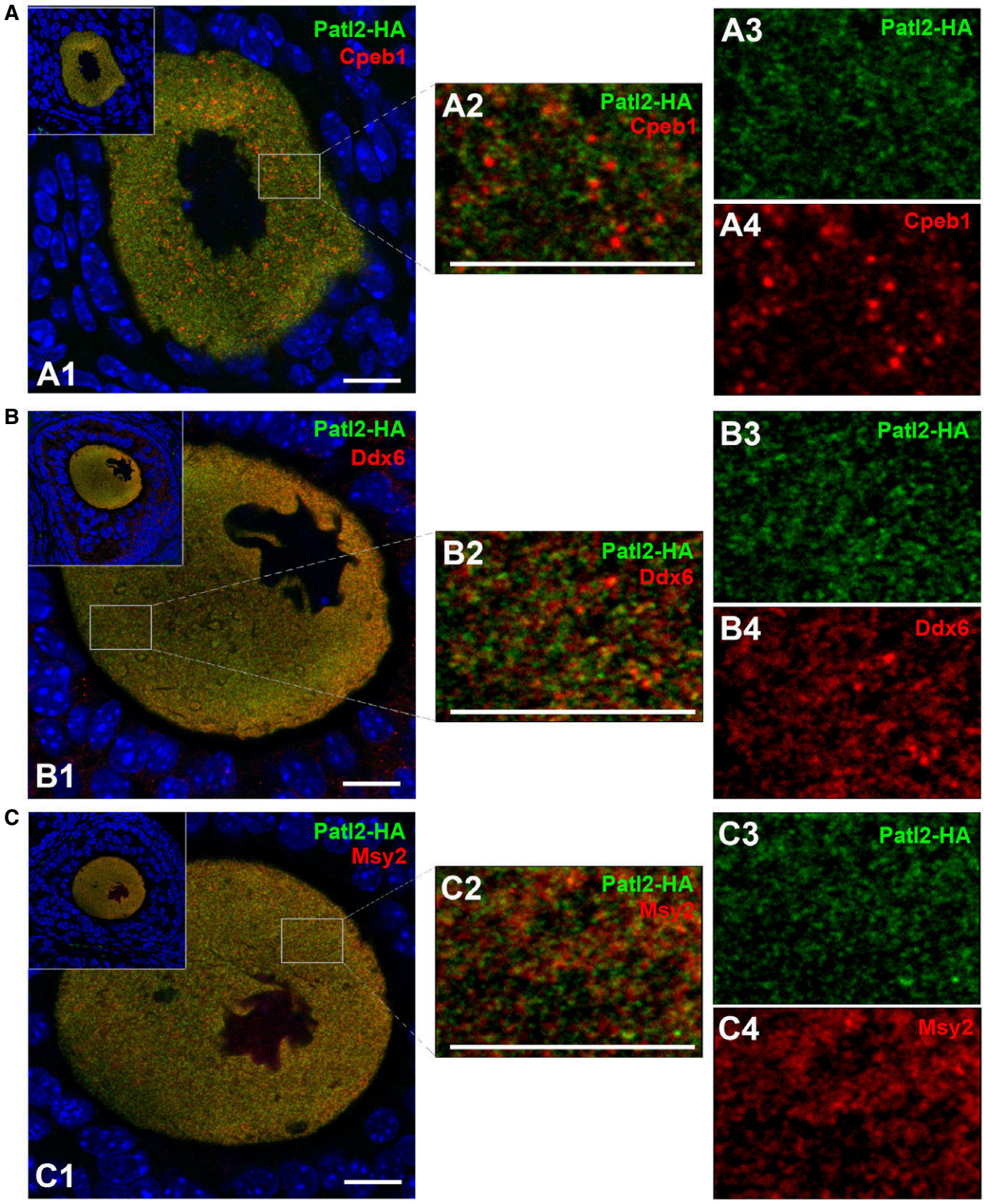


Figure 3. Analysis of possible colocalisation of Patl2 with Cpeb1, Ddx6 and Msy2.

A Confocal image of secondary follicle oocytes from Patl2-HA homozygous females co-stained with antibodies against HA tag (green) and Cpeb1 (red) and counterstained with Hoechst to reveal the nucleus (A1). Insets show the follicular environment of the oocyte studied. White rectangles indicate the zones of enlargement, shown on the right (A2). This image corresponds to a merge of Patl2 (A3) and Cpeb1 (A4) signals. Scale bars = 10 μm.

B Similar experiments performed with secondary follicle oocytes from Patl2-HA homozygous females co-stained with antibodies against HA tag (green) and Ddx6 (red). Scale bars = 10 μm.

C Similar experiments performed with secondary follicle oocytes from Patl2-HA homozygous females co-stained with antibodies against HA tag (green) and Msy2 (red). Scale bars = 10 μm.

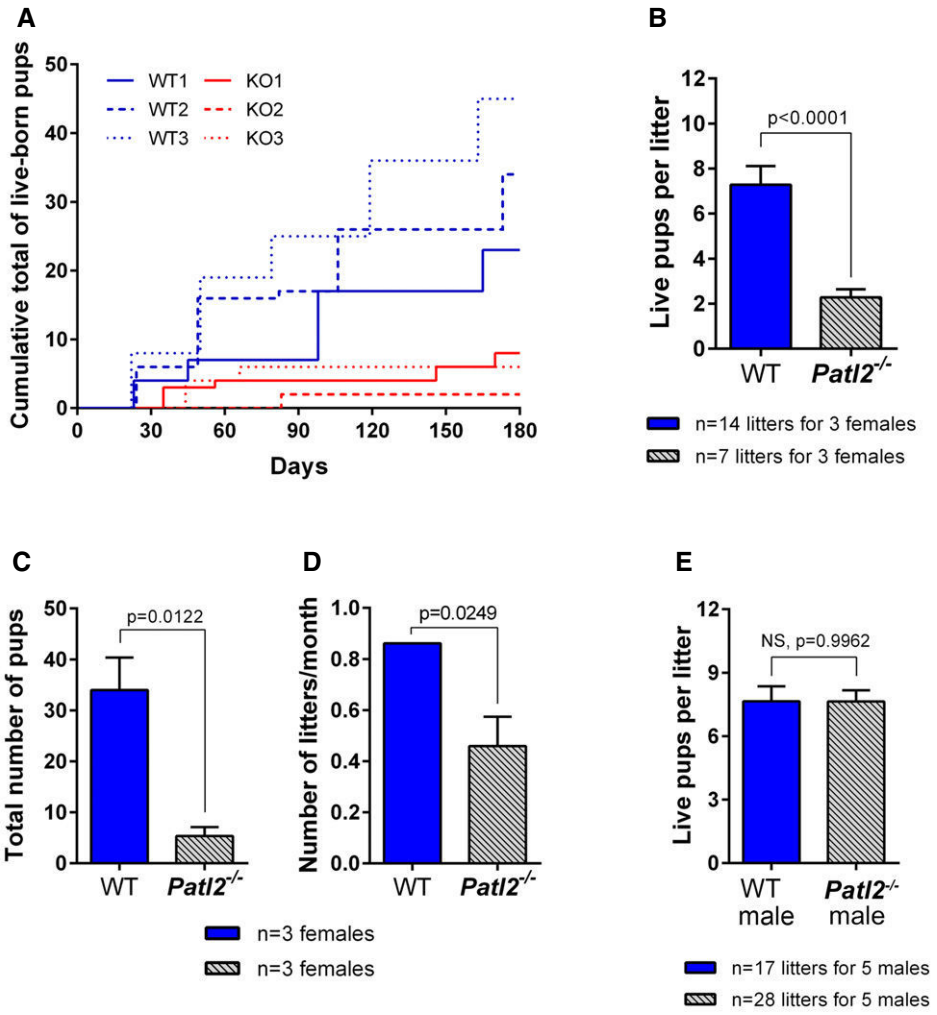


Figure 4. *Patl2* knockout (*Patl2*$^{-/-}$) females exhibit a severe subfertility phenotype when mated with WT males, whereas *Patl2*$^{-/-}$ males are fertile.

A Comparative accumulation of live pups over a period of 6 months from three WT and three *Patl2*$^{-/-}$ females crossed with WT males shows severe hypofertility of *Patl2*$^{-/-}$ females.

B Histograms showing the number of pups per litter (mean ± SEM) obtained by crossing three WT (n = 14 total litters) and three *Patl2*$^{-/-}$ females (n = 7 total litters) with WT males. Statistical test: two-tailed unpaired t-test with Welch's correction.

C Total number (mean ± SEM) of pups produced by a WT or *Patl2*$^{-/-}$ female over a 6-month period (n = 3 per genotype). Statistical test used: two-tailed unpaired t-test.

D The number of litters per month (mean ± SEM) is significantly decreased for *Patl2*$^{-/-}$ versus WT females. Statistical test: two-tailed unpaired t-test.

E *Patl2*$^{-/-}$ males produce comparable litter sizes to WT males when mated with WT females (mean ± SEM). Statistical test: two-tailed unpaired t-test.

reaching the two-cell stage was 67.7% ± 8.1 (n = 5 experiments, 10 females) (Fig 6A). Given that only 74% of ovulated oocytes can be assumed to be at the MII stage (Fig 5A), this proportion translates to 90% of WT MII oocytes reaching the two-cell stage. In contrast, for *Patl2*$^{-/-}$ females the percentage of eggs reaching the two-cell stage dropped to 36.4% ± 6.4, which translates to 65% success if we consider that only 55% of ovulated oocytes from *Patl2*$^{-/-}$ females are MII oocytes (Fig 5A). The IVF outcomes were therefore significantly altered in *Patl2*$^{-/-}$ females, indicating compromised developmental competence for *Patl2*$^{-/-}$ oocytes. This decrease in numbers of two-cell *Patl2*$^{-/-}$ embryos correlates with the abnormal development of *Patl2*$^{-/-}$ zygotes, which exhibited numerous defects, including delayed pronucleus formation, absence of sperm DNA decondensation and/or polyspermy. In contrast, almost all fertilised WT zygotes contained two pronuclei (2PN) (Fig 6B–D). Finally, the reduced developmental competence of *Patl2*$^{-/-}$ eggs and their altered fertilisation also severely affected pre-implantation development since only 27.2% ± 5.1 (n = 4 experiment, eight females) of two-cell embryos generated with *Patl2*$^{-/-}$ eggs reached the blastocyst stage in contrast to 87.1% ± 5.6 with WT eggs (Fig 6A).

Absence of Patl2 significantly alters the transcriptome of GV and MII oocytes

Since x-Pat1a, the *Xenopus* orthologue of *PATL2*, is a RNA-binding protein (Marnef *et al*, 2010; Nakamura *et al*, 2010), we next

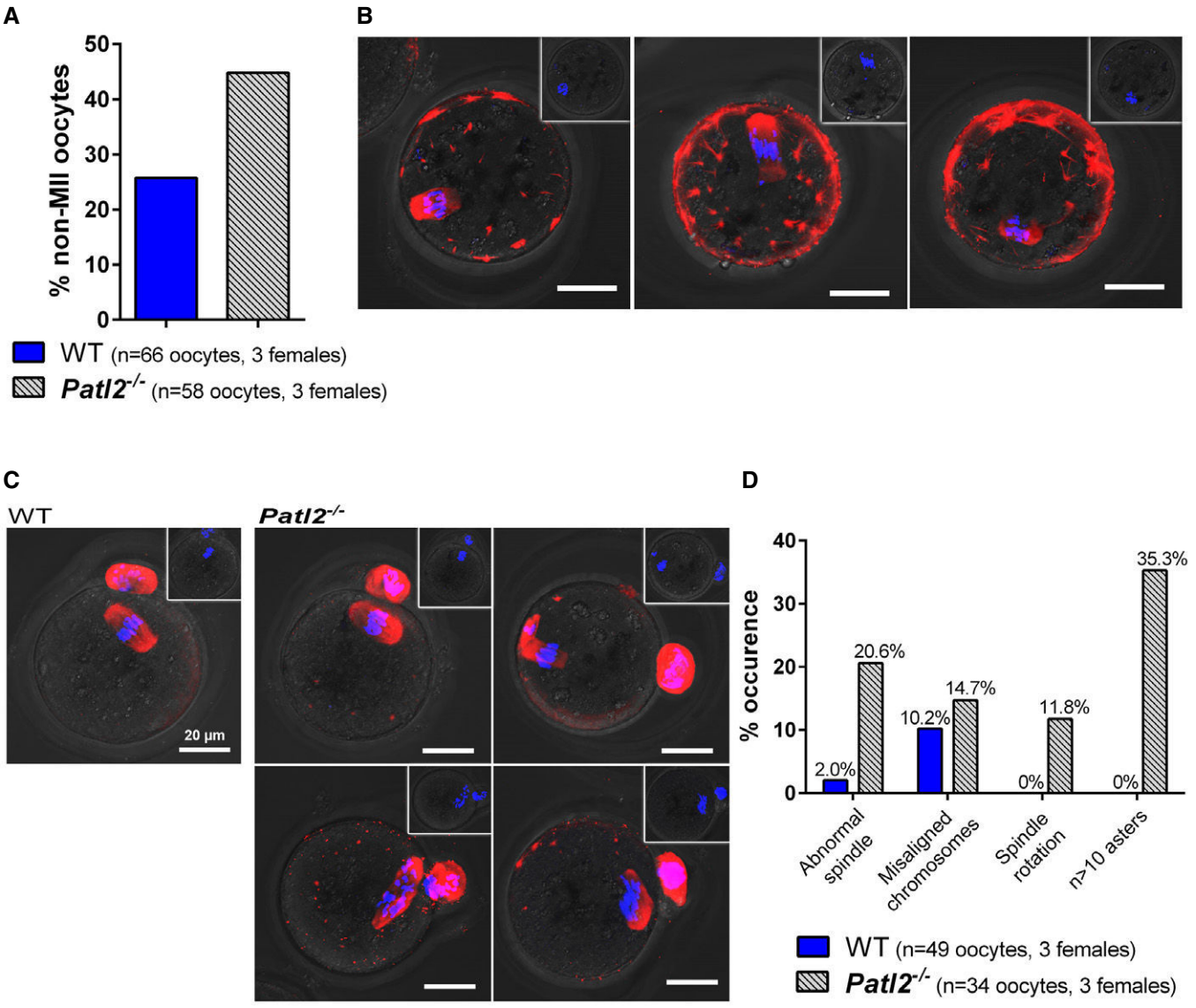


Figure 5. Infertility of *Patl2*-deficient female mice is due to oocyte maturation defects.

A Oocytes collected after ovarian stimulation were labelled with a tubulin antibody (red) and counterstained with DAPI to reveal DNA (blue). An increase in non-MII oocytes (MI arrest, absence of PB) after full ovarian stimulation was observed in *Patl2*$^{-/-}$ mice (n = 3 females per genotype).

B IF images of tubulin-stained *Patl2*$^{-/-}$ oocytes arrested at MI stage showing various defects such as irregular spindle shape and abnormal chromosome distribution. Scale bar = 20 µm. Inset in each panel shows overlay of phase contrast image and Hoechst staining of the corresponding oocyte. No polar bodies were observed for MI oocytes.

C IF images of tubulin-stained WT and *Patl2*$^{-/-}$ MII oocytes, as evidenced by PB1. In control MII oocytes, stack projections of confocal images show that the spindle was symmetric and the chromosomes distributed in the middle of the spindle. In contrast, in *Patl2*$^{-/-}$ MII oocytes various defects were observed such as irregular spindle shape, spindle rotation and numerous cytoplasmic asters. Slightly greater numbers of oocytes with abnormal chromosome distribution were also observed. Scale bar = 20 µm. Inset in each panel shows overlay of phase contrast image and Hoechst staining of the corresponding oocyte. One polar bodies was observed for MII oocytes.

D Histograms quantifying the % defects observed in *Patl2*$^{-/-}$ MII oocytes.

assessed how the absence of Patl2 affected the oocyte transcriptome during oocyte maturation in mice. To do so, global gene expression analysis was performed on oocytes collected at GV and MII stages from WT and *Patl2*$^{-/-}$ females.

Expression levels for nearly 66,000 transcripts were measured across the different oocyte groups using Affymetrix microarrays. First, we verified that oocyte RNA purification was not contaminated by RNA from follicular cells, by comparing expression levels of genes specific to follicular cells and oocytes, respectively (Appendix Fig S6A). The absence of exon 7 transcription in *Patl2*$^{-/-}$ oocyte extracts was also verified in the microarray data (Appendix Fig S6B). The results of this analysis revealed no

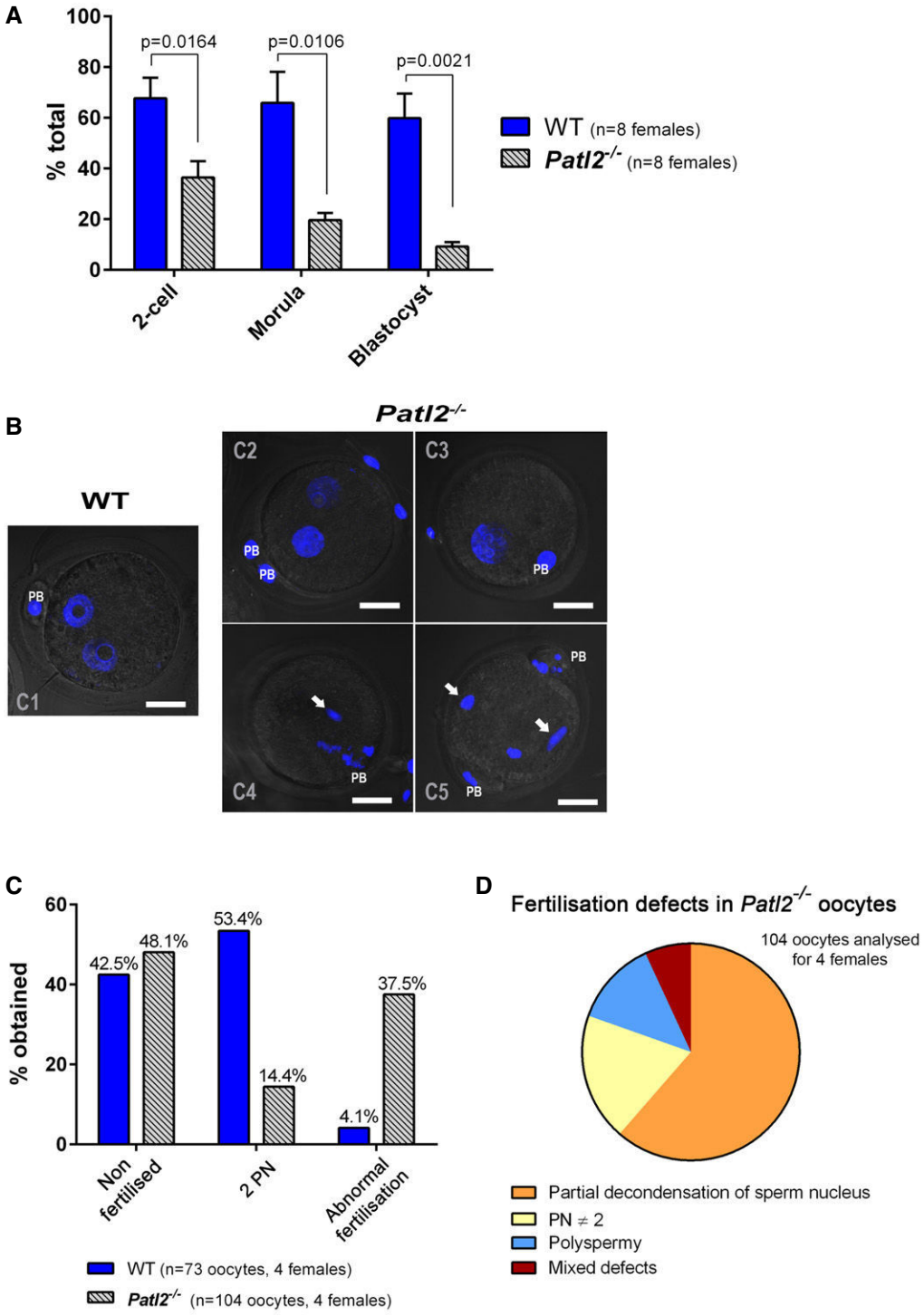


Figure 6. MII oocytes from *Patl2*$^{-/-}$ female mice exhibit abnormal fertilisation preventing normal embryo development.

A IVF outcomes (mean ± SEM) measured at the two-cell, morula and blastocyst stages show that the developmental competence of *Patl2*$^{-/-}$ oocytes is compromised. Oocytes were collected from stimulated WT and *Patl2*$^{-/-}$ females and sperm from WT males. For each IVF replicate (different WT males), IVF outcomes at different stages were compared for WT and *Patl2*$^{-/-}$ oocytes. Statistical differences were assessed using unpaired two-tailed *t*-tests.

B *Z*-stack projections of confocal images of 2PN zygotes obtained from WT and *Patl2*$^{-/-}$ oocytes 6-8 h after sperm–egg mixing. Fertilised WT oocytes exhibit normal 2PN stage (C1) whereas the number of fertilised *Patl2*$^{-/-}$ oocytes exhibiting normal 2PN stage is strongly reduced (C2) and most of them show defects such as absence of a male pronucleus (C3), partial decondensation of male PN (C4, C5, white arrows) or polyspermy (C5 white arrows). Scale bar = 20 μm, PB = polar body, PN = pronucleus.

C, D The percentage of 2PN obtained 6–8 h after fertilisation drops from 53.4% for WT to 14.4% for *Patl2*$^{-/-}$ oocytes, which exhibit various fertilisation abnormalities including partial decondensation of sperm DNA, polyspermy, abnormal number of PN or mixed defects (D).

difference in overall purified RNA concentration between WT and *Patl2*$^{-/-}$ oocytes, for both GV and MII stages, indicating that unlike Msy2, Patl2 is not a global regulator of mRNA stability and translation (Yu *et al*, 2001) (Appendix Fig S6C).

However, significant changes were observed for specific transcripts, both at GV and MII stages (Fig 7A). At the GV stage, lack of Patl2 induced a > twofold decrease for 95 transcripts ($P < 0.05$) and > twofold increase for 39 transcripts (Dataset EV1). At the MII stage, a > twofold decrease in 124 transcripts and > twofold increase in 122 transcripts were observed (Dataset EV2). Approximately one-third of the genes down-regulated at the GV stage (32) were also down-regulated at the MII stage, and half of the genes up-regulated at the GV stage (19) were also up-regulated at the MII stage (Dataset EV3 and Fig 7B). The impact of Patl2 deletion on gene expression at the GV and MII stages was then visualised using hierarchical clustering of genes with an absolute fold-change ((aFC))> 2, $P < 0.05$ (Fig 7C).

Phosphorylation, oxidation and other pathways are down-regulated in the absence of Patl2

Literature mining to determine the functions of the genes down-regulated in *Patl2*$^{-/-}$ oocytes identified several groups of genes reported to be involved in oocyte maturation. The proteins encoded are involved in several signalling pathways implicated in oocyte differentiation, oxidative stress, transcription and translation, exocrine modulation, meiosis and spindle formation (Table 2). We therefore suspect that the decreased expression of these genes interferes with normal oocyte differentiation. Regarding signalling pathways that are activated during oocyte maturation, significant down-regulation of transcripts for proteins involved in mTORC1, Wnt, NF-kB, MAP kinase and phosphatase signalling pathways was observed. It is worth noting that we also detected a more than twofold decrease in Pgrmc1, a receptor required to slow down oocyte meiotic progression (Guo *et al*, 2016; Table 2). This decrease in Pgrmc1 in GV oocytes was confirmed by RT–qPCR experiments (Appendix Fig S7).

Oocyte maturation is controlled by bidirectional crosstalk between the follicular cells and the oocyte: the secretion of oocyte-produced factors is necessary for follicle cell differentiation; follicle cells in turn secrete factors activating different signalling pathways within the oocyte (Li & Albertini, 2013). Interestingly, two factors, Cxcl14 and Adm2, known to play a crucial role in cumulus cell maturation (Bobe *et al*, 2006; Chang *et al*, 2011), were down-regulated in *Patl2*$^{-/-}$ oocytes (Table 2). We also observed a strong deregulation of several other interesting transcription factors, such as Sohlh2, which was down-regulated 2.2-fold and 5.3-fold at the GV and MII stages, respectively, and Eef1e1 (Table 2). We also noticed that two glutathione-S-transferases were repressed in *Patl2*$^{-/-}$ GV oocytes (Table 2). This down-regulation may increase oxidative stress within the oocyte. Spindle defects may also be aggravated by repression of Pak4 and Ccdc69 (Table 2), two proteins known to affect spindle assembly (Pal *et al*, 2010; Bompard *et al*, 2013). RT–qPCR experiments confirmed the down-regulation of Ccdc69 and Eef1e1 in GV oocytes (Appendix Fig S7).

Some transcripts, such as *Fgf9* and *Cdc25a* which control meiosis II (Assou *et al*, 2006), are known to be strongly up-regulated in oocytes after GVBD. Interestingly, both these transcripts were significantly down-regulated in *Patl2*$^{-/-}$ MII oocytes (Table 2). This decreased expression probably hampers final oocyte maturation. Among the list of genes up-regulated at both GV and MII (Dataset EV3), two particular genes, *Prr11* and *Ska2* (spindle and kinetochore associated complex subunit 2), are known to be required for spindle stability (Zhang *et al*, 2012). Ska2 up-regulation was confirmed by RT–qPCR for GV oocytes (Appendix Fig S7).

Finally, the lists of all down- and up-regulated proteins in *Patl2*$^{-/-}$ oocytes at the GV and MII stages were uploaded directly into Ingenuity Pathway Analysis (IPA) software, to explore their molecular and biological functions. At the GV stage, the most significantly down-regulated pathways were the ephrin receptor pathway and ephrin A signalling, which control embryo/trophoblast development (Fig EV4A and Table 2). Interestingly, at the MII stage, the cell cycle G1/S checkpoint was also significantly down-regulated (Fig EV4B), in accordance with the cell cycle defects visible in Fig 5.

The transcripts differentially expressed between MII and GV in WT and *Patl2*$^{-/-}$ samples affect a similar set of biological pathways and functions

To assess a possible role for Patl2 in the GV-MII transition, we identified the up- or down-regulated differential expression transcripts (DET) ($P < 0.05$, absolute fold-change > 2) between the GV and MII transcript lists for WT and *Patl2*$^{-/-}$ samples (Datasets EV4 and EV5). The lists of corresponding proteins were directly uploaded into IPA software to explore their molecular and biological functions. The functional pathways or networks with the highest confidence scores were then determined by right-tailed Fisher's exact tests. This analysis showed that a large portion of DET in WT samples were involved in energy production (oxidative phosphorylation and mitochondrial (dys)function, Appendix Fig S8) protein synthesis (EIF2 signalling, regulation of eIF4 and p7056K signalling) and DNA replication, recombination and repair (Fig EV5A). It is worth noting that the up- and down-regulated transcripts included in the DET lists were documented to functionally interact with each other, forming tightly connected networks. This observation strengthens the relevance of the data (Appendix Fig S9A). A similar analyse of DET from *Patl2*$^{-/-}$ samples indicated that the eight pathways with highest confidence scores were identical to those recorded for WT samples (Fig EV5B), suggesting that Patl2 only plays a minor role (or no role) in the

Figure 7. Transcriptional analysis of GV and MII oocytes from WT and Patl2$^{-/-}$ mice.

A Comparison of the transcriptional profiles in *Patl2*$^{-/-}$ oocytes versus WT oocytes at the GV or MII stages. GV oocytes were collected 44 h after PMSG injection and MII 13 h after hCG injections. For both MII and GV, two replicates for WT and three replicates for *Patl2*$^{-/-}$ oocytes were analysed.

B Venn diagrams representing down- or up-regulated genes in *Patl2*$^{-/-}$ oocytes (absolute fold-change ((aFC))> 2, $P < 0.05$) with respect to WT oocytes at the GV and MII stages.

C Hierarchical clustering of gene expression data for the down- and up-regulated genes (aFC> 2, $P < 0.05$) of *Patl2*$^{-/-}$ and WT oocytes at GV (left) and MII (right) stages, demonstrating the clustering of replicates to their respective groups.

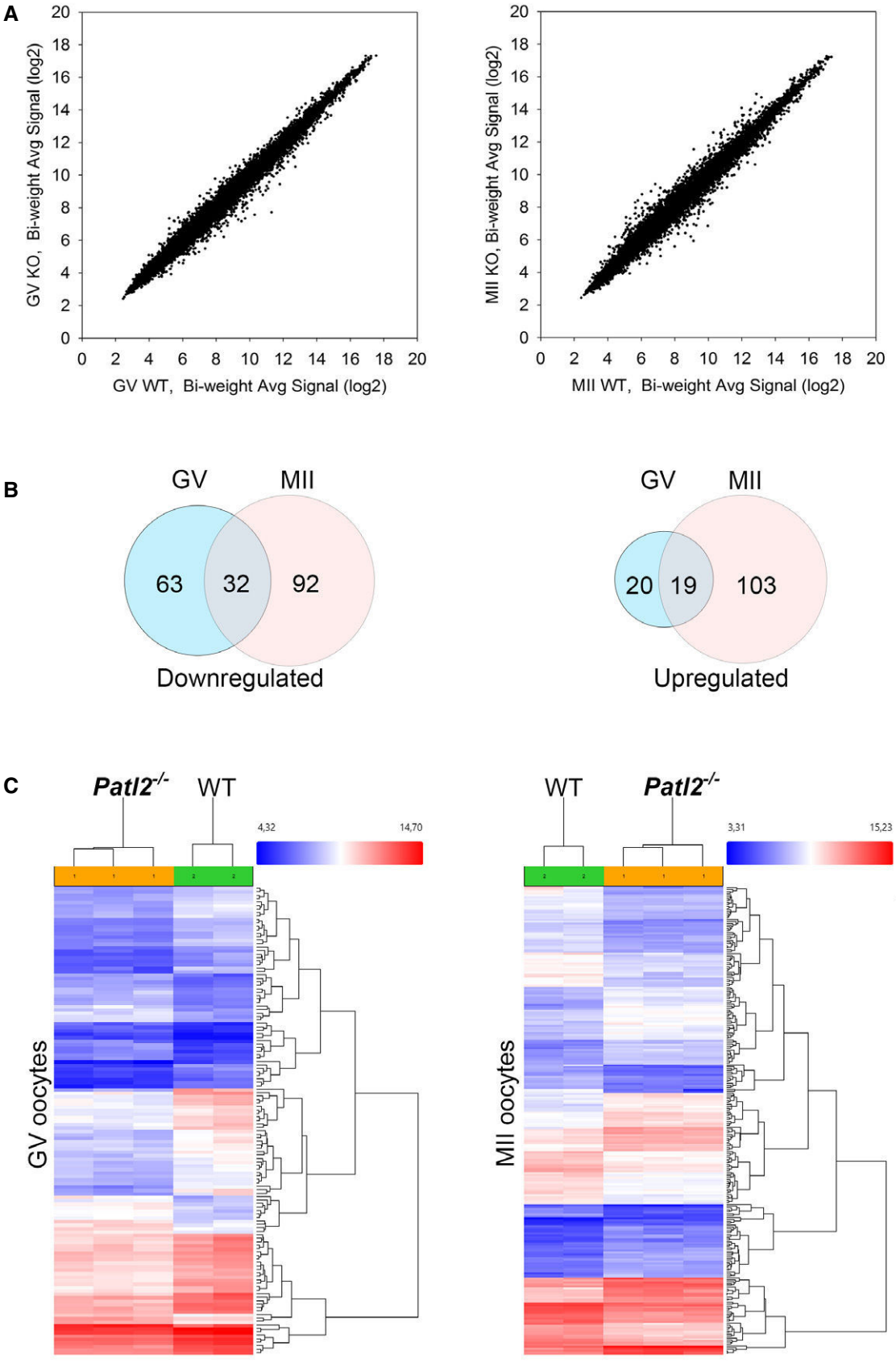

A
GV KO, Bi-weight Avg Signal (log2)
GV WT, Bi-weight Avg Signal (log2)
MII KO, Bi-weight Avg Signal (log2)
MII WT, Bi-weight Avg Signal (log2)
B
GV
MII
63
32
92
Downregulated
20
19
103
Upregulated
C
Patl2-/-
WT
4,32
14,70
GV oocytes
3,31
15,23
MII oocytes

Figure 7.

Table 2. Gene ontology analysis of microarray data.

Gene symbol	Gene name	WT GV	KO GV	Fold-change	WT MII	KO MII	Fold-change	(Possible) role in oocyte maturation and early embryo development	References
Elements of signalling pathways activated during oocyte maturation									
Gatsl3	GATS protein-like 3	9.64	8.12	−2.88	9.26	8.99	−1.21	CASTOR (gatsl3) is an arginine sensor for the mTORC1 pathway. Arginine breaks the CASTOR/GATOR2 complex. GATOR2 activates mTORC1. Activated mTORC1 will phosphorylate translation inhibitor 4E-BP1, releasing it from eukaryotic translation initiation factor 4E (eIF4E) which is then free to join the translation initiation complex. When the mTORC pathway is inhibited, bovine oocytes are blocked at the M1 stage	Chantranupong *et al* (2016), Mayer *et al* (2014)
Pygo1	Pygopus 1	8.68	6.9	−3.43	8.25	6.73	−2.87	Pygopus is a Wnt transcriptional component. Knockdown of Axin-1, a negative regulator of Wnt signalling leads to defective spindles, misaligned chromosomes, PB1 extrusion failure and impaired PN formation. Embryo development is also impacted since maternal Wnt/STOP signalling promotes cell division during early *Xenopus* embryogenesis. WNT signalling pathway is important for proper oocyte maturation	He *et al* (2016), Huang *et al* (2015), Spate *et al* (2014)
Ppp1r14b	Protein phosphatase 1, regulatory (inhibitor) subunit 14B	12.62	11.11	−2.84	12.65	11.34	−2.49	PP1 is an important protein involved in the cell cycle and controls dephosphorylation of numerous proteins such as proteins phosphorylated by Cdc2 and downstream mitotic kinases. Likely impacts meiotic control	
Ppp1r15a	Protein phosphatase 1, regulatory (inhibitor) subunit 15A	8.26	7.29	−1.96	9.12	8.17	−1.92		
Dusp1	Dual specificity phosphatase 1	11.94	10.49	−2.72	12.78	11.78	−2.01	Dusp1 is able to dephosphorylate MAP kinase, known to be very important in oocyte maturation and meiosis	Liang *et al* (2007)
Nfkbia	Nuclear factor of kappa light polypeptide gene enhancer in B cells inhibitor, alpha	9.78	8.33	−2.73	9.43	7.7	−3.31	NFKBIA is a gene involved in maintaining meiotic transcriptional arrest. Inhibits the NF-κB transcription factor. Change during GV-MII transition. Highly expressed at embryonic genome activation	Paciolla *et al* (2011)
Pgrmc1	Progesterone receptor membrane component 1	9.16	7.97	−2.27	9.53	8.01	−2.86	P4–PGRMC1 interaction attenuated oocyte meiotic progression and primordial follicle formation by decreasing intra-oocyte cAMP levels. When PGRMC1 is low, oocytes are not blocked by P4 and mature too fast	Guo *et al* (2016)
Secreted factors									
Cxcl14	Chemokine (C-X-C motif) ligand 14	8.08	6.99	−2.13	8.59	7.5	−2.13	Important factor for oocyte maturation in fish	Bobe *et al* (2006)

Table 2 (continued)

Gene symbol	Gene name	WT GV	KO GV	Fold-change	WT MII	KO MII	Fold-change	(Possible) role in oocyte maturation and early embryo development	References
Adm2	Adrenomedullin 2	11.75	10.64	−2.16	12.4	11.91	−1.4	ADM2 is a novel oocyte-derived ligand important for the regulation of cell interactions in COCs that functions, in part, by suppressing cumulus cell apoptosis	Chang *et al* (2011)
Transcription translation factors									
Med11	Mediator complex subunit 11;	9.84	8.28	−2.94	9.78	8.39	−2.62	Mediator of RNA polymerase II transcription, subunit 11 homologue (*S. cerevisiae*)	
Med29	Mediator complex subunit 29	11.01	9.33	−3.2	7.97	7.32	−1.57	Mediator complex subunit 29 (MED29) is part of a large multiprotein coactivator complex that mediates regulatory signals from gene-specific activators to general transcription machinery in RNA polymerase II mediated transcription	
Sohlh2	Spermatogenesis and oogenesis specific basic helix-loop-helix 2	7.18	6.06	−2.18	8	5.58	−5.33	SOHLH2 is a transcription factor that coordinates oocyte differentiation without affecting meiosis I and drives oocyte growth and survival but not meiosis I	Choi *et al* (2008), Shin *et al* (2017)
Slx2	Sycp3 like X-linked	8.04	6.81	−2.36	5.97	5.77	−1.15	SLX2 might be involved in DNA recombination, synaptonemal complex formation as well as sex body maintenance during meiosis	Shi *et al* (2013)
Prmt5	Protein arginine N-methyltransferase 5	9.05	7.88	−2.25	8.72	8.16	−1.47	PRMT5 negatively affects cyclin E1 promoter activity. Cyclin E1 is activated during meiosis in *Xenopus*. Prmt5 is critical in biological function in a wide range of cellular processes including development and methylates histones H2A and H4 in oocytes	Stopa *et al* (2015), Wilczek *et al* (2011)
Eef1e1	Eukaryotic translation elongation factor 1 epsilon 1	13.39	12.26	−2.19	13.24	11.78	−2.76	DNA damage response. Regulator of translation	Uniprot
E2f4	E2F transcription factor 4	13.61	12.5	−2.17	14.02	13.18	−1.8	Regulator of translation	
Oxidative stress									
Gstp2	Glutathione S-transferase, pi 2	12.04	10.41	−3.11	10.28	9.51	−1.71	Important in oxidative stress. A high level of gluthatione transferase is associated with higher oocyte developmental quality. Glutathione S-transferase is a marker of oocyte maturity in pigs	Paczkowski and Krisher (2010), Rausell *et al* (2007)
Gstp1	Glutathione S-transferase, pi 1	11.43	9.91	−2.86	9.74	9.14	−1.51		
Spindle assembly									
Pak4	p21 protein (Cdc42/Rac)-activated kinase 4	11.32	10.32	−2.01	12.02	11.41	−1.52	Ran is a substrate for p21-activated kinase 4 (PAK4). RanGTP is an important actor of spindle formation and asymmetric division during meiosis	Bompard *et al* (2013)

Table 2 (continued)

Gene symbol	Gene name	WT GV	KO GV	Fold-change	WT MII	KO MII	Fold-change	(Possible) role in oocyte maturation and early embryo development	References
Ccdc69	Coiled-coil domain containing 69	14.63	13.54	−2.13	14.8	13.87	−1.91	CCDC69 regulates central spindle assembly	Pal *et al* (2010)
Embryo/trophoblast factors									
Phlda2	Pleckstrin homology-like domain, family A, member 2	10.3	7.56	−6.7	7.7	7.05	−1.57	PHLDA2 is an imprinted gene, and only the maternal copy is expressed. This gene is associated with placental dysfunction. KO mice exhibit foetal growth deficiency	Frank *et al* (2002), Jensen *et al* (2014)
Efna1	Ephrin A1	9.73	7.34	−5.24	9.05	8.02	−2.04	Ephrins A1-4 were expressed in blastocysts. The ephrin A system is involved in regulating contact between blastocysts and endometrium during embryo implantation	Fu *et al* (2012), Fujii *et al* (2006)
Efna4	Ephrin A4	8.7	7.08	−3.07	8.55	7.21	−2.54		
Efna2	Ephrin A2	8.9	7.49	−2.66	8.81	7.86	−1.93		
Cstb	Cystatin B	11.81	9.44	−5.17	10.85	9.5	−2.55	The cathepsin–cystatin system plays an important role in trophoblast cell invasion and normal embryonic growth	Nakanishi *et al* (2005)
Prl8a2	Prolactin family 8, subfamily a, member 2	9.51	7.17	−5.05	9.12	5.93	−9.16	In Prl8a2 null tissues, genes expressed in the trophoblast are down-regulated	Alam *et al* (2015)
Crabp2	Cellular retinoic acid binding protein II	12.73	10.5	−4.69	12.34	10.73	−3.06	Altered expression level of endometrial CRABP2 is involved in abnormal endometrium-trophoblast interaction, which leads to implantation failure	Lee *et al* (2011)
Tlr8	Toll-like receptor 8	7.34	5.49	−3.61	7.15	5	−4.46	TLR8 is expressed in the trophoblast. Its inhibition suggests that it is necessary for successful establishment of early gestation in ewes	Kaya *et al* (2017), Ruiz-Gonzalez *et al* (2015)
Specific MII regulation									
Fgf9	Fibroblast growth factor 9	6.64	6.1	−1.45	7.21	5.88	−2.51	FGF9 counteracts retinoic acid, which promotes entry into meiosis. Its expression increases at MI stage and reaches highest level at the MII stage	Feng *et al* (2014)
Cdc25a	Cell division cycle 25A	11.59	11.55	−1.03	12.31	11.18	−2.19	Its expression increases significantly at the MII stage. Cdc25a is crucial in the MI-MII transition and its down-regulation results in fewer oocytes resuming meiosis and reaching MII	Solc *et al* (2008)

GV-MII transition. Similarly, in *Patl2*$^{-/-}$ samples, DET formed tightly connected networks (Appendix Fig S9B).

Discussion

Genetic basis of OMD

The genetic basis of OMD was unexplored until 2016 when heterozygous mutations of *TUBB8* were reported to account for approximately 30% (7/23) of OMD Han Chinese familial cases (Feng *et al*, 2016). In our 22-subject cohort, only one *TUBB8* deletion variant was identified by WES analysis, suggesting that *TUBB8* mutations may not be a common cause of OMD in North African women. No down-regulation of the expression of tubulin genes was observed in GV or MII oocytes from *Patl2*$^{-/-}$ mice, indicating that *PATL2*-dependent OMD does not involve tubulin deregulation. These results suggest geographical heterogeneity of the genetic causes of OMD. Indeed, 26% of our patients were found to present the same homozygous truncation of *PATL2*, demonstrating that *PATL2* mutation is a major cause of OMD

in North Africa. Interestingly, the same mutation was detected in all our patients, suggesting a founder effect. This hypothesis was reinforced by the analysis of the variants from WES data in the regions surrounding *PATL2*, which revealed a common homozygous haplotype (Dataset EV6). During the assessment of our data, two other studies were published associating *PATL2* with human OMD. The first was carried out on a cohort of 180 Han Chinese patients: it identified 5 *PATL2* nonsense variants and four mis-sense variants in five unrelated subjects. If the mis-sense variants are proven to be deleterious, this corresponds to a frequency of ~3% in the subjects tested (5/180) (Chen *et al*, 2017). In the second study, two homozygous *PATL2* variants were identified in two OMD subjects each from two different Saudi Arabian families. One of these variants is the same p.Arg160Ter variant presented here, whereas the other is a new mis-sense variant (Maddirevula *et al*, 2017). With cases across two continents, these studies, in combination with the present study, provide strong evidence that *PATL2* mutation is a major cause of human OMD whose phenotype is characterised by arrested development at the germinal vesicle stage.

Comparison of expression profiles and distribution for Patl2 and other RNA-binding proteins suggest a niche role for Patl2

By IF, we showed that (i) Patl2 is undetectable in primordial follicle oocytes; (ii) significant Patl2 translation begins at the primary follicle stage; (iii) Patl2 concentration peaks in secondary follicle oocytes; (iv) Patl2 is expressed only in the cytoplasm; (v) staining is relatively homogenous; and (vi) taking oocyte volume into account, the total amount of Patl2 increases as oocytes grow. These results clearly indicate that Patl2 plays a key role during oocyte growth and that it does not contribute to the maintenance of the primordial follicle pool in mouse ovaries. This conclusion is corroborated by our data showing comparable numbers of primordial follicles in control and Patl2$^{-/-}$ ovaries at 12 dpp. Nevertheless, the number of primordial follicles was found to be significantly higher at 26 dpp in Patl2$^{-/-}$ ovaries. This puzzling result deserves further investigation; it may indicate that expression of Patl2 is required for the primordial/primary follicle transition. The importance of Patl2 in oocyte growth is reinforced by our data showing that GV and MII oocytes are smaller in diameter in the absence of Patl2. Interestingly, although signal intensity for Patl2 declined during the GV NSN to SN transition, the protein remained clearly detectable in MII oocytes. This observation contrasts with data from *Xenopus* oocytes, where x-Pat1a was shown to completely disappear upon progesterone-induced meiotic maturation (Marnef *et al*, 2010; Nakamura *et al*, 2010). This discrepancy suggests that the function of Patl2 may have evolved in mammals to include a role during fertilisation and/or early development. A similar evolution has been reported for the germ-cell-specific translational repressor Msy2, which is detectable until the late two-cell embryo stage, undergoing changes in its phosphorylation state which presumably influence its interaction with mRNA (Yu *et al*, 2001).

Oocyte growth is characterised by extensive mRNA synthesis and accumulation within the cytoplasm. Although most of the mRNA synthesised during oocyte growth is immediately translated to support the growing oocyte, up to 30% of mRNAs are stored for subsequent translation and are required at meiotic resumption or for early zygote development (Pique *et al*, 2008). Several molecular mechanisms controlling mRNA stability have been described and involve a number of RNA-binding proteins (Clarke, 2012), although their relative contributions remain poorly understood. The best-described translational repression mechanism in oocytes involves a specific mRNA sequence, the cytoplasmic polyadenylation element (CPE), which is present in the 3′ UTR of the affected mRNAs. It associates with its binding factor, CPEB, as well as the cytoplasmic polyadenylation machinery (Racki & Richter, 2006; Sousa Martins *et al*, 2016). The role of CPEB in mediating translational repression is well documented, but its precise mechanism of action—in particular the roles of its molecular partners—remains to be fully elucidated. x-Pat1a, for example, was described to be an RNA-binding protein and a partner of CPEB in *Xenopus* oocytes, but its precise role in translational repression remains unknown (Nakamura *et al*, 2010). Apart from CPEB, the RNA-binding protein Msy2 also plays a key role in RNA stability. The *Xenopus* orthologue of Msy2 (Ybx2) has been reported to participate in the same complex as x-Pat1a and CPEB in *Xenopus* oocytes (Medvedev *et al*, 2011). Absence of Msy2 in mice leads to infertility and modulates the expression of around 7,000 transcripts (> twofold absolute change), that is approximately 30% of all transcripts. This broad effect could be explained if it acts as a sequence-independent RNA-binding protein or global regulator of mRNA stability. Finally, P-bodies, central to RNA processing in somatic cells, have been described in mammalian oocytes. The RNA helicase and translational repressor Ddx6 is a key component of all known types of P-body. Several studies have reported the existence of P-bodies containing Ddx6 in immature mouse oocytes, which disappear during oocyte growth, dispersing Ddx6 as well as Msy2 and Cpeb within the cytoplasm (Swetloff *et al*, 2009; Flemr *et al*, 2010). We therefore endeavoured to find out whether Patl2 may be part of a similar, temporary, structure.

To assess the possibility that Patl2 works in synergy with these proteins which are important for mRNA stability, we compared their expression profiles and abundance and performed colocalisation experiments. Patl2 localisation was determined using Patl2-HA mice. Msy2 and Ddx6 proteins were found to be expressed at significant levels in primordial follicle oocytes, whereas Patl2 was undetectable at this stage. The Msy2 concentration appeared constant throughout oocyte growth, and its abundance in antral follicle oocytes is in line with its role in the oocyte-embryo transition (Yu *et al*, 2001). Ddx6 is strongly expressed in primordial and primary follicle oocytes, with signal intensity gradually decreasing in secondary and tertiary follicle oocytes. As Ddx6 is a P-body component, this result is compatible with the disappearance of P-bodies in fully grown oocytes (Flemr *et al*, 2010). Unlike Patl2, Cpeb1 was detectable in primordial follicle oocytes, but its signal intensity pattern was otherwise similar to that of Patl2, peaking in secondary follicle oocytes and weakening in tertiary follicle oocytes. However, Cpeb1 appears to be localised in small RNPs, as evidenced by its clear punctiform staining (Fig 3). In colocalisation experiments performed on secondary follicle stage oocytes, Patl2 failed to clearly colocalise with any of the three mRNA regulatory proteins. Taken together, these results suggest that Patl2 may be involved in a new pathway controlling the stability of specific mRNA. This new pathway merits further study.

Impact of *Patl2* invalidation on the oocyte transcriptome

In this study, we assessed the impact of the absence of Patl2 on the transcriptomic profiles of GV and MII oocytes and on the

transcriptomic variation observed at the GV-MII transition. The GV-MII transition had previously been reported to be associated with specific degradation of a large number of transcripts associated with protein synthesis, DNA replication and energy production (Su *et al*, 2007). Our results confirm these previous findings, as the pathways that underwent the most significant changes during the GV-MII transition in WT oocytes were associated with oxidative phosphorylation and mitochondrial (dys)function (energy production), EIF2 and regulation of eIF4 and p7056K signalling (protein synthesis) and DNA repair (Su *et al*, 2007). Su *et al* (2007) also showed that 62 out of 88 transcripts in the oxidative phosphorylation pathways were down-regulated at the GV-MII transition. Here, we report down-regulation of 45 out of 88 transcripts, of which 32 are included in the set reported by Su *et al* (2007). Interestingly, our IPA analysis revealed other pathways that are modulated during the GV-MII transition, such as sirtuin signalling and mTOR signalling. The similarity of our results to previous reports, together with RT–qPCR analysis of selected genes, reinforces the conclusions drawn from the transcriptomic analyses. Importantly, we observed effects on the same major pathways in the GV-MII transition in *Patl2*$^{-/-}$ oocytes and in WT oocytes, suggesting that Patl2 is not involved in this transitional phase.

Our data show that the absence of Patl2 induces transcriptomic deregulation, affecting 135 genes at the GV stage and 248 genes at the MII stage with an absolute fold-change > 2, ($P < 0.05$). This relatively small list includes genes of highly relevant function (Table 2). Indeed, several of the corresponding proteins, such as Prgmc1 and Slohlh2, are reported to be involved in oocyte maturation. Sohlh2 is of particular interest as it has been described to be required for oocyte growth, and Sohlh2$^{-/-}$ females were found to be infertile (Choi *et al*, 2008). Interestingly, this factor does not affect meiosis I (Shin *et al*, 2017), which is line with the observation that *Patl2*$^{-/-}$ females are able to generate MII oocytes. The factors Cxcl14 and Adm2, known to be necessary for cumulus cell maturation, were also found to be down-regulated. However, no decrease in follicle size or any other visible defects were observed in ovarian sections at 26 dpp, suggesting that any follicular defects are subtle. Importantly, several key proteins protecting cells against oxidative stress were down-regulated (GST forms). Oocytes are very sensitive to oxidative stress, which causes spindle abnormalities (Choi *et al*, 2007) and affects developmental competence (Rausell *et al*, 2007); in addition, glutathione S-transferase has been reported to be a marker of oocyte maturity in pigs (Paczkowski & Krisher, 2010).

Some of the up-regulated transcripts could also interfere with normal oocyte maturation. The most extensively up-regulated gene at the MII stage was *Prr11*. In WT MII oocytes, its expression was low, suggesting that its role in normal oogenesis is minimal or null. *Prr11* deregulation has however been shown to dramatically modify the cell cycle, although how it does so remains unclear (Li *et al*, 2017). Its strong up-regulation may therefore interfere with meiosis or early development. Another remarkable gene up-regulated at both GV and MII is *Ska2* (spindle and kinetochore associated complex subunit 2), which is known to control spindle stability during meiosis (Zhang *et al*, 2012). This up-regulation could explain the numerous defects observed in MI and MII meiotic spindles. Up-regulation of these genes could be a cellular response attempting to address the transcript deregulation induced by Patl2-deficiency. Taken together, these results underline the importance of PATL2 in the regulation of a specific subset of mRNAs required for the generation of mature oocytes.

Moreover, a considerable number of genes known to be involved in pre-implantation embryonic development were also found to be down-regulated in both GV and MII oocytes from *Patl2*$^{-/-}$ mice. This down-regulation may contribute to the poor pre-implantation developmental competence observed for *Patl2*$^{-/-}$ embryos (Fig 6). These results suggest that Patl2 plays an essential role in maintaining the integrity of a small pool of mRNAs synthesised during oocyte growth and necessary after fertilisation and during early embryo development. This role is compatible with the presence of Patl2 in MII oocytes and its function as a translation repressor, as described for its *Xenopus* homologue (Marnef *et al*, 2010).

In conclusion, we report that PATL2 is a vital player in oocyte growth and maturation, where it regulates the expression of mRNAs encoding proteins crucial for oocyte meiotic progression and early embryonic development, and that its invalidation causes OMD in humans. Unravelling the molecular basis of OMD will help patients by improving diagnosis and our understanding of their disease. This work will also be of tremendous interest to the fast-growing field of clinical *in vitro* oocyte maturation, necessary for the development of a number of applications, such as fertility preservation for cancer patients (in particular for young prepubertal girls; Kim *et al*, 2016) before commencing reprotoxic chemotherapy, or *in vitro* maturation of primary/secondary follicles for patients with premature ovarian failure (Kim, 2012; Yin *et al*, 2016).

Materials and Methods

Ethics

After sperm analysis and IVF treatment, performed independently of the research presented in this paper, couples presenting an interesting phenotype (male and/or female infertility) were selected and referred by their physician. The physician explained the study and its objectives before subjects signed an informed consent form in line with the local IRB protocols and the principles of the Declaration of Helsinki. DNA samples from couples of interest were collected and the identity of patients coded in such a manner that subjects could not be identified. All medical records were saved. No specific treatment was given to any female patient for research purposes, and all oocytes collected were used for IVF/ICSI treatment only.

Patients and biological samples

A total of 23 patients were recruited. All subjects were of North African descent, mainly Tunisia and Algeria, with one patient from Mauritania. All had undergone one or two cycles of ovarian stimulation to allow egg collection for use in IVF. Human DNA samples were obtained from patients consulting for diagnosis and assisted reproductive techniques at the PolyClinique des Jasmins (Tunis, Tunisia) ($n = 21$) or at the reproductive unit located in Bondy, France ($n = 2$).

Patients underwent standard controlled ovarian stimulation by the administration of gonadotropins using either the long agonist or the antagonist protocol. Women were given between 150 and 225 IU recombinant follicle-stimulating hormone (Gonal-F; Merck-Serono). Subsequently, ultrasound was performed and follicular response

recorded from day 5 of gonadotropin stimulation. When at least two follicles were ≥ 18 mm in diameter, 6,500 IU or 10,000 IU human chorionic gonadotropin was administered. Oocytes were retrieved 34–36 h later by vaginal ultrasound-guided follicular puncture.

Control cohort: Since 2007, and in parallel to this study on OMD, we also performed a large study on male infertility, involving several hundred couples. Most of these couples had IVF/ICSI treatments, and female partners of infertile men with normal fertility levels were considered as fertile controls. Their characteristics were recorded and used anonymously. These cohorts have already been described in our previous works concerning male infertility (Dieterich *et al*, 2007; Harbuz *et al*, 2011; Ben Khelifa *et al*, 2014; Coutton *et al*, 2018; Kherraf *et al*, 2018).

Exome sequencing and bioinformatics analysis

Genomic DNA was isolated from saliva using Oragene saliva DNA collection kit (DNAgenotek Inc., Ottawa, Canada). Exome capture was performed using NimblGen SeqCap EZ Kit version 2 (Roche). Sequencing was conducted on Illumina HiSeq2000 instruments with paired-end 76-nt reads. Sequence reads were aligned against the reference genome (hg19) using MAGIC (SEQC/MAQC-III Consortium, 2014). Duplicate reads and reads mapping to multiple locations in the genome were excluded from further analysis. Positions with a sequence coverage of < 10 on either the forward or reverse strand were excluded. Single nucleotide variations (SNV) and small insertions/deletions (indels) were identified and quality-filtered using in-house scripts. The most promising candidate variants were identified using an in-house bioinformatics pipeline which is described in (Coutton *et al*, 2018). Variants with a minor allele frequency > 5% in the NHLBI ESP6500 or in 1,000 Genomes Project phase 1 data sets, or > 1% in ExAC, were discarded. These variants were also compared to an in-house database of 56 control exomes. All variants present in a homozygous state in this database were excluded. Variant Effect Predictor (Ensembl) was used to predict the impact of the selected variants. Only variants affecting splice donor/acceptor sites or causing frameshift, inframe insertions/deletions, stop gain, stop loss or mis-sense variants were retained, except for those scored as "tolerated" by SIFT (sift.jcvi.org) and as "benign" by Polyphen-2 (genetics.bwh.harvard.edu/pph2). All steps from sequence mapping to variant selection were performed using the ExSQLibur pipeline (https://github.com/tkaraouzene/ExSQLibur).

SANGER sequencing

Sanger sequencing was carried out using the primers listed in Appendix Table S1A. PCR amplification (35 cycles) was performed with a melting temperature of 60°C. Sequencing reactions were performed using BigDye Terminator v3.1 (Applied Biosystems). Sequence analyses were carried out on ABI 3130XL (Applied Biosystems). Sequences were analysed using seqscape software (Applied Biosystems). The primers used for Sanger verification of *PATL2* mutations in patients are listed in Appendix Table S1A.

Mice

All animal procedures were performed according to the French and Swiss guidelines on the use of animals in scientific investigations after approval of the study protocol by the local Ethics Committee (ComEth Grenoble No. 318, ministry agreement number #7128 UHTA-U1209-CA) and the Direction Générale de la Santé (DGS) for the State of Geneva. *Patl2*$^{FL/FL}$ mice (C57BL/6NTac-Patl2 < tm1a) were generated by the EUCOMM Consortium (http://www.mousephenotype.org/about-ikmc/eucomm). They were obtained from the Mouse Clinical Institute—MCI, Strasbourg, France. Deletion of exon 7 was obtained by crossing them with adult heterozygous EIIaCre transgenic mice (Lakso *et al*, 1996), obtained from Institut Cochin—Inserm 1016—CNRS 8104—Paris. The ellaCre carries a Cre transgene under the control of the adenovirus EIIa promoter. This construction targets expression of Cre recombinase to early mouse embryos. Cre expression is thought to occur prior to implantation in the uterine wall. Cre-mediated recombination thus affects a wide range of tissues, including the germ cells that transmit the genetic alteration to progeny.

Patl2-HA knock-in mice were generated by CRISPR/Cas9 technology (Kherraf *et al*, 2018). Twenty-seven nucleotides encoding the HA tag were inserted in the c-terminal of the *Patl2* gene, immediately before the TAG stop codon (TAC CCA TAC GAT GTT CCA GAT TAC GCT TAG) in C57BL/6 mice. One plasmid containing one sgRNA and Cas9 was injected (5 ng/μL) with the single-stranded DNA (50 ng/μl) for homology-directed repair. The single-stranded DNA was 187 nt long and contained two 80 nt long arms homologous to the nucleotides located before and after the *Patl2* stop codon, surrounding the 27 nucleotides encoding the HA tag, and ending with a TAG stop codon. The sgRNA sequence overlapped the native Patl2 stop codon which, eight nucleotides later, was followed by a TGG representing a suitable protospacer adjacent motif (PAM) sequence. The synthetic ssDNA was purchased from IDT (Leuven, Belgium). After injection, zygotes were left for 4–6 h before introducing them into pseudopregnant host females where they were carried to term. Edited founders were identified by Sanger sequencing from digit biopsies. Mice carrying the desired modification were crossed with C57BL6/J to verify germline transmission and eliminate any possible mosaicism. Heterozygous animals with the same modification were then mated to produce homozygous offspring.

Mice were housed with unlimited access to food and water in a facility with 12-h light per day. Animals were sacrificed by cervical dislocation at the ages indicated.

Genotyping

DNA for genotyping was isolated from tail biopsies. Tail biopsies (2 mm long) were digested in 200 μl lysis buffer (Direct PCR Lysis Reagent (Tail); Viagen Biotech inc, CA, USA) and 0.2 mg proteinase K for 12–15 h at 55°C, followed by 1 h at 85°C to inactivate proteinase K. The DNA was directly used to perform PCRs. Multiplex PCR was done for 35 cycles, with an annealing temperature of 62°C and an elongation time of 45 s at 72°C. PCR products were separated by 2% agarose gel electrophoresis. Genotypes were determined depending on the migration pattern. Primers for Patl2-KO and Patl2-HA-tagged mice are listed in Appendix Table S1B,C.

RT–qPCR

To verify the microarray results, RT–qPCR was performed on GV-stage oocytes from wild-type and knockout mice using the TaqMan® Gene Expression Cells-to-CT™ Kit (Ambion). Ten cells were washed

twice in PBS before lysis in 50 µl lysis/DNase buffer, giving a final lysate concentration of 0.2 cells/µl. 10 µl of cell lysate was added to 1× RT buffer for a final volume of 50 µl, resulting in a final concentration of 0.04 cell/µl. Gene expression was quantified using 4 µl of the resulting cDNA. Reactions were performed in 96-well plates on a StepOnePlus instrument (Applied Biosystems). Primers and probes (TaqMan Gene Expression assays) were ordered form Applied Biosystems and consist of a pair of unlabelled PCR primers and a TaqMan probe with a dye label (FAM) on the 5′ end, and a minor groove binder (MGB) and non-fluorescent quencher (NFQ) on the 3′ end. Normalisation was performed relative to *Gapdh* (fold-change found equal to 1 by microarray analysis). The ΔCt, which is determined by subtracting the Ct number for the reference gene from that of the target gene was statistically analysed. Relative quantification (RQ) was calculated ($2^{-\Delta\Delta Ct}$) and used to graphically present the results. Primers are listed in Appendix Table S1B.

Phenotypic analysis of mutant mice

Collection of GV and MII oocytes

GV oocytes were collected from 3- to 8-week-old females by puncturing ovaries with a 26-gauge needle in M2 medium (Sigma-Aldrich, Lyon, France) containing 150 µM dibutyryl cyclic AMP to prevent GV oocyte maturation and GVBD. Follicular cells surrounding GV oocytes were enzymatically (hyaluronidase 0.1%) and mechanically removed using a pipette with an inner diameter of around 100 µm. To extract GV oocytes from pre-antral follicles, follicles were treated with 2 mg/ml collagenase IV and 1 mg/ml hyaluronidase and ruptured with a fine pipette. For MII oocytes, 48 h after PMSG injection, 5 UI of human chorionic gonadotropin (hCG) was injected (Chorulon, Intervet), and cumulus-oocytes complexes (COCs) released from the ampullae were collected in M2 medium after 13 h. COCs were incubated in hyaluronidase enzyme (0.1 mg/ml, Sigma-Aldrich) for 5 min, and follicle-cell-free MII oocytes were obtained by pipetting.

In vitro fertilisation (IVF)

Eggs were collected from 4- to 8-week-old females, synchronised by exposing to 5 units of PMSG (Synchropart, Intervet, Beaucouze, France) and 5 units of hCG. Sperm from healthy males (B6D2 F1) capacitated for 80 min in M16 + 2% BSA (A3803 Sigma-Aldrich) were simultaneously added to the COCs and incubated in M16 medium for 4 h at 37°C under 5% CO_2. Unbound sperm were washed away after 4-h incubation. 24 h after fertilisation, the different stages, that is unfertilised oocytes and two-cell embryos (as an indication of successful fertilisation), were scored. Embryos were cultured in potassium simplex optimised medium (KSOM) (1×, Life technologies) supplemented with essential and non-essential amino acids (KSOM/EAA): two-cell embryos were incubated in KSOM/EAA medium at 37°C under 5% CO_2 and cultured up to the blastocyst stage.

Analysis of fertilised oocytes

After fertilisation, zygotes were transferred to a clean well containing M16 media and left for a further 2 h before being fixed in PFA 4% for 10 min and stained with Hoechst 33342 (2 µg/ml in PBS-PVP 0.5%). Oocytes were observed under confocal microscopy in drops of PBS-PVP covered with mineral oil on LabTeK chambered coverglass plates (Thermo Fisher Scientific, Villebon sur Yvette, France).

Immunostaining of GV and MII oocytes

Cumulus-free MII or GV oocytes prepared as described above were fixed in 4% PFA for 15 min, washed twice in PBS-PVP (0.5%) and permeabilised in PBS-Triton 0.25% for 20 min. Oocytes were then blocked in blocking solution (2% NGS, 0.1% Triton) for 90 min at room temperature and stained overnight at 4°C with primary antibodies (Appendix Table S2). All antibodies were diluted in blocking solution. Following staining, oocytes were washed twice with PBS-PVP 0.5% and incubated for 1 h at 37°C in secondary antibody and Hoechst 33342 at 2 µl/ml in blocking solution. Oocytes were finally washed twice and observed by confocal microscopy (Zeiss LSM 710) in drops of PBS-PVP covered with mineral oil on LabTeK coverglass plates. Images were processed using Zen 2.1 software.

Immunostaining of ovarian sections for oocyte counting

Mice were sacrificed by cervical dislocation, and ovaries were collected and fixed for 4 h in paraformaldehyde (4%). Ovaries were dehydrated embedded in paraffin, and 3-µm sections were cut. For histological analysis, after deparaffinisation, HIER and blocking slides were incubated with anti-Msy2 antibody (Appendix Table S2) for 1 h followed by a fluorophore-conjugated secondary antibody and Hoechst 33342 counterstaining. Stained sections were digitised at ×40 magnification using an Axioscope microscope (Zeiss, Germany) equipped with a motorised X–Y-sensitive stage. For each ovary, three sampling zones separated by 100–200 µm were studied. For each sampling zone, six to seven consecutive sections were stained and follicles counted, and the mean number of each class of follicle was calculated per sampling zone for WT and *Patl2*$^{-/-}$ mice.

Protein extraction and Western Blot

WT and PATL2-HA-tagged mice were sacrificed by cervical dislocation, and the hypothalamus, pituitary glands and livers were collected, snap-frozen in liquid nitrogen and stored at −80°C. The day of the experiment, organs were thawed in RIPA lysis buffer (50 mM Tris–HCl pH 7.5, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, Complete EDTA-free protease inhibitor cocktail tablet (Roche)) and homogenised using a Dounce tissue grinder before mixing the supernatant with loading buffer in equal volumes (4% SDS, 62.5 mM Tris pH 6.8, 0.1% bromophenol blue, 15% glycerol, 5% bromophenol blue). The cytoplasmic fraction was isolated by centrifugation at 3,000 × *g* for 10 min. GV oocytes were collected in M2 medium supplemented with 150 mg/ml dbcAMP and prepared as described above. Oocytes were washed three times in PBS-PVP 0.5% to remove proteins from the M2 medium and added to an equal volume of loading buffer for a final volume of 20 µL. All protein samples were heated to 65°C for 15 min before loading onto a 4–20% TGX Mini-PROTEAN stain-free precast gel (Bio-Rad). Proteins were transferred onto a PVDF membrane, blocked in 5% milk and incubated overnight at 4°C in anti-HA antibody (Appendix Table S2) in blocking solution.

Oocyte RNA profiling

Total RNAs were purified from wild-type or *Patl2*$^{-/-}$ mouse oocytes using Norgen Biotek's Single cell Purification kit (Cat. 51800; Thorold, ON, Canada) according to the manufacturer's protocol. The GV oocyte samples (two WT and three *Patl2*$^{-/-}$ samples)

The paper explained

Problem
Infertility is considered a global public health issue since it affects more than 70 million couples worldwide. The common treatment for infertile couples is IVF (*in vitro* fertilisation) or ICSI (intracytoplasmic sperm injection); however, both of these techniques require mature egg cells having correctly completed both meiotic divisions. In rare cases, women who undergo hormonal stimulation for IVF or ICSI produce only immature eggs which cannot be fertilised or made to mature *in vitro*. We named this condition "oocyte meiotic deficiency". This condition is very poorly understood.

Results
We analysed a group of 23 women with oocyte meiotic deficiency by whole exome sequencing and identified a genetic mutation in the gene *PATL2* in 26% of patients. *PATL2* encodes an RNA-binding protein that has been shown to play an important role in egg cell maturation in frogs, but its function in mammals has not been studied until now. Using a knockout mouse model, we showed that Patl2 deficiency leads to defective oocyte maturation due to the deregulation of important genes involved in egg cell growth, meiosis and early embryo development. We also showed that Patl2 has a unique expression profile in comparison with other known oocyte RNA-binding proteins. These results indicate a specific, niche function for Patl2 in mRNA regulation during egg cell maturation in mammals.

Impact
We have identified *PATL2* mutation as a major cause of oocyte meiotic deficiency. This finding can benefit patients through an improved understanding of their condition and allowing for better informed decisions regarding treatment options. We have also demonstrated that mouse Patl2 plays an important role in mRNA regulation, furthering our understanding of the process of mammalian oocyte maturation. This finding is highly relevant to the fast-growing field of clinical *in vitro* oocyte maturation, with wide-ranging applications including fertility treatment for patients with premature ovarian failure and fertility preservation for cancer patients facing reprotoxic chemotherapy.

contained a total number of cells ranging between 20–32 oocytes/sample; the MII oocyte samples (2 WT and 3 $Patl2^{-/-}$ samples) contained between 32 and 46 oocytes. Purified RNA concentrations were assessed by RT–qPCR using the Affymetrix RNA quantification kit. Sample RNAs (250 pg) were converted to biotin-labelled single-strand cDNAs using the Affymetrix Genechip WT pico kit. Labelled and fragmented ss-cDNA (5.5 μg) were hybridised to Affymetrix arrays (Genechip Clariom D mouse). The array complexity allowed analysis of the expression of more than 66,000 different transcripts (including transcript variants and non-coding-cRNA) by measuring the expression level of their individual exons.

The data obtained for the whole set of samples were normalised by applying the RMA process (Affymetrix Expression Consol). Raw and normalised data were uploaded to the GEO database (Accession number GSE100117). Probe-set annotation, quantitative expressions of all the transcripts and comparisons between the different groups of samples (GV-KO versus GV-WT; MII-KO versus MII-WT) were analysed using TAC.v3 software (Affymetrix). Expression levels are reported as log2 conversions of the intensities measured for the ss-cDNA hybridised to the arrays. Differential expression of transcripts between two groups of samples was considered significant when the expression ratios of the transcripts between the two groups were at least twofold higher or lower (in linear scale) with a P-value < 0.05 (ANOVA test).

Transcriptomic analysis by IGA
IPA software (https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/) was used for the functional assessment of DET that were deregulated in $Patl2^{-/-}$ versus WT samples and to construct molecular interaction networks. IPA is software application that helps classify the molecular networks and functions most relevant to transcripts of interest. DET were imported to IPA for analysis. IPA generates pathways based on the transcripts contained in a data set and the information stored in the Ingenuity Pathways Knowledge Base. The significance of the transcripts' annotation is indicated by a P-value of < 0.05, as determined by a right-tailed Fisher's exact test.

RT–qPCR validation experiments
Quantification of gene expression levels by RT–qPCR was performed on GV-stage oocytes from WT and $Patl2^{-/-}$ mice. cDNAs were produced from oocyte lysates using the TaqMan® Gene Expression Cells-to-CT™ Kit. Oocytes were isolated from the ovaries of 4-week-old mice (average of 50 GV per mouse) and washed twice in PBS-PVP 0.5% before lysis. Each lysis reaction was performed with 10 oocytes in a total volume of 50 μl of lysis buffer supplemented with DNase at a ratio of 1:100. RT reactions were performed using 10 μl of the lysate in a total volume of 50 μl using the same kit. Real-time PCR was performed in 96-well plates on the StepOnePlus system (Applied Biosystems). Reactions were performed in a total volume of 20 μl, comprising 4 μl of cDNA, 1 μl of TaqMan Assays and 10 μl of Master Mix. Data were normalised relative to expression levels measured for the *Gapdh* reference gene. ΔCt, which is determined by subtracting the Ct of the reference gene from that of the target gene, was the subject of statistical analysis.

Statistical analyses

Statistical analyses were performed using SigmaPlot or GraphPad prism 6. Unpaired t-tests were used to compare WT and $Patl2^{-/-}$ samples. Data are represented as mean ± SEM. Statistical tests for which the two-tailed P-value ≤ 0.05 were considered significant.

Data availability

Transcriptomic data are available in the GEO database: accession number GSE100117 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi). Clinical exomic data are available in the EGA data base: EGAS00001002903 (https://www.ebi.ac.uk/ega/home).

Expanded View for this article is available online.

Acknowledgements
We thank the IAB microscopy platform and Mylene PEZET, Alexei GRICHINE, Jacques MAZZEGA for their technical help. We thank Emeline FONTAINE PELLETIER (INSERM 1209, CNRS UMR 5309) for her generous donation of antibodies. We thank Marie-Christine BIRLING for helping with mouse genotyping. We thank Marcio CRUZEIRO (Institut Cochin, Paris, France) for providing the EIIaCre transgenic mice and Julien Fauré for access to molecular biology facility. Lastly, we thank all patients and control subjects for their participation.

We are also grateful to Jacques Brocard from the Photonic Imaging Centre at Grenoble Institute Neuroscience (Univ Grenoble Alpes—Inserm U1216—ISdV core facility and certified by the IBiSA label). This work was supported by the following grants: "Investigation of the genetic aetiology of oocyte meiotic failure (OMF) by exome sequencing" (awarded by the Fondation Maladies Rares (FMR) for the "High throughput sequencing and rare diseases 2012" programme), funding for the "MAS-Flagella" project (French National agency for research (ANR)) and the "Whole genome sequencing of patients with Flagellar Growth Defects (FGD)" (DGOS for the PRTS 2014 programme).

Author contributions

CA and PFR analysed the data and wrote the manuscript; Z-EK, AA-Y, CC and MB performed molecular work; TK and NT-M analysed genetic data; MC-K performed IF and histological experiments; MC-K, JE, ELB and GM performed IVF experiments; JPI, AG and SA performed transcriptome analyses; EL and SPB performed biochemistry experiments; MC-K, EL and ELB performed histological study. SN, BC and Z-EK made HA-tagged mice. SFBM, IC-D, LH, OM, MM, HL, MK and RZ provided clinical samples and data; CA and PFR designed the study, supervised all laboratory work, had full access to all the data in the study and took responsibility for the integrity of the data and its accuracy. All authors contributed to the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

For more information

OMIM gene number 614661: http://omim.org/entry/614661, phenotype entry 617743: http://omim.org/entry/617743.

References

Alam SM, Konno T, Soares MJ (2015) Identification of target genes for a prolactin family paralog in mouse decidua. *Reproduction* 149: 625 – 632

Assou S, Anahory T, Pantesco V, Le Carrour T, Pellestor F, Klein B, Reyftmann L, Dechaud H, De Vos J, Hamamah S (2006) The human cumulus–oocyte complex gene-expression profile. *Hum Reprod* 21: 1705 – 1719

Beall S, Brenner C, Segars J (2010) Oocyte maturation failure: a syndrome of bad eggs. *Fertil Steril* 94: 2507 – 2513

Ben Khelifa M, Coutton C, Zouari R, Karaouzene 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

Bobe J, Montfort J, Nguyen T, Fostier A (2006) Identification of new participants in the rainbow trout (*Oncorhynchus mykiss*) oocyte maturation and ovulation processes using cDNA microarrays. *Reprod Biol Endocrinol* 4: 39

Bompard G, Rabeharivelo G, Cau J, Abrieu A, Delsert C, Morin N (2013) P21-activated kinase 4 (PAK4) is required for metaphase spindle positioning and anchoring. *Oncogene* 32: 910 – 919

Braun JE, Tritschler F, Haas G, Igreja C, Truffault V, Weichenrieder O, Izaurralde E (2010) The C-terminal alpha-alpha superhelix of Pat is required for mRNA decapping in metazoa. *EMBO J* 29: 2368 – 2380

Chang CL, Wang HS, Soong YK, Huang SY, Pai SY, Hsu SY (2011) Regulation of oocyte and cumulus cell interactions by intermedin/adrenomedullin 2. *J Biol Chem* 286: 43193 – 43203

Chantranupong L, Scaria SM, Saxton RA, Gygi MP, Shen K, Wyant GA, Wang T, Harper JW, Gygi SP, Sabatini DM (2016) The CASTOR proteins are arginine sensors for the mTORC1 pathway. *Cell* 165: 153 – 164

Chen B, Zhang Z, Sun X, Kuang Y, Mao X, Wang X, Yan Z, Li B, Xu Y, Yu M *et al* (2017) Biallelic mutations in PATL2 cause female infertility characterized by oocyte maturation arrest. *Am J Hum Genet* 101: 609 – 615

Choi WJ, Banerjee J, Falcone T, Bena J, Agarwal A, Sharma RK (2007) Oxidative stress and tumor necrosis factor-alpha-induced alterations in metaphase II mouse oocyte spindle structure. *Fertil Steril* 88: 1220 – 1231

Choi Y, Yuan D, Rajkovic A (2008) Germ cell-specific transcriptional regulator sohlh2 is essential for early mouse folliculogenesis and oocyte-specific gene expression. *Biol Reprod* 79: 1176 – 1182

Clarke HJ (2012) Post-transcriptional control of gene expression during mouse oogenesis. *Results Probl Cell Differ* 55: 1 – 21

Coutton C, Vargas AS, Amiri-Yekta A, Kherraf ZE, Ben Mustapha SF, Le Tanno P, Wambergue-Legrand C, Karaouzene T, Martinez G, Crouzy S *et al* (2018) Mutations in CFAP43 and CFAP44 cause male infertility and flagellum defects in trypanosoma and human. *Nat Commun* 9: 686

De La Fuente R (2006) Chromatin modifications in the germinal vesicle (GV) of mammalian oocytes. *Dev Biol* 292: 1 – 12

Dieterich K, Soto RR, Faure AK, Hennebicq S, Ben Amar B, Zahi M, Perrin J, Martinez D, Sele B, Jouk PS *et al* (2007) Homozygous mutation of AURKC yields large-headed polyploid spermatozoa and causes male infertility. *Nat Genet* 39: 661 – 665

Feng CW, Bowles J, Koopman P (2014) Control of mammalian germ cell entry into meiosis. *Mol Cell Endocrinol* 382: 488 – 497

Feng R, Sang Q, Kuang Y, Sun X, Yan Z, Zhang S, Shi J, Tian G, Luchniak A, Fukuda Y *et al* (2016) Mutations in TUBB8 and human oocyte meiotic arrest. *N Engl J Med* 374: 223 – 232

Flemr M, Ma J, Schultz RM, Svoboda P (2010) P-body loss is concomitant with formation of a messenger RNA storage domain in mouse oocytes. *Biol Reprod* 82: 1008 – 1017

Frank D, Fortino W, Clark L, Musalo R, Wang W, Saxena A, Li CM, Reik W, Ludwig T, Tycko B (2002) Placental overgrowth in mice lacking the imprinted gene Ipl. *Proc Natl Acad Sci USA* 99: 7490 – 7495

Fu Y, Fu J, Ren Q, Chen X, Wang A (2012) Expression of Eph A molecules during swine embryo implantation. *Mol Biol Rep* 39: 2179 – 2185

Fujii H, Tatsumi K, Kosaka K, Yoshioka S, Fujiwara H, Fujii S (2006) Eph-ephrin A system regulates murine blastocyst attachment and spreading. *Dev Dyn* 235: 3250 – 3258

Furuya M, Tanaka M, Teranishi T, Matsumoto K, Hosoi Y, Saeki K, Ishimoto H, Minegishi K, Iritani A, Yoshimura Y (2007) H1foo is indispensable for meiotic maturation of the mouse oocyte. *J Reprod Dev* 53: 895 – 902

Guo M, Zhang C, Wang Y, Feng L, Wang Z, Niu W, Du X, Tang W, Li Y, Wang C *et al* (2016) Progesterone receptor membrane component 1 mediates progesterone-induced suppression of oocyte meiotic prophase i and primordial folliculogenesis. *Sci Rep* 6: 36869

Harbuz R, Zouari R, Pierre V, Ben Khelifa M, Kharouf M, Coutton C, Merdassi G, Abada F, Escoffier J, Nikas Y *et al* (2011) A recurrent deletion of DPY19L2 causes infertility in man by blocking sperm head elongation and acrosome formation. *Am J Hum Genet* 88: 351 – 361

Hartshorne G, Montgomery S, Klentzeris L (1999) A case of failed oocyte maturation *in vivo* and *in vitro*. *Fertil Steril* 71: 567 – 570

He XQ, Song YQ, Liu R, Liu Y, Zhang F, Zhang Z, Shen YT, Xu L, Chen MH, Wang YL *et al* (2016) Axin-1 regulates meiotic spindle organization in mouse oocytes. *PLoS ONE* 11: e0157197

Hourvitz A, Maman E, Brengauz M, Machtinger R, Dor J (2010) *In vitro* maturation for patients with repeated *in vitro* fertilization failure due to "oocyte maturation abnormalities". *Fertil Steril* 94: 496 – 501

Huang YL, Anvarian Z, Doderlein G, Acebron SP, Niehrs C (2015) Maternal Wnt/STOP signaling promotes cell division during early *Xenopus* embryogenesis. *Proc Natl Acad Sci USA* 112: 5732 – 5737

Jensen AB, Tunster SJ, John RM (2014) The significance of elevated placental PHLDA2 in human growth restricted pregnancies. *Placenta* 35: 528 – 532

Kaya MS, Kose M, Guzeloglu A, Kiyma Z, Atli MO (2017) Early pregnancy-related changes in toll-like receptors expression in ovine trophoblasts and peripheral blood leukocytes. *Theriogenology* 93: 40 – 45

Kherraf ZE, Conne B, Amiri-Yekta A, Christou-Kent M, Coutton C, Escoffier J, Nef S, Arnoult C, Ray PF (2018) Combining whole exome sequencing and CRISPR-Ca9 technology to study male infertility and spermatogenesis defects. *Mol Cell Endocrinol* https://doi.org/10.1016/j.mce.2018.03.002

Kim JY (2012) Control of ovarian primordial follicle activation. *Clin Exp Reprod Med* 39: 10 – 14

Kim SY, Kim SK, Lee JR, Woodruff TK (2016) Toward precision medicine for preserving fertility in cancer patients: existing and emerging fertility preservation options for women. *J Gynecol Oncol* 27: e22

Lakso M, Pichel JG, Gorman JR, Sauer B, Okamoto Y, Lee E, Alt FW, Westphal H (1996) Efficient *in vivo* manipulation of mouse genomic sequences at the zygote stage. *Proc Natl Acad Sci USA* 96: 5860 – 5865

Lee J, Oh JS, Cho C (2011) Impaired expansion of trophoblast spheroids cocultured with endometrial cells overexpressing cellular retinoic acid-binding protein 2. *Fertil Steril* 95: 2599 – 2601

Levran D, Farhi J, Nahum H, Glezerman M, Weissman A (2002) Maturation arrest of human oocytes as a cause of infertility: case report. *Hum Reprod* 17: 1604 – 1609

Li R, Albertini DF (2013) The road to maturation: somatic cell interaction and self-organization of the mammalian oocyte. *Nat Rev Mol Cell Biol* 14: 141 – 152

Li J, Sun P, Yue Z, Zhang D, You K, Wang J (2017) miR-144-3p induces cell cycle arrest and apoptosis in pancreatic cancer cells by targeting proline-rich protein 11 expression via the mitogen-activated protein kinase signaling pathway. *DNA Cell Biol* 36: 619 – 626

Liang CG, Su YQ, Fan HY, Schatten H, Sun QY (2007) Mechanisms regulating oocyte meiotic resumption: roles of mitogen-activated protein kinase. *Mol Endocrinol* 21: 2037 – 2055

Libby BJ, De La Fuente R, O'Brien MJ, Wigglesworth K, Cobb J, Inselman A, Eaker S, Handel MA, Eppig JJ, Schimenti JC (2002) The mouse meiotic mutation mei1 disrupts chromosome synapsis with sexually dimorphic consequences for meiotic progression. *Dev Biol* 242: 174 – 187

Lincoln AJ, Wickramasinghe D, Stein P, Schultz RM, Palko ME, De Miguel MP, Tessarollo L, Donovan PJ (2002) Cdc25b phosphatase is required for resumption of meiosis during oocyte maturation. *Nat Genet* 30: 446 – 449

Lipkin SM, Moens PB, Wang V, Lenzi M, Shanmugarajah D, Gilgeous A, Thomas J, Cheng J, Touchman JW, Green ED *et al* (2002) Meiotic arrest and aneuploidy in MLH3-deficient mice. *Nat Genet* 31: 385 – 390

Maddirevula S, Coskun S, Alhassan S, Elnour A, Alsaif HS, Ibrahim N, Abdulwahab F, Arold ST, Alkuraya FS (2017) Female infertility caused by mutations in the oocyte-specific translational repressor PATL2. *Am J Hum Genet* 101: 603 – 608

Marnef A, Maldonado M, Bugaut A, Balasubramanian S, Kress M, Weil D, Standart N (2010) Distinct functions of maternal and somatic Pat1 protein paralogs. *RNA* 16: 2094 – 2107

Mayer S, Wrenzycki C, Tomek W (2014) Inactivation of mTor arrests bovine oocytes in the metaphase-I stage, despite reversible inhibition of 4E-BP1 phosphorylation. *Mol Reprod Dev* 81: 363 – 375

Medvedev S, Pan H, Schultz RM (2011) Absence of MSY2 in mouse oocytes perturbs oocyte growth and maturation, RNA stability, and the transcriptome. *Biol Reprod* 85: 575 – 583

Nakamura Y, Tanaka KJ, Miyauchi M, Huang L, Tsujimoto M, Matsumoto K (2010) Translational repression by the oocyte-specific protein P100 in *Xenopus*. *Dev Biol* 344: 272 – 283

Nakanishi T, Ozaki Y, Blomgren K, Tateyama H, Sugiura-Ogasawara M, Suzumori K (2005) Role of cathepsins and cystatins in patients with recurrent miscarriage. *Mol Hum Reprod* 11: 351 – 355

Paciolla M, Boni R, Fusco F, Pescatore A, Poeta L, Ursini MV, Lioi MB, Miano MG (2011) Nuclear factor-kappa-B-inhibitor alpha (NFKBIA) is a developmental marker of NF-kappaB/p65 activation during *in vitro* oocyte maturation and early embryogenesis. *Hum Reprod* 26: 1191 – 1201

Paczkowski M, Krisher R (2010) Aberrant protein expression is associated with decreased developmental potential in porcine cumulus-oocyte complexes. *Mol Reprod Dev* 77: 51 – 58

Pal D, Wu D, Haruta A, Matsumura F, Wei Q (2010) Role of a novel coiled-coil domain-containing protein CCDC69 in regulating central spindle assembly. *Cell Cycle* 9: 4117 – 4129

Pique M, Lopez JM, Foissac S, Guigo R, Mendez R (2008) A combinatorial code for CPE-mediated translational control. *Cell* 132: 434 – 448

Racki WJ, Richter JD (2006) CPEB controls oocyte growth and follicle development in the mouse. *Development* 133: 4527 – 4537

Rausell F, Pertusa JF, Gomez-Piquer V, Hermenegildo C, Garcia-Perez MA, Cano A, Tarin JJ (2007) Beneficial effects of dithiothreitol on relative levels of glutathione S-transferase activity and thiols in oocytes, and cell number, DNA fragmentation and allocation at the blastocyst stage in the mouse. *Mol Reprod Dev* 74: 860 – 869

Ruiz-Gonzalez I, Minten M, Wang X, Dunlap KA, Bazer FW (2015) Involvement of TLR7 and TLR8 in conceptus development and establishment of pregnancy in sheep. *Reproduction* 149: 305 – 316

Ryu KY, Sinnar SA, Reinholdt LG, Vaccari S, Hall S, Garcia MA, Zaitseva TS, Bouley DM, Boekelheide K, Handel MA *et al* (2008) The mouse polyubiquitin gene Ubb is essential for meiotic progression. *Mol Cell Biol* 28: 1136 – 1146

SEQC/MAQC-III Consortium (2014) A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *NatBiotechnol* 32: 903 – 914

Shi YQ, Zhuang XJ, Xu B, Hua J, Liao SY, Shi Q, Cooke HJ, Han C (2013) SYCP3-like X-linked 2 is expressed in meiotic germ cells and interacts with synaptonemal complex central element protein 2 and histone acetyltransferase TIP60. *Gene* 527: 352 – 359

Shin YH, Ren Y, Suzuki H, Golnoski KJ, Ahn HW, Mico V, Rajkovic A (2017) Transcription factors SOHLH1 and SOHLH2 coordinate oocyte differentiation without affecting meiosis I. *J Clin Invest* 127: 2106 – 2117

Solc P, Saskova A, Baran V, Kubelka M, Schultz RM, Motlik J (2008) CDC25A phosphatase controls meiosis I progression in mouse oocytes. *Dev Biol* 317: 260 – 269

Sousa Martins JP, Liu X, Oke A, Arora R, Franciosi F, Viville S, Laird DJ, Fung JC, Conti M (2016) DAZL and CPEB1 regulate mRNA translation synergistically during oocyte maturation. *J Cell Sci* 129: 1271 – 1282

Spate LD, Brown AN, Redel BK, Whitworth KM, Murphy CN, Prather RS (2014) Dickkopf-related protein 1 inhibits the WNT signaling pathway and improves pig oocyte maturation. *PLoS ONE* 9: e95114

Stopa N, Krebs JE, Shechter D (2015) The PRMT5 arginine methyltransferase: many roles in development, cancer and beyond. *Cell Mol Life Sci* 72: 2041 – 2059

Su YQ, Sugiura K, Woo Y, Wigglesworth K, Kamdar S, Affourtit J, Eppig JJ (2007) Selective degradation of transcripts during meiotic maturation of mouse oocytes. *Dev Biol* 302: 104 – 117

Swetloff A, Conne B, Huarte J, Pitetti JL, Nef S, Vassalli JD (2009) Dcp1-bodies in mouse oocytes. *Mol Biol Cell* 20: 4951 – 4961

Takabayashi S, Yamauchi Y, Tsume M, Noguchi M, Katoh H (2009) A spontaneous smc1b mutation causes cohesin protein dysfunction and sterility in mice. *Exp Biol Med* 234: 994 – 1001

Vaccari S, Horner K, Mehlmann LM, Conti M (2008) Generation of mouse oocytes defective in cAMP synthesis and degradation: endogenous cyclic AMP is essential for meiotic arrest. *Dev Biol* 316: 124 – 134

Wilczek C, Chitta R, Woo E, Shabanowitz J, Chait BT, Hunt DF, Shechter D (2011) Protein arginine methyltransferase Prmt5-Mep50 methylates histones H2A and H4 and the histone chaperone nucleoplasmin in *Xenopus laevis* eggs. *J Biol Chem* 286: 42221 – 42231

Williams CJ, Erickson GF (2012) *Morphology and physiology of the ovary*. South Dartmouth, MA: MDText.com, Inc.; 2000

Yin O, Cayton K, Segars JH (2016) *In vitro* activation: a dip into the primordial follicle pool? *J Clin Endocrinol Metab* 101: 3568 – 3570

Yu J, Hecht NB, Schultz RM (2001) Expression of MSY2 in mouse oocytes and preimplantation embryos. *Biol Reprod* 65: 1260 – 1270

Zhang QH, Qi ST, Wang ZB, Yang CR, Wei YC, Chen L, Ouyang YC, Hou Y, Schatten H, Sun QY (2012) Localization and function of the Ska complex during mouse oocyte meiotic maturation. *Cell Cycle* 11: 909 – 916

Expanded View Figures

Figure EV1. Quality of oocytes collected from patients harbouring *PATL2* mutation and control patients after ovarian stimulation.

A The mean age of the six patients harbouring a *PATL2* mutation at the time of 11 hormonal stimulations was compared to the non-*PATL2* patients within the cohort and also to a control cohort corresponding to women from infertile couples of similar geographical origin where the male was diagnosed with a male infertility (mean ± SEM, n = 234). There was no significant (NS) age difference between the *PATL2* patients, the non-*PATL2* patients and the control cohort.

B Numbers of oocytes retrieved after hormonal stimulation (mean ± SEM) were similar in *PATL2* patients, non-*PATL2* patients and the control cohort.

C Collected oocytes were sorted according to their maturation stage. For patients harbouring *PATL2* mutation, the mean numbers of GV and atretic oocytes were significantly increased and no MII oocytes were collected. Non-*PATL2* patients from the same cohort showed a comparably larger proportion of MI-arrested oocytes.

Data information: Statistical differences were assessed using unpaired two-tailed t-tests.

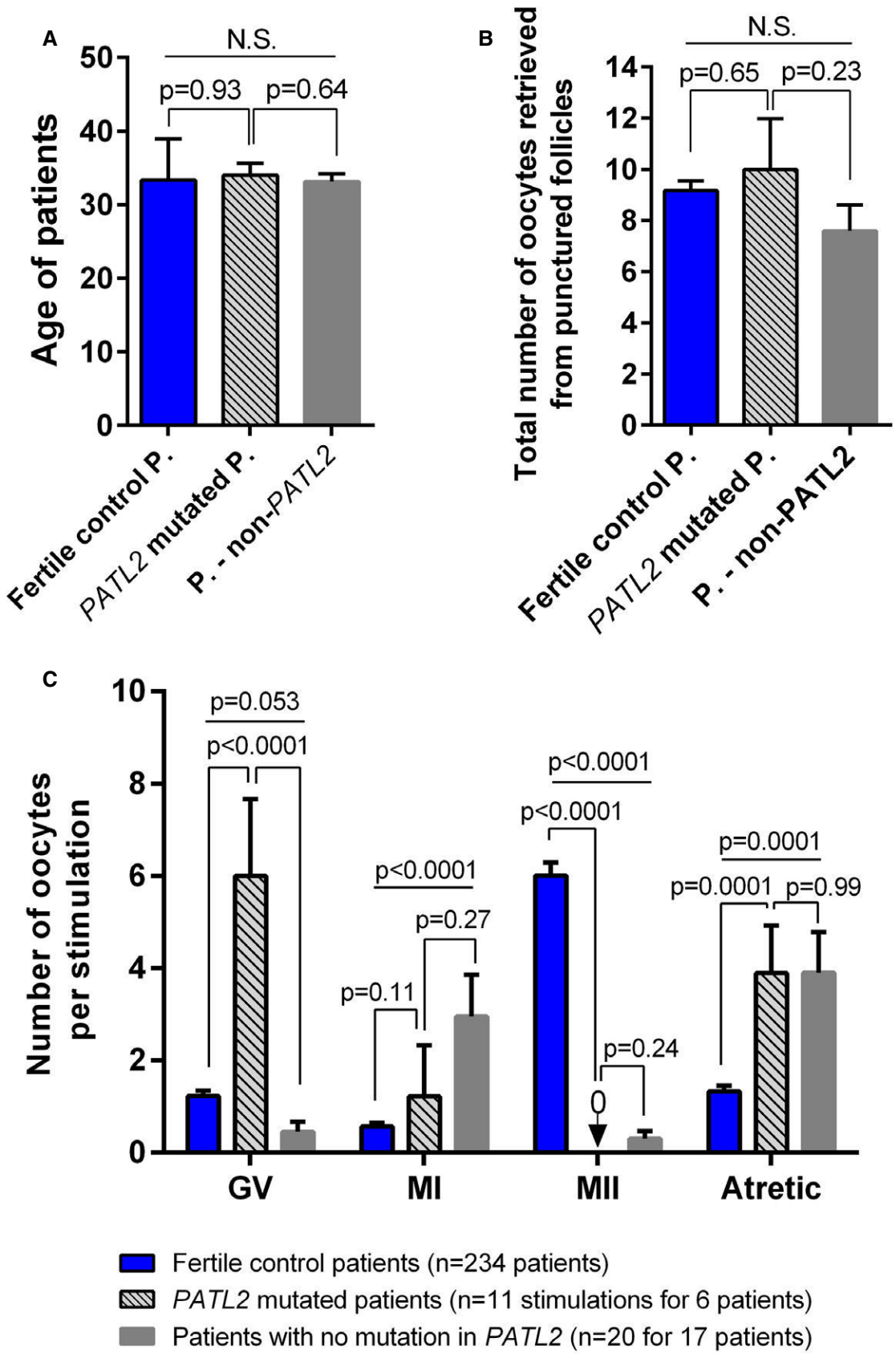


Figure EV1.

Figure EV2. Histological analysis of ovaries from control and *Patl2*$^{-/-}$ females at 26 days postpartum (dpp).

A Whole section (3 μm thick) of an ovary from a control female at 26 ddp. Section was stained with an antibody against Msy2 (orange staining) to make oocytes clearly visible. Sections were counterstained with Hoechst to reveal the nucleus (blue staining). Right image, corresponding to the enlargement of the red square on the left image, shows the different classes of follicle: primordial (1al), primary (1ry) and secondary (2ry).

B Similar images for an ovary from a *Patl2*$^{-/-}$ female at 26 ddp.

C Comparative numbers of primordial and primary follicle oocytes per section in *Patl2*$^{-/-}$ and control ovaries. Nine different 3-μm sections from three different mice (for each section, four-seven technical replicates corresponding to successive sections were counted). Data are presented on box and whisker plots indicating min. and max. values. Statistical differences were assessed using *t*-test, *P*-value as indicated.

D Comparative number of follicles per section from control and *Patl2*$^{-/-}$ ovaries. Only follicles where oocytes were visible were counted (n = 9 sections per genotype, between four to seven technical replicates). Data are presented on box and whisker plots indicating min. and max. values. Statistical differences were assessed using *t*-test, *P*-value as indicated.

E Comparative amplitude histograms of follicle size per section from control and *Patl2*$^{-/-}$ ovaries (n = 9 sections per genotype with four to seven technical replicates). Statistical differences were assessed using Kolmogorov-Smirnov test.

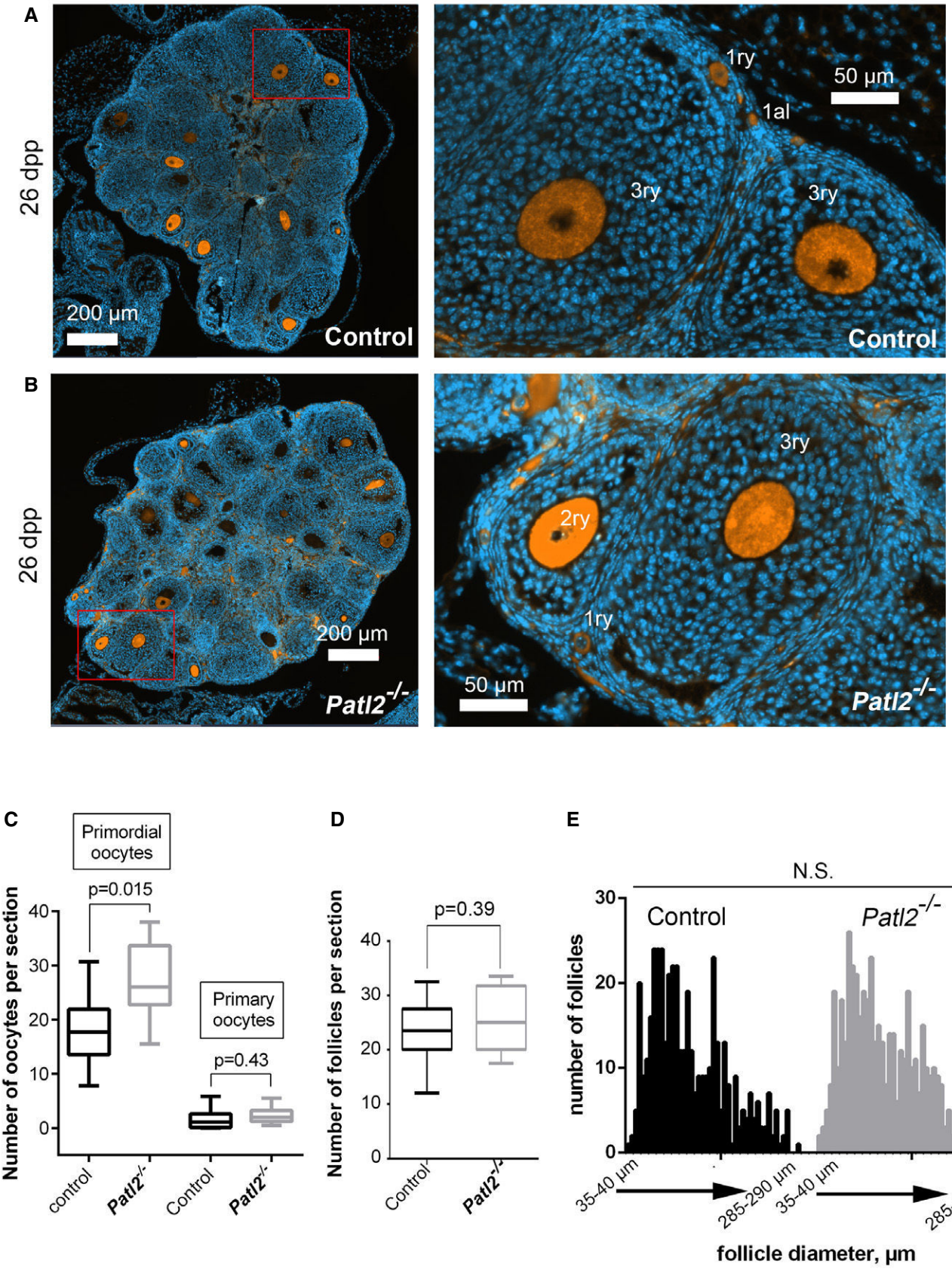


Figure EV2.

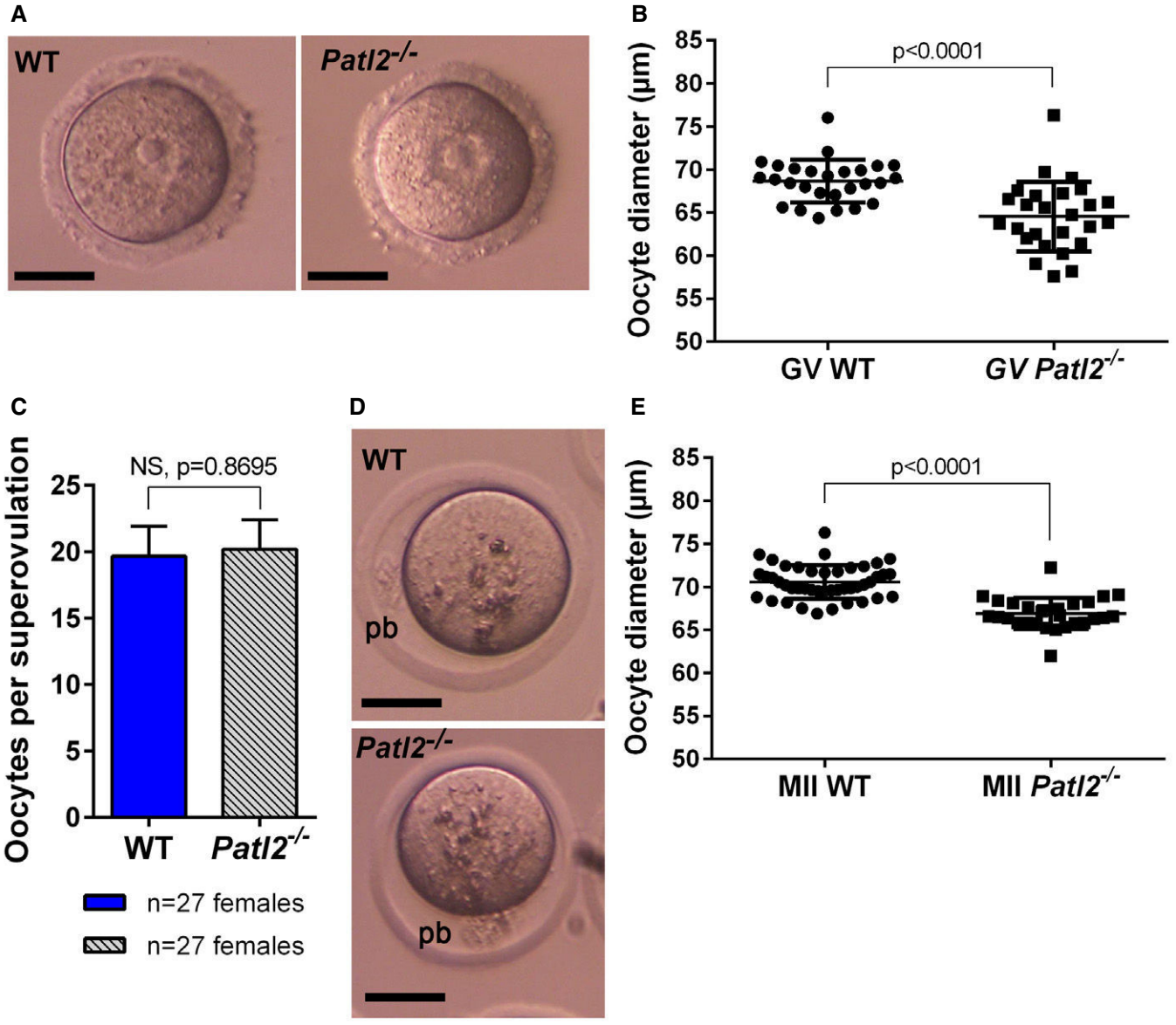


Figure EV3. Lack of Patl2 does not affect number of oocytes produced, but impairs oocyte growth.

A Representative images for a GV oocyte from WT and *Patl2*$^{-/-}$ females stimulated with 5 UI PSMG. Scale bars = 35 µm.

B Comparison of WT and *Patl2*$^{-/-}$ GV oocyte diameter. In the absence of Patl2, the diameter (mean ± SEM) dropped from 68.6 ± 0.5 to 64.5 ± 0.8 µm ($P < 0.0001$). Statistical difference was assessed using unpaired two-tailed *t*-test with Welch's correction.

C The number of oocytes (mean ± SEM) harvested 13 h after full hormonal stimulation (PMSG followed 48 h later by 5 UI HCG) was similar in WT and Patl2-deficient females. Statistical difference was assessed using unpaired two-tailed *t*-test.

D Images of a MII oocytes from WT and *Patl2*$^{-/-}$ females after stimulation (pb = 1st polar body). Scale bars = 35 µm.

E Diameter of WT and *Patl2*$^{-/-}$ MII oocytes (mean ± SEM), the mean drops from 70.6 ± 0.3, $n = 44$ in the WT to 66.9 ± 0.3, $n = 28$ in *Patl2*$^{-/-}$ ($P < 0.0001$), oocytes were collected from three females in each case. Statistical difference was assessed using unpaired two-tailed *t*-test.

Figure EV4. Biological functions and pathways associated with up- or down-regulated differential expression transcripts in *Patl2*$^{-/-}$ versus WT samples at the GV and MII stages.

To investigate possible interactions of up and down-regulated transcripts ($P < 0.05$, absolute fold-change > 2), genes with altered expression profile identified by the Affymetrix microarray were imported into the Ingenuity Pathway Analysis software (IPA) for analysis.

A Canonical pathways identified by IPA that were significantly enriched among transcripts deregulated in *Patl2*$^{-/-}$ GV oocytes with respect to GV-WT oocytes.
B Canonical pathways identified by IPA that were significantly enriched among transcripts deregulated in *Patl2*$^{-/-}$ MII oocytes with respect to MII-WT oocytes.

Data information: *Y*-axis indicates the significance (−log *P*-value) of the functional pathway association, which depends on the number of genes in a class as well as biological relevance. The threshold line represents a *P*-value of 0.05 and was calculated by applying Fischer's test.

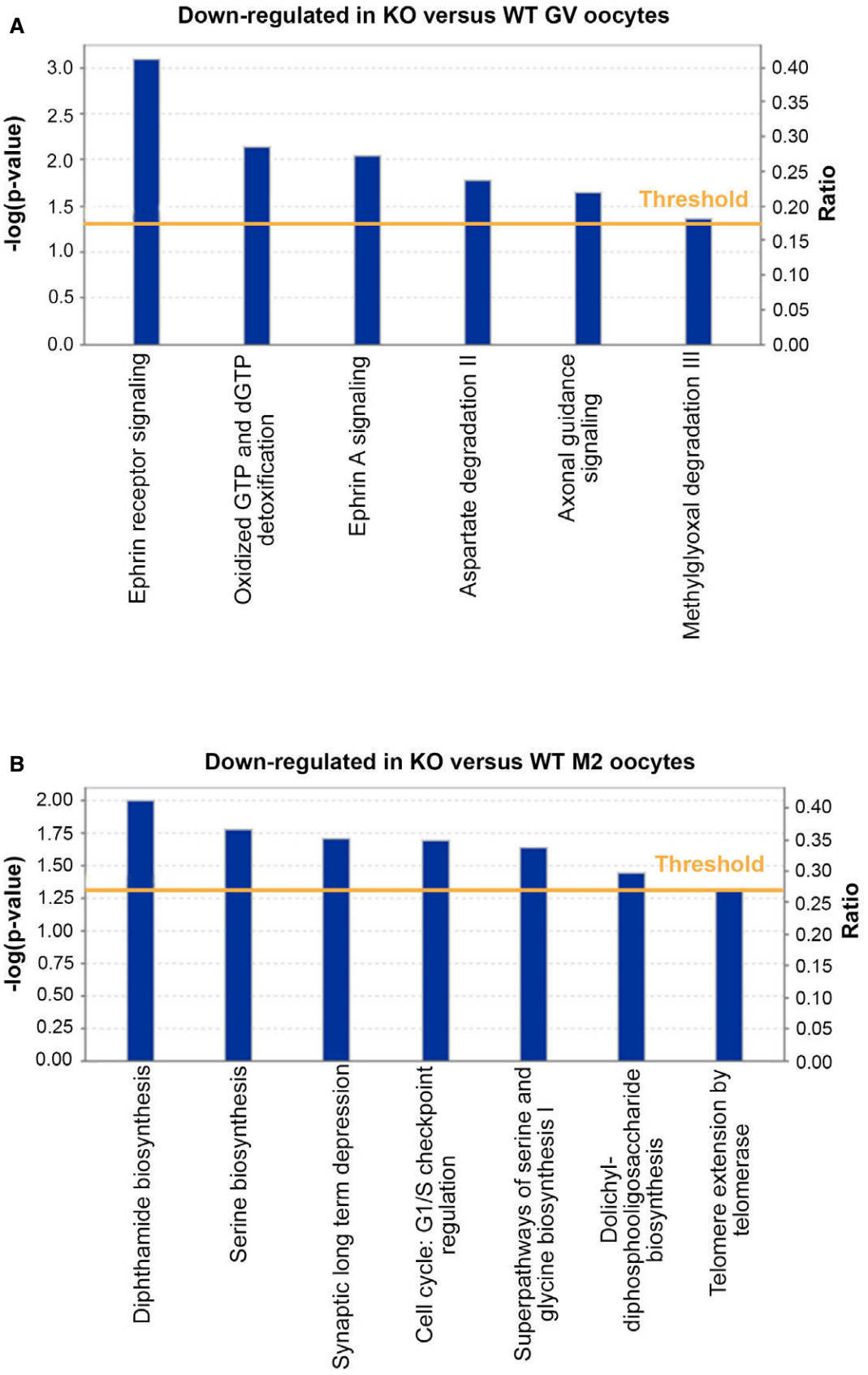

A
Down-regulated in KO versus WT GV oocytes
-log(p-value)
Ratio
Threshold
Ephrin receptor signaling
Oxidized GTP and dGTP detoxification
Ephrin A signaling
Aspartate degradation II
Axonal guidance signaling
Methylglyoxal degradation III
B
Down-regulated in KO versus WT M2 oocytes
-log(p-value)
Ratio
Threshold
Diphthamide biosynthesis
Serine biosynthesis
Synaptic long term depression
Cell cycle: G1/S checkpoint regulation
Superpathways of serine and glycine biosynthesis I
Dolichyl-diphosphooligosaccharide biosynthesis
Telomere extension by telomerase

Figure EV4.

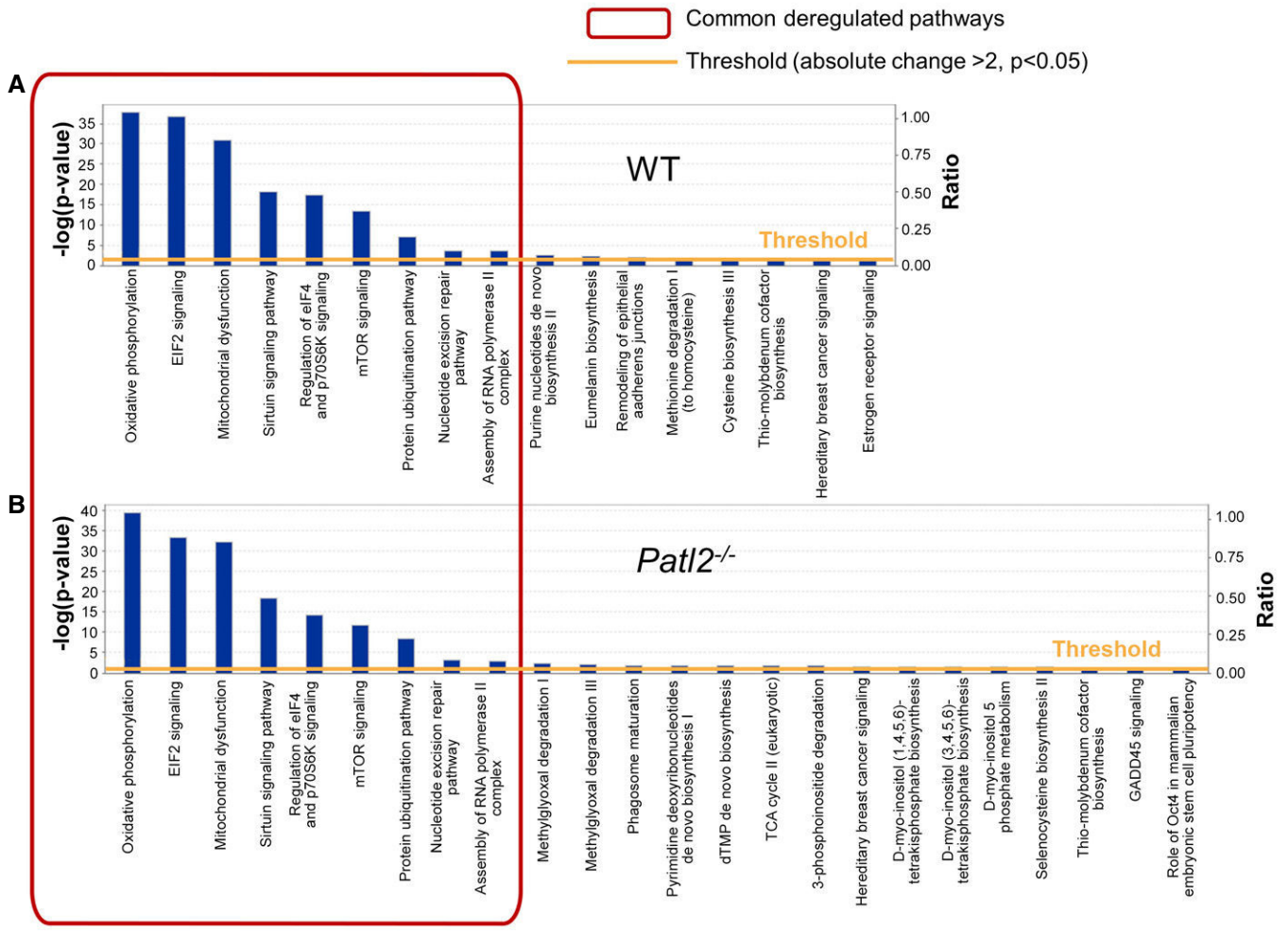


Figure EV5. Transcriptomic analysis of the GV-MII transition. Biological functions and pathways associated with up- or down-regulated differential expression in MII versus GV oocytes from WT and *Patl2*$^{-/-}$ samples.

To investigate possible interactions of up- and down-regulated transcripts ($P < 0.05$, absolute fold-change > 2) between GV and MII stages, genes with an altered expression profile identified by the Affymetrix microarray were imported into the Ingenuity Pathway Analysis software (IPA) for analysis.

A Canonical pathways identified by IPA that were significantly enriched among changed transcripts identified in WT samples at the GV-MII transition.
B Canonical pathways identified by IPA that were significantly enriched among changed transcripts identified in *Patl2*$^{-/-}$ samples at the GV-MII transition.

Data information: *Y*-axis indicates the significance (−log *P*-value) of the functional pathway association, which depends on the number of genes in a class as well as biological relevance. The threshold line represents a *P*-value of 0.05 and was calculated by applying Fischer's test.

APPENDIX

PATL2 is a key actor of oocyte maturation whose invalidation causes infertility in women and mice

Table of contents

SUPPLEMENTARY FIGURES 2
Figure S1. 2
Figure S2. 3
Figure S3. 4
Figure S4 5
Figure S5. 6
Figure S6. 7
Figure S7. 8
Figure S8. 9
Figure S9. 10
SUPPLEMENTARY TABLES 11
Table S1 11
Table S2 12

SUPPLEMENTARY FIGURES

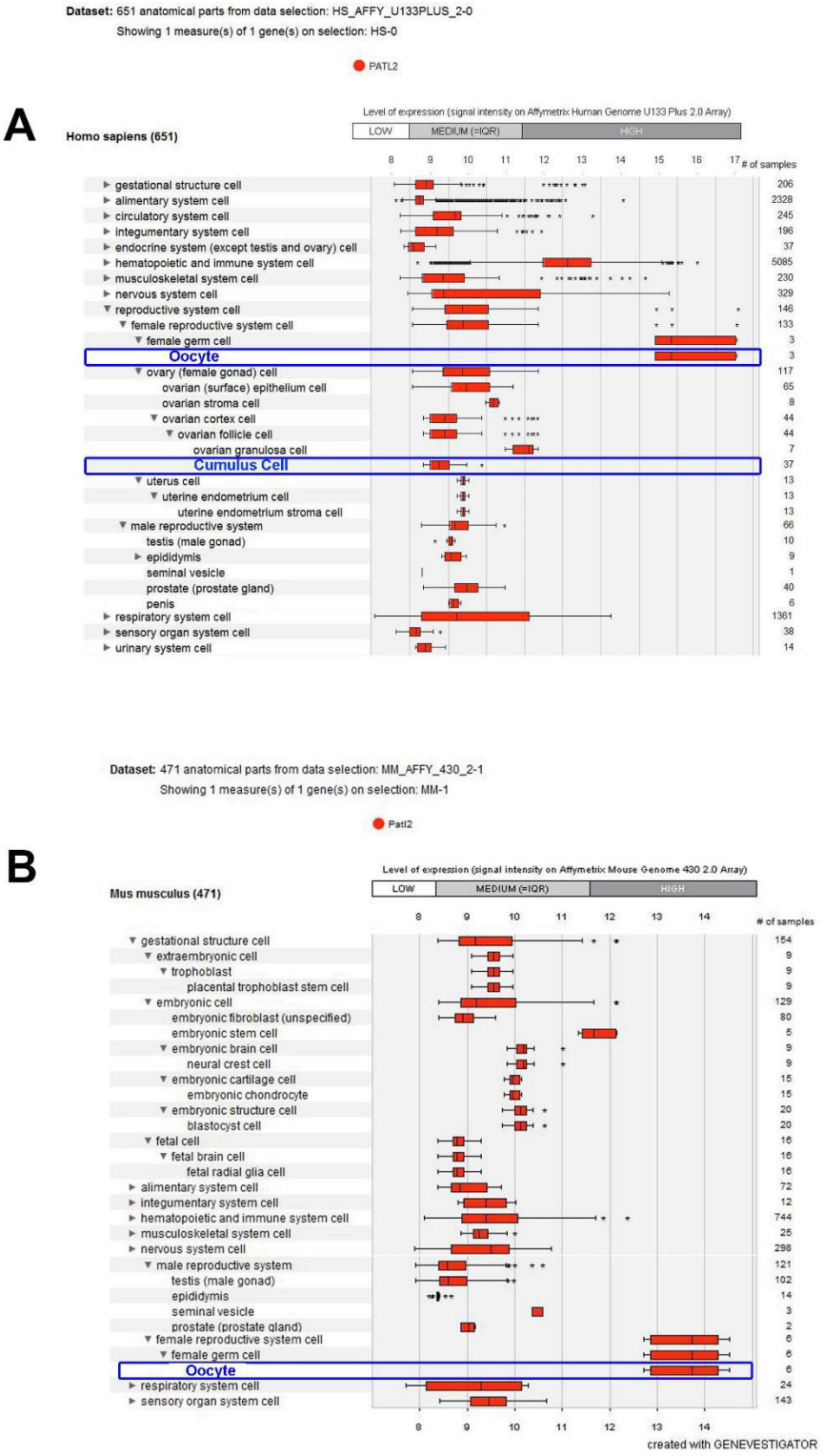


Figure S1. **Relative mRNA expression levels for human and mouse *PATL2* transcripts.**

(A) *PATL2* mRNA levels were measured in different tissues/cells in Humans using Affymetrix microarrays (data available from the Genevestigator database, https://genevestigator.com). Blue rectangles highlight the expression level in oocytes and follicular cells. Oocytes exhibit the highest level of *PATL2* expression, whereas its level is very low in follicular cells. **(B)** Similar data for mice. Data were generated with Genevestigator (Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, Oertle L, Widmayer P, Gruissem W and P Zimmermann (2008) Genevestigator V3: a reference expression database for the meta-analysis of transcriptomes. Advances in Bioinformatics 2008, 420747)

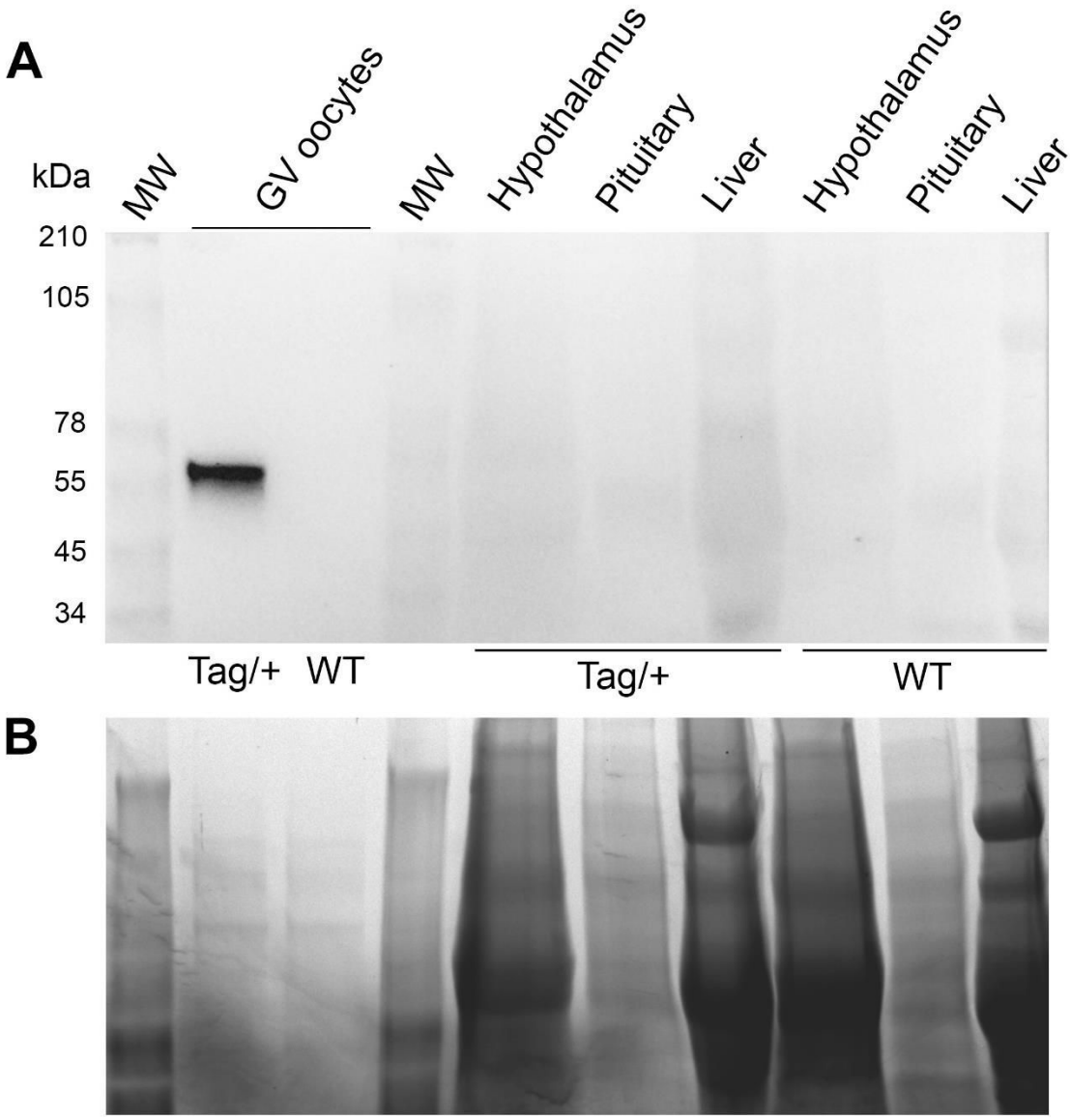


Figure S2. **Western blot comparison of expression of Patl2-HA in oocyte, hypothalamus and pituitary gland extracts.**

(**A**) Western blot of protein extracts from different tissues as indicated and revealed with anti-HA antibodies. Tissues were collected from WT or from Patl2-HA tagged females (tag/+) and 130-140 GV oocytes were loaded per well. Protein extracts were obtained by pooling the organs of 3 different WT or Patl2-HA tagged animals. In the lane loaded with GV oocyte extract from Patl2-HA tag mice, a band at around 60 kDa was observed (expected MW for mouse Patl2-HA = 60.74 KDa), whereas no bands were observed in the other extracts. MW = Molecular Weight (**B**) Protein loads of the Western blot presented in (A) were controlled with TGX stain free™ precast gels.

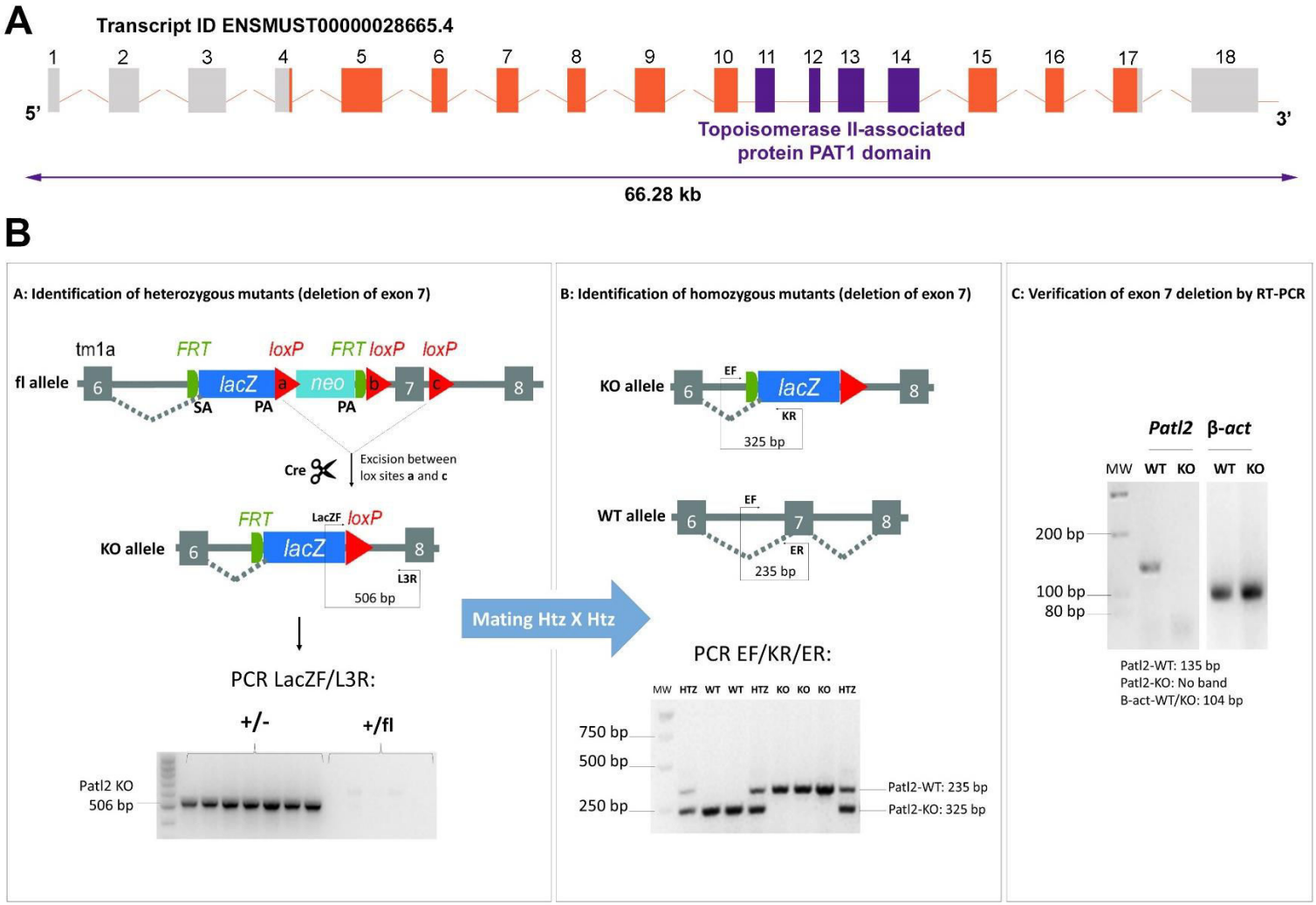


Figure S3. **Genetic modification of *Patl2*-deficient mice and genotyping strategy**

(A) Exon map showing the localization of the topoisomerase II associated Pat1 domain, downstream of exon 7 **(B1)** *Patl2*$^{-/-}$ mice (C57BL/6NTac-Patl2<tm1a) were generated by the EUCOMM Consortium (http://www.mousephenotype.org/about-ikmc/eucomm). The allele map shows the insertion of a LacZ and a neomycin cassette between exons 6 and 7. The critical exon 7 is surrounded by two LoxP sequences. *Patl2*$^{-/-}$ mice were crossed with adult heterozygous EIIaCre transgenic mice which express the CRE-recombinase enzyme ubiquitously from an early stage. Offspring were screened by PCR using LacZF/L3R primers to detect exon 7 deletion. **(B2)** Heterozygous mutant mice showing deletion of exon 7 were mated together and homozygous exon-7-deleted animals were identified by multiplex PCR using EF/KR/ER primers. **(B3)** RT-PCR was carried out on selected F2 animals to verify deletion of exon 7 in *Patl2*$^{-/-}$ animals using primers within exon 7. The band at 135 bp was sequenced and corresponded to *Patl2*. Sequences of all primers are indicated in Table S6.

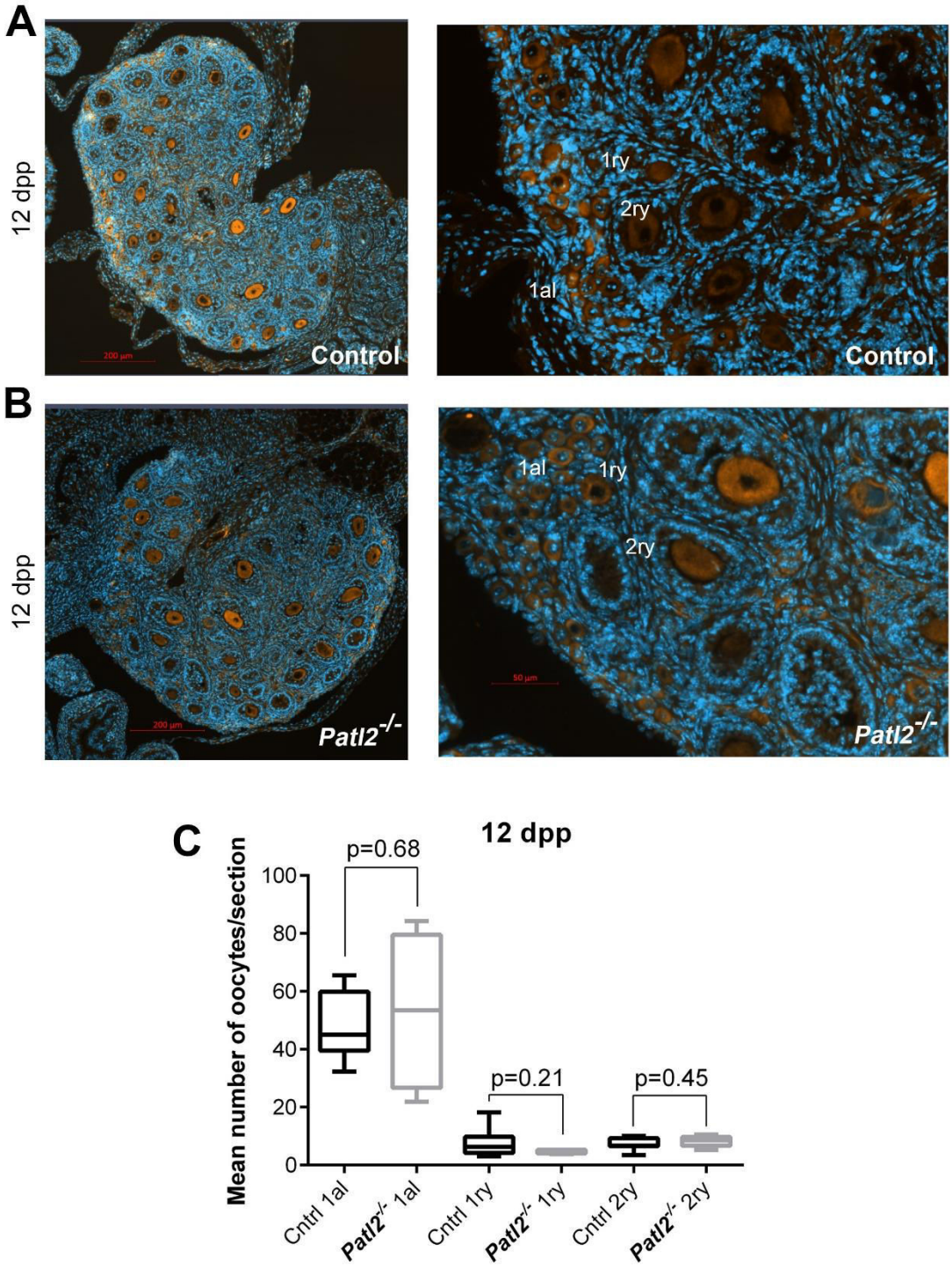


Figure S4 **Ovaries from WT and *Patl2* deficient females have a similar anatomical structure and cellular composition at 12 dpp.**

12-day-old females were euthanised and ovaries were collected and subjected to histological studies by IF. Sections (3 µm thick) were stained with Hoechst to reveal nuclei (blue staining) and an anti-Msy2 antibody (orange staining) to identify oocytes. All oocyte stages were stained by anti-Msy2 **(A)** WT ovaries and **(B)** *Patl2*$^{-/-}$ ovaries. No obvious morphological differences were observed between WT and *Patl2*$^{-/-}$ ovaries. (**C**) The different classes of oocytes (primordial, primary and secondary) were counted in 9 different 3-µm sections from 3 different mice (for each section, 4-7 technical replicates, corresponding to successive sections were counted). n=9, Statistical differences were measured with t-test. P value as indicated.

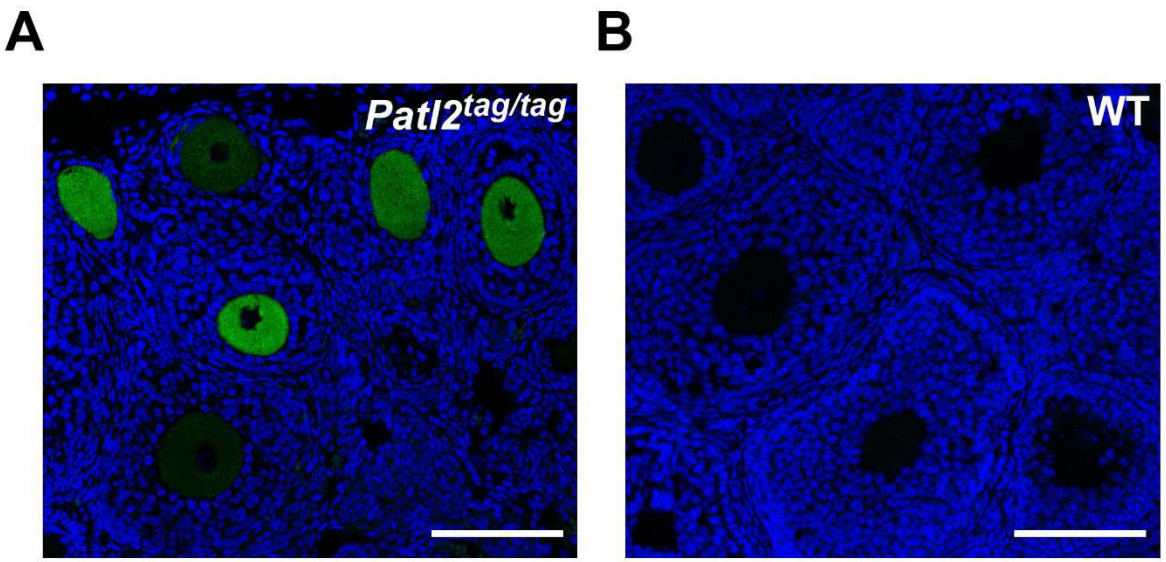


Figure S5. **Validation of the specificity of the HA-antibody in IF experiments**

Comparative fluorescent signals of sections of ovaries from HA-tagged Patl2 (A) and WT (B) mice. Ovaries were stained with Hoechst and an anti-HA tag antibody and observed under confocal microscopy.

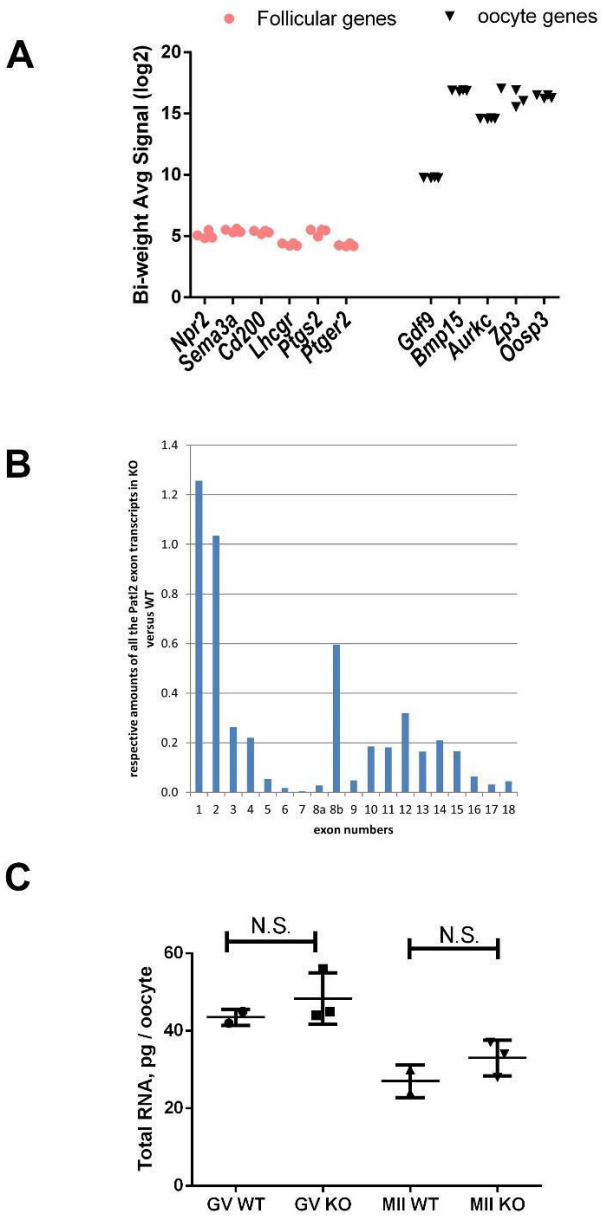


Figure S6. **Characterisation of the RNA sample used in transcriptomic approach.**

(A) Expression levels for specific follicular and oocyte genes. The expression levels for follicle-specific genes were around 5, which is considered a very low level. In contrast, the expression levels for oocyte genes were above 9 and reached the highest values (17). The follicular genes analysed were: Natriuretic peptide receptor 2 (*Npr2*), semaphorin 3A (*Sema3a*), CD200 antigen (C*d200*), luteinizing hormone/choriogonadotropin receptor (*Lhcgr*), prostaglandin G/H synthase 2 (*Ptgs2*) and prostaglandin E receptor 2 (*Ptger*2). The oocyte-specific genes were: growth differentiation factor 9 (*Gdf9*), bone morphogenetic protein 15 (*Bmp15*), aurora kinase C (*Aurkc*), zona pellucida glycoprotein 3 (*Zp3*), oocyte secreted protein 3 (*Oosp3*). **(B)** Comparative expression levels for the different *Patl2* exons. As expected, the level of exon 7 is close to 0, confirming deletion of exon 7 in knock-out females. The expression levels for all 19 Patl2 exons (according to the exon map nomenclature from Affymetrix) were expressed as the ratio (y axis and in linear scale) between exon transcript levels in Patl2$^{-/-}$ oocytes compared to wild type oocytes. In *Patl2*$^{-/-}$ oocytes, *Patl2* exons 1 and 2 were expressed in nearly identical amounts relative to levels measured for wild type oocytes. All the other exons were more or less severely underexpressed or absent. **(C)** Total RNA was extracted from WT and *Patl2*$^{-/-}$ oocytes (both at GV and MII stages) as described in materials and methods. After RT-qPCR assays, the RNA amounts recovered from the purification columns were expressed in pg purified RNA per lysed oocyte. Assuming a consistent RNA recovery yield from the columns, we can conclude that there are non-significant differences in RNA concentrations per oocyte between WT and *Patl2*$^{-/-}$ GV oocytes, as well as between WT and *Patl2*$^{-/-}$ MII oocytes. However, there is a nearly 30 to 40% decrease in the RNA level between MII stage oocytes and GV stage oocytes, as already reported (Su YQ, Sugiura K, Woo Y, et al. Selective degradation of transcripts during meiotic maturation of mouse oocytes. Developmental biology 2007;302:104-17).

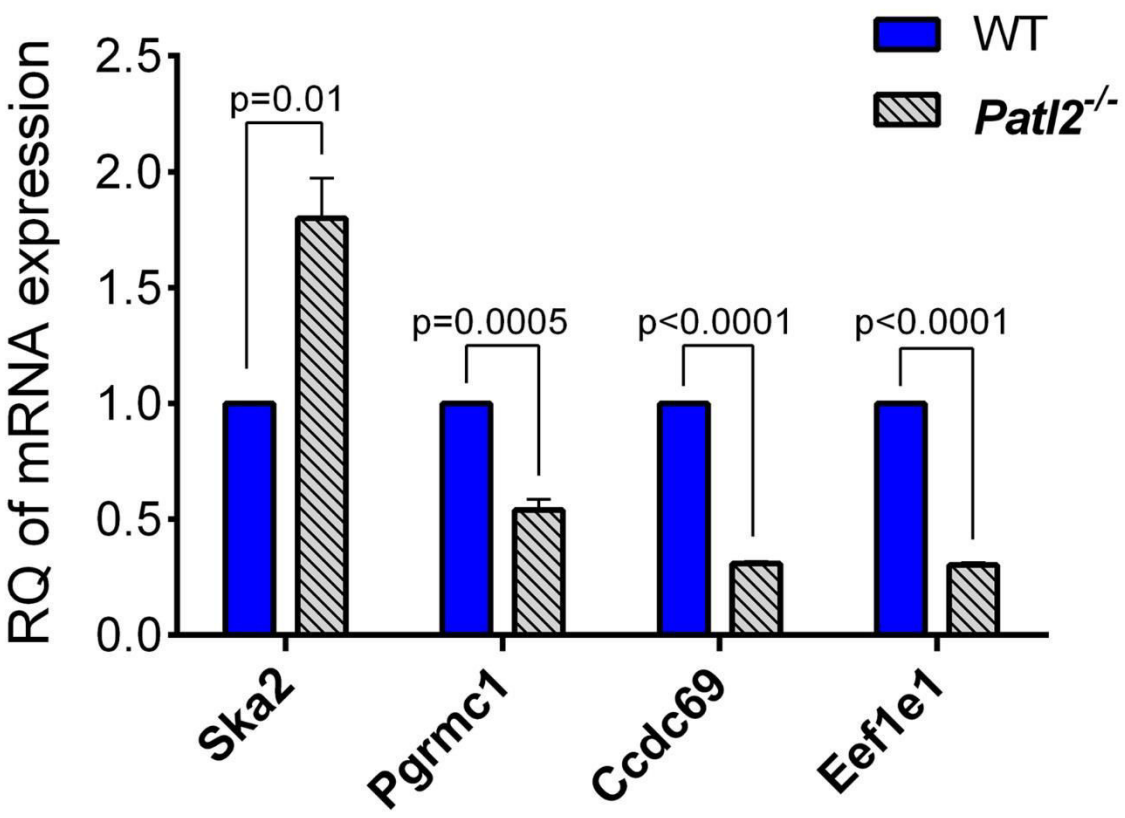


Figure S7. **RT-qPCR experiments validating Affymetrix measures**

Relative quantification (RQ) of 4 genes (*Ska2*, *Pgrmc1, Ccdc69* and *Eef1e1*) in germinal vesicle (GV) oocytes from both wild-type (blue) and *Patl2*$^{-/-}$ (grey) mice. n = 8.

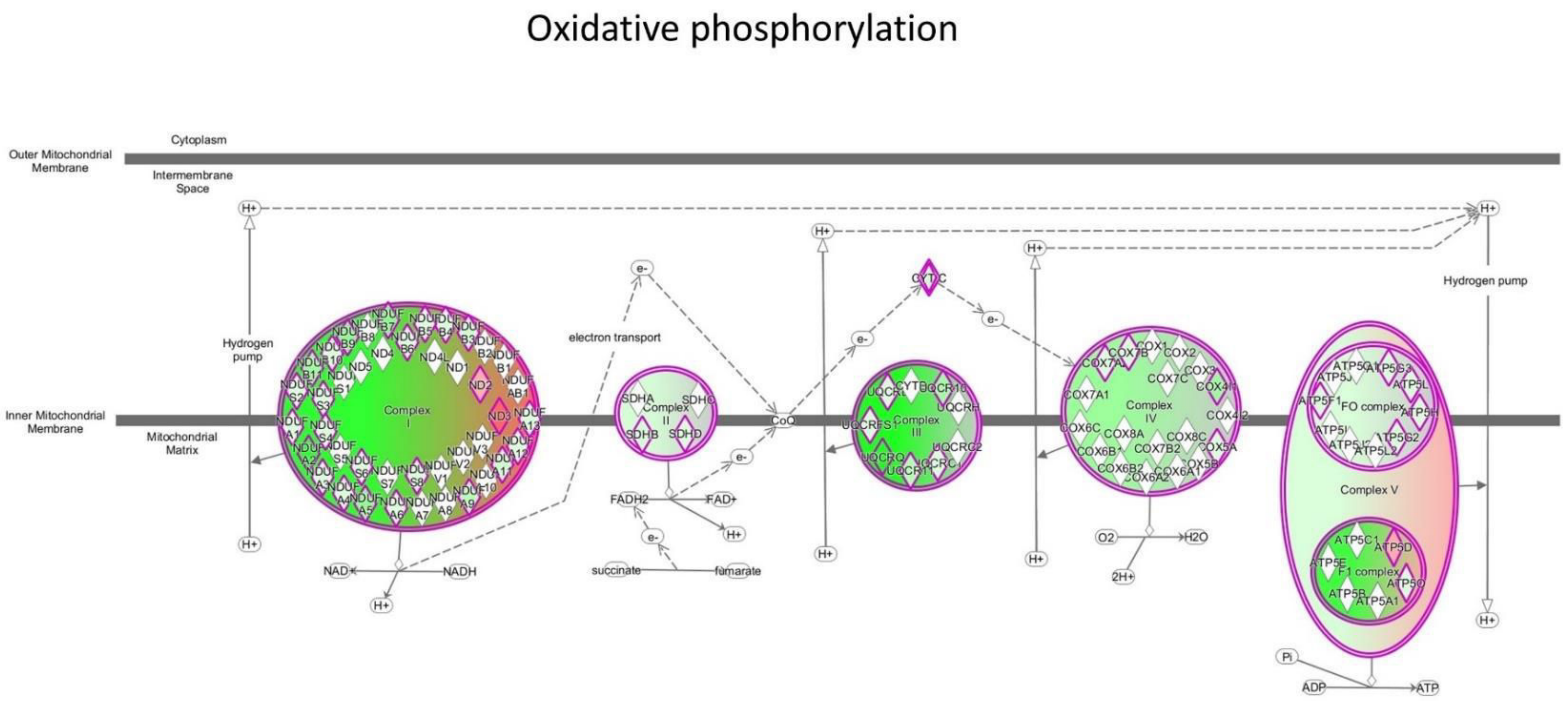


Figure S8. **Deregulated transcripts related to the oxidative phosphorylation pathway**. The interaction network was generated using Ingenuity software and shows that many genes from the oxidative phosphorylation pathway (the major pathway for conversion of energy from NADH oxidation into ATP) are deregulated in both WT and KO samples. Up-regulated transcripts are shown in pink and down-regulated transcripts are indicated in green.

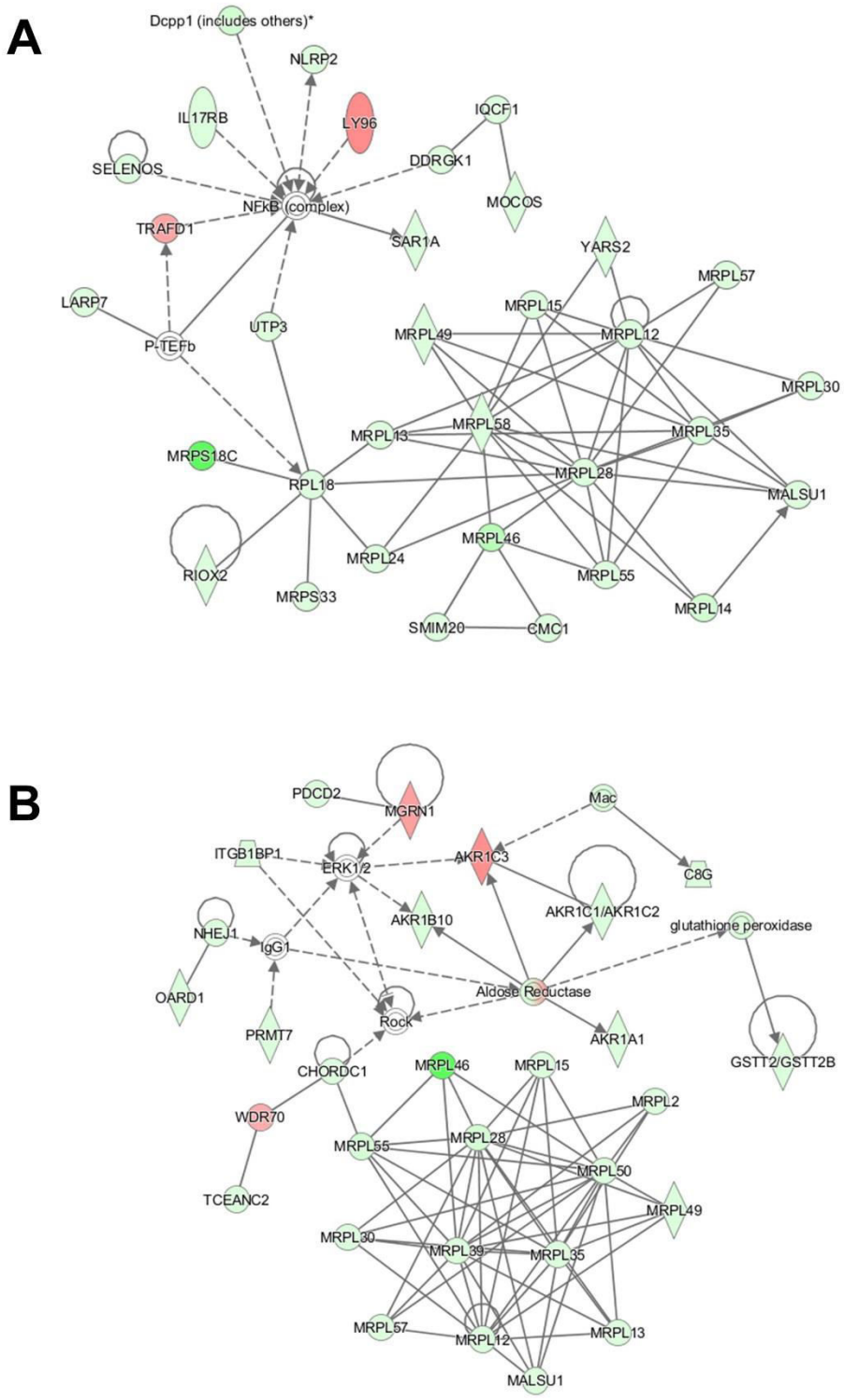


Figure S9. **Functional Network analyses by Ingenuity pathway analysis.** (A) The most significant networks identified in WT included transcripts related to gene expression, protein synthesis, cellular assembly and organisation, and maintenance. (B) The most significant networks in *Patl2*$^{-/-}$ mice assembled transcripts related to gene expression, protein synthesis and lipid metabolism. Up-regulated transcripts are shown in pink and down-regulated transcripts are indicated in green. Continuous lines between nodes indicate direct molecular interactions between connected transcripts; dotted lines indicate indirect functional interactions between transcripts.

SUPPLEMENTARY TABLES

Table S1A. Primer list for Sanger sequencing verification of PATL2 mutation

Oligo name	Sequence	Product size (bp)
PATL2H-Ex6F2	ATGTGCCATGTGGCTGACTT	286
PATL2H-Ex6R2	CAACTGGTCACAAGGGGAGA	

Table S1B. List of primers used to genotype *Patl2 KO* mice and for RT-PCR

	Oligo name	Sequence	Product size (bp)
Genotyping	Patl2-Ef	AAGCTCTGTTGGGTTTGAGGAGAAAA	WT allele: 235 KO allele: 325
	Patl2-Er	CAGCCTCTTTCCCCTGAATAATTTCA	
	Patl2-Kr	GGGCAAGAACATAAAGTGACCCTCC	
	C-LacZ-F	CCCGTCAGTATCGGCGGAAT	KO allele: 506
	L3R	TGTAATCTGGCTGCAGACAATCTAGGC	
RT-PCR	Patl2-RTseq-Ex7F	TCTGTGTTTCCAAGAGCCAGTTC	WT: 135 KO: no band
	Patl2-RTseq-Ex8R	GAGCTGGGTCAGATGACTGG	
	βact-F	ACCAGAGGCATACAGGGACA	104
	βact-R	CTAAGGCCAACCGTGAAAAG	

Table S1C. List of primers used to genotype *Patl2-HA tagged* mice and for RT-PCR

Oligo name	sequence	Product size (bp)
gRNA	5'-CAAGCAATTAGTTCAGCAGC-3'	
Patl2_HA-F primer	5'-CAGACTTTGCCTGGACATCA-3'	351 (WT) 378 (KI)
Patl2_HA-R primer	5'-GACCATGCTTGGCTCATAG-3'	

Table S1**. List of primers used for mouse genotyping**

Target /Antigen	Species	Supplier	REF	DILUTION
Tubulin	Mouse	Sigma Aldrich	T4026	1/800
MSY2	Mouse	Santa cruz Technology	SC393840	1/200 for IF
HA High Affinity	Rat	Roche/Sigma	11815016001	1/800 for IF, 1/1000 for WB
DDX6	Rabbit	Biotechne	NB200-191	1/500 for IF
CPEB1	Rabbit	Abcam	ab3465	1/200 for IF
Secondary antibody DyLight549 Goat anti mouse	Goat	Jackson Immuno Research	115-505-062	1/800
Secondary antibody AlexaFluor488 Goat anti rabbit	Goat	Jackson Immuno Research	115-545-144	1/800
Secondary antibody AlexaFluor488 Goat anti rat	Goat	Life Technologies	A-11006	1/800
Secondary antibody cy3 (?) Goat anti rabbit	Goat	Jackson Immuno Research	111-165-144	1/400
Secondary antibody HRP conjugate goat anti-rat	Goat	Millipore	AP136P	1/5000

Table S2. **List of antibodies**

5- Article Annexe 2 : Bases moléculaires et physiopathologiques des tératozoospermies monomorphes

Zine-Eddine KHERRAF, Christophe ARNOULT et Pierre F RAY.

Gynécologie Obstétrique Pratique. Soumis.

Bases moléculaires et physiopathologiques des tératozoospermies monomorphes

Zine-Eddine KHERRAF[1,2], Christophe ARNOULT[1] et Pierre F RAY[1,2*]

[1] Université Grenoble Alpes, INSERM U1209, CNRS UMR 5309, Institute for Advanced Biosciences, Team Genetics Epigenetics and Therapies of Infertility, 38000 Grenoble, France.

[2] CHU Grenoble Alpes, UM GI-DPI, Grenoble, F-38000, France

*Auteur pour correspondance : pray@chu-grenoble.fr

Résumé :

Ces dernières années ont vu l'avènement de nouvelles technologies de séquençage haut débit qui sont en train de révolutionner la génétique médicale et tout particulièrement la génétique de l'infertilité masculine. Le nombre de nouvelles causes génétiques identifiées ne cesse d'augmenter. Parmi les phénotypes étudiés les tératozoospermies monomorphes représentent un des meilleurs modèles d'étude en raison de leur forte composante génétique et d'une relative homogénéité génétique observée pour chaque phénotype. Une meilleure compréhension de la pathogénie moléculaire de l'infertilité masculine permet une meilleure orientation du couple dans le parcours de soins et d'envisager le développement de nouvelles thérapies personnalisées.

L'infertilité masculine :

L'infertilité masculine est un trouble reproductif fréquent caractérisé par une pathogénie extrêmement hétérogène mais souvent due à une altération de la production des gamètes mâles au cours de la spermatogenèse. Ce processus très complexe, évolue en trois étapes successives : la prolifération mitotique des cellules souche spermatogéniques ou spermatogonies, la division méiotique des spermatocytes et finalement la spermiogénèse, une phase de différentiation cellulaire au cours de laquelle, les spermatides rondes se différentient en spermatides allongées puis en spermatozoïdes. Cette différentiation est marquée par trois modifications morphologiques majeures qui sont : la condensation du noyau, La biogenèse de l'acrosome et la formation du flagelle. Pour être considéré comme normal, le spermatozoïde doit avoir une tête ovale avec un acrosome normal localisé sur le pole antérieur du noyau et un flagelle d'environ 50 µm de long pourvu d'une pièce intermédiaire de 4-5 µm de diamètre.

Les tératozoospermies monomorphes :

Bien que le la majorité des cas d''infertilité masculine soit liés à un trouble quantitatif de la spermatogenèse, beaucoup de cas sont liés à une altération qualitative de ce processus qui aboutit à la production de spermatozoïdes non fonctionnels. Lorsque les anomalies spermatiques constatées sont morphologiques, on parle alors de tératozoospermie. Les tératozoospermies monomorphes sont définies par la présence dans l'éjaculat de spermatozoïdes atypiques concernés par les mêmes anomalies morphologiques (Coutton et al., 2015; Ray et al., 2017). Trois formes de tératozoospermie monomorphe, représentant

environ 1 % des infertilités masculines, sont présentées ici : la macrozoospermie, la globozoospermie et le phénotype MMAF (multiple morphological anomalies of the sperm flagellum).

1- La macrozoospermie :

La macrozoospermie est un phénotype très rare d'infertilité masculine caractérisé par la présence dans l'éjaculat d'un pourcentage élevé de spermatozoïdes présentant une tête large atypique et plusieurs flagelles. Depuis sa description il y a 40 ans, plusieurs études ont rapporté un taux élevé d'aneuploïdie et de polyploïdie dans ces spermatozoïdes. Ces observations suggèrent que cette anomalie est liée à la non-disjonction des chromosomes au cours de la méiose.

En 2007, Dieterich et al., ont identifié la même délétion (c.144delC) dans le gène *AURKC (Aurora kinase C*) chez l'ensemble des patients inclus dans l'étude soit 14 patients nord-africains (Dieterich et al., 2007). Le gène *AURKC* code pour une protéine kinase sérine-thréonine-dépendante qui fait partie du CPC (chromosomal passenger complex). Ce complexe agit comme un régulateur de la méiose en contrôlant essentiellement l'alignement et la ségrégation des chromosomes. En 2012, Ben Khelifa et al., avaient génotypé 44 individus macrozoospermiques et identifié une nouvelle mutation homozygote délétère (c.744C>G, p.Y248X) sur le gène *AURKC* chez 11 individus d'origines européenne et nord-africaine. L'étude cytogénétique de ces spermatozoïdes a révélé l'existence d'une tétraploïdie homogène expliquant que les cellules germinales ne peuvent pas se diviser après la réplication de leur ADN. Malgré ce défaut méiotique, la spermiogénèse n'est pas abolie chez ces patients et aboutie à des spermatozoïdes macrocéphales multi-flagellés (figure 1) (Ben Khelifa et al., 2011).

La prise en charge de ces patients devrait commencer par la recherche de cette mutation dans le gène *AURKC*. La FIV-ICSI est alors formellement contre-indiquée chez les patients présentant des mutations homozygotes en raison de la présence constante de polyploïdie incompatible avec le développement embryonnaire. Concernant les patients qui ne présentent pas de mutations du gène *AURKC*, l'ICSI pourrait être envisagée après une analyse chromosomique par FISH (Fluorescence in situ hybridization) des spermatozoïdes afin d'évaluer le taux d'aneuploïdie. Un suivi rapproché de la grossesse devrait alors être instauré. Un diagnostic préimplantatoire pourrait également être envisagé dans certains cas présentant un taux d'aneuploïdie intermédiaire.

2- La globozoospermie :

La globozoospermie est aussi un phénotype rare observé chez moins de 0.1% des hommes infertiles. Elle est caractérisée par la présence dans l'éjaculat d'une majorité de spermatozoïdes présentant une tête globulaire dépourvue d'acrosome et une enveloppe nucléaire anormale. La présence de mariages consanguins dans les familles affectées par la globozoospermie suggère la contribution d'une composante génétique transmise sous un mode autosomique récessif. Deux gènes impliqués dans la biogenèse de l'acrosome et l'organisation de la tête du spermatozoïdes ont été identifiés dans la globozoospermie (Coutton et al., 2015).

En 2007, Dam et al., avaient identifié une mutation homozygote (c.848G>A, p.R283Q) dans le gène *SPATA16* (*spermatogenesis-associated protein 16*) chez trois frères d'une famille consanguine ashkénaze (Dam et al., 2007). La protéine produite est exprimée pendant la spermiogénèse. Elle est localisée dans l'appareil de Golgi et les vésicules proacrosomales qui fusionnent sur le pôle antérieur du noyau pour former l'acrosome. En 2011, Harbuz et al., avaient étudié une cohorte de 20 individus tunisiens globozoospermiques et identifié une délétion homozygote de 200 kb emportant le gène *DPY19L2* (*dpy-19 like 2*) chez 15 individus (Harbuz et al., 2011). D'autres études ont par la suite confirmé la prévalence élevée des altérations du gène *DPY19L2* dans la globozoospermie dans d'autres régions du monde (Coutton et al., 2012; ElInati et al., 2012; Zhu et al., 2013). La protéine DPY19L2 est localisée dans la membrane interne du noyau à partir du stade de spermatide ronde où elle est impliquée dans le positionnement de l'acroplaxome, une structure dérivée du cytosquelette et localisée au niveau du pôle antérieur du noyau spermatique qui va accueillir l'acrosome. L'absence de DPY19L2 fragilise cette structure d'attache et cause la perte de l'acrosome au cours de la spermiogénèse (figure 2).

L'absence de l'acrosome rend les spermatozoïdes globocéphales incompatibles avec la fécondation in vitro (FIV) standard. Quant à la FIV-ICSI, l'échec de fécondation observé est attribué à un déficit en facteur d'activation ovocytaire. Il s'agit d'une phospholipase spécifique des cellules germinales testiculaires, la PLCζ qui est perdu avec l'acrosome au cours de la spermiogénèse. Cependant, l'absence de la PLCζ peut être remplacée par l'injection concomitante d'un ionophore calcique avec le spermatozoïde lors de l'ICSI. Cependant les taux de grossesses sont modérés en raison de la présence de défauts de condensation de la chromatine et une fragmentation de l'ADN spermatique. Par ailleurs l'utilisation d'un ionophore calcique n'est actuellement pas permise en France.

3- Les anomalies multiples du flagelle spermatique (phénotype MMAF)

Le flagelle spermatique acquis au cours de la spermiogénèse joue un rôle essentiel dans la fécondation grâce à la mobilité qu'il procure au spermatozoïde pour parvenir à sa cible, l'ovocyte, et traverser la zone pellucide qui l'entoure. Le flagelle spermatique est composé de trois pièces morphologiquement distinctes : la pièce intermédiaire, la pièce principale et la pièce terminale. Sa structure interne est marquée par la présence d'un axonème, une structure microtubulaire de type 9+2 qu'on retrouve aussi dans les cils mobiles et des structures paraxonémales propres au flagelle spermatique comme le manchon mitochondrial, la gaine fibreuse et les fibres denses.

Le phénotype MMAF est caractérisé par une mosaïque d'anomalies morphologiques du flagelle spermatique incluant des flagelles absents, courts, enroulés ou de calibre irrégulier et une architecture centrale profondément altérée concernant à la fois l'axonème et les structure para-axonémales (figure 3). En 2014, Ben Khelifa et al. avaient montré que le gène *DNAH1* (*dynein axonemal heavy chain 1*) est muté chez approximativement un tiers des patients analysés porteurs de ce syndrome. Lorsque la protéine DNAH1 est absente, l'architecture axonémale est désorganisée avec souvent une perte de la paire centrale de microtubules (Ben Khelifa et al., 2014). Plusieurs études ont montré une élévation du taux d'aneuploïdie et une altération de la qualité nucléaire spermatique chez les patients MMAF. Cependant, chez les patients mutés pour *DNAH1*, ces anomalies sont rares et la FIV-ICSI a un bon pronostic (Wambergue et al., 2016). D'autres gènes ont été récemment identifiés dans le phénotype

MMAF comme *CFAP43* et *CFAP44* (Coutton et al., 2018). Ces résultats soulignent l'hétérogénéité génétique de ce phénotype et suggèrent l'implication d'autres causes génétiques jusqu'à présent inconnues. On peut penser alors que de nouveaux gènes associés à ce phénotype seront identifiés dans les années à venir par l'étude de plus larges cohortes de patients par séquençage haut débit.

En conclusion :

Plus d'un millier de gènes sont exprimés pendant la spermatogenèse dont environ 500 sont spécifiques de ce processus. Il est donc attendu que l'altération de l'un de ces gènes soit responsable d'anomalies spermatiques et conduise à une infertilité masculine. Le développement du séquençage de nouvelle génération et la diffusion de son utilisation en clinique va permettre l'identification de nouveaux gènes dans l'infertilité masculine et on peut s'attendre en particulier à ce que ces analyses permettent également des avancées majeures pour diagnostiquer les azoospermies (absence de spermatozoïdes dans l'éjaculat).

Bien que les tératozoospermies monomorphes soient rares, la macrozoospermie, la globozoospermie et le phénotype MMAF ouvrent une nouvelle voie pour explorer la spermatogenèse. L'identification des causes génétiques permet d'affiner le diagnostic de l'infertilité masculine et comprendre la pathogénie moléculaire sous-jacente pour, *in fine*, développer de nouvelles thérapies ciblées, éviter des explorations invasives et raccourcir le parcours de soins des couples infertiles.

Pour en savoir plus :

Ben Khelifa, M., Zouari, R., Harbuz, R., Halouani, L., Arnoult, C., Lunardi, J., and Ray, P.F. (2011). A new AURKC mutation causing macrozoospermia: implications for human spermatogenesis and clinical diagnosis. Mol. Hum. Reprod. *17*, 762–768.

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.

Coutton, C., Zouari, R., Abada, F., Khelifa, M.B., Merdassi, G., Triki, C., Escalier, D., Hesters, L., Mitchell, V., Levy, R., et al. (2012). MLPA and sequence analysis of DPY19L2 reveals point mutations causing globozoospermia. Hum. Reprod. *27*, 2549–2558.

Coutton, C., Escoffier, J., Martinez, G., Arnoult, C., and Ray, P.F. (2015). Teratozoospermia: spotlight on the main genetic actors in the human. Hum. Reprod. Update *21*, 455–485.

Coutton, C., Vargas, A.S., Amiri-Yekta, A., Kherraf, Z.-E., Mustapha, S.F., Tanno, P., Wambergue-Legrand, C., Karaouzène, T., Martinez, G., Crouzy, S., et al. (2018). Mutations in CFAP43 and CFAP44 cause male infertility and flagellum defects in Trypanosoma and human. Nat. Commun. *9*, 686.

Dam, A.H.D.M., Koscinski, I., Kremer, J.A.M., Moutou, C., Jaeger, A.-S., Oudakker, A.R., Tournaye, H., Charlet, N., Lagier-Tourenne, C., Van Bokhoven, H., et al. (2007). Homozygous mutation in SPATA16 is associated with male infertility in human globozoospermia. Am. J. Hum. Genet. *81*, 813–820.

Dieterich, K., Soto Rifo, R., Karen Faure, A., Hennebicq, S., Amar, B.B., Zahi, M., Perrin, J., Martinez, D., Sèle, B., Jouk, P.-S., et al. (2007). Homozygous mutation of AURKC yields large-headed polyploid spermatozoa and causes male infertility. Nat. Genet. *39*, 661–665.

ElInati, E., Kuentz, P., Redin, C., Jaber, S., Meerschaut, F.V., Makarian, J., Koscinski, I., Nasr-Esfahani, M.H., Demirol, A., Gurgan, T., et al. (2012). Globozoospermia is mainly due to DPY19L2 deletion via non-allelic homologous recombination involving two recombination hotspots. Hum. Mol. Genet. dds200.

Harbuz, R., Zouari, R., Pierre, V., Ben Khelifa, M., Kharouf, M., Coutton, C., Merdassi, G., Abada, F., Escoffier, J., Nikas, Y., et al. (2011). A recurrent deletion of DPY19L2 causes infertility in man by blocking sperm head elongation and acrosome formation. Am. J. Hum. Genet. *88*, 351–361.

Ray, P.F., Toure, A., Metzler-Guillemain, C., Mitchell, M.J., Arnoult, C., and Coutton, C. (2017). Genetic abnormalities leading to qualitative defects of sperm morphology or function. Clin. Genet. *91*, 217–232.

Wambergue, C., Zouari, R., Mustapha, S.F.B., Martinez, G., Devillard, F., Hennebicq, S., Satre, V., Brouillet, S., Halouani, L., Marrakchi, O., et al. (2016). Patients with multiple morphological abnormalities of the sperm flagella due to DNAH1 mutations have a good prognosis following intracytoplasmic sperm injection. Hum. Reprod. dew083.

Zhu, F., Gong, F., Lin, G., and Lu, G. (2013). DPY19L2 gene mutations are a major cause of globozoospermia: identification of three novel point mutations. Mol. Hum. Reprod. *19*, 395–404.

Légendes des figures :

Figure 1 : La macrozoospermie.

A) Au cours de la méiose, les chromosomes homologues s'apparient et s'alignent au niveau de la plaque équatoriale grâce à l'équilibre des tensions subies par chacun des kinétochores (en jaune) d'un chromosome. A la fin de chaque division méiotique (ici la méiose II n'est pas présentée), les cellules filles se séparent (cytokinèse). Le gamète mâle produit est haploïde.
B) Chez les patients porteurs de mutations homozygotes du gène *AURKC*, on observe un déséquilibre d'alignement et une non-disjonction des chromosomes au cours de la méiose I. On observe aussi un échec de la cytokinèse et les spermatozoïdes produits sont tétraploïdes et pourvus de plusieurs flagelles.

Figure 2 : La globozoospermie.

A) Représentation schématique de la morphologie d'un spermatozoïde globocéphale dépourvu d'acrosome (à droite) comparé à un spermatozoïde normal (à gauche).
B) Biogenèse normale de l'acrosome au cours de la spermiogénèse : l'appareil de Golgi produit des vésicules proacrosomales qui fusionnent sur l'acroplaxome positionné sur le pôle antérieur du noyau de la spermatide pour former l'acrosome.

C) Le déficit en SPATA16 chez les patients globozoospermiques empêche la fusion des vésicules proacrosomales et l'acrosome ne se forme pas.
D) En cas de déficit en DPY19L2, l'acroplaxome est fragilisé et l'acrosome en formation se détache du pole antérieur du noyau des spermatides. Finalement, l'acrosome est perdu avec l'excès cytoplasmique du spermatozoïde au cours de sa maturation.

Figure 3 : Le phénotype MMAF.

A) A gauche, un spermatozoïde normal et une coupe transversale au niveau de la pièce principale du flagelle. On notre du centre vers la périphérie : un axonème avec un complexe central et 9 doublets périphériques de microtubules, sept fibres denses, une gaine fibreuse et une membrane cytoplasmique.

B) A droite, un spermatozoïde typique d'un patient MMAF avec un flagelle de morphologie anormale et une structure interne profondément désorganisée.

Figure 1

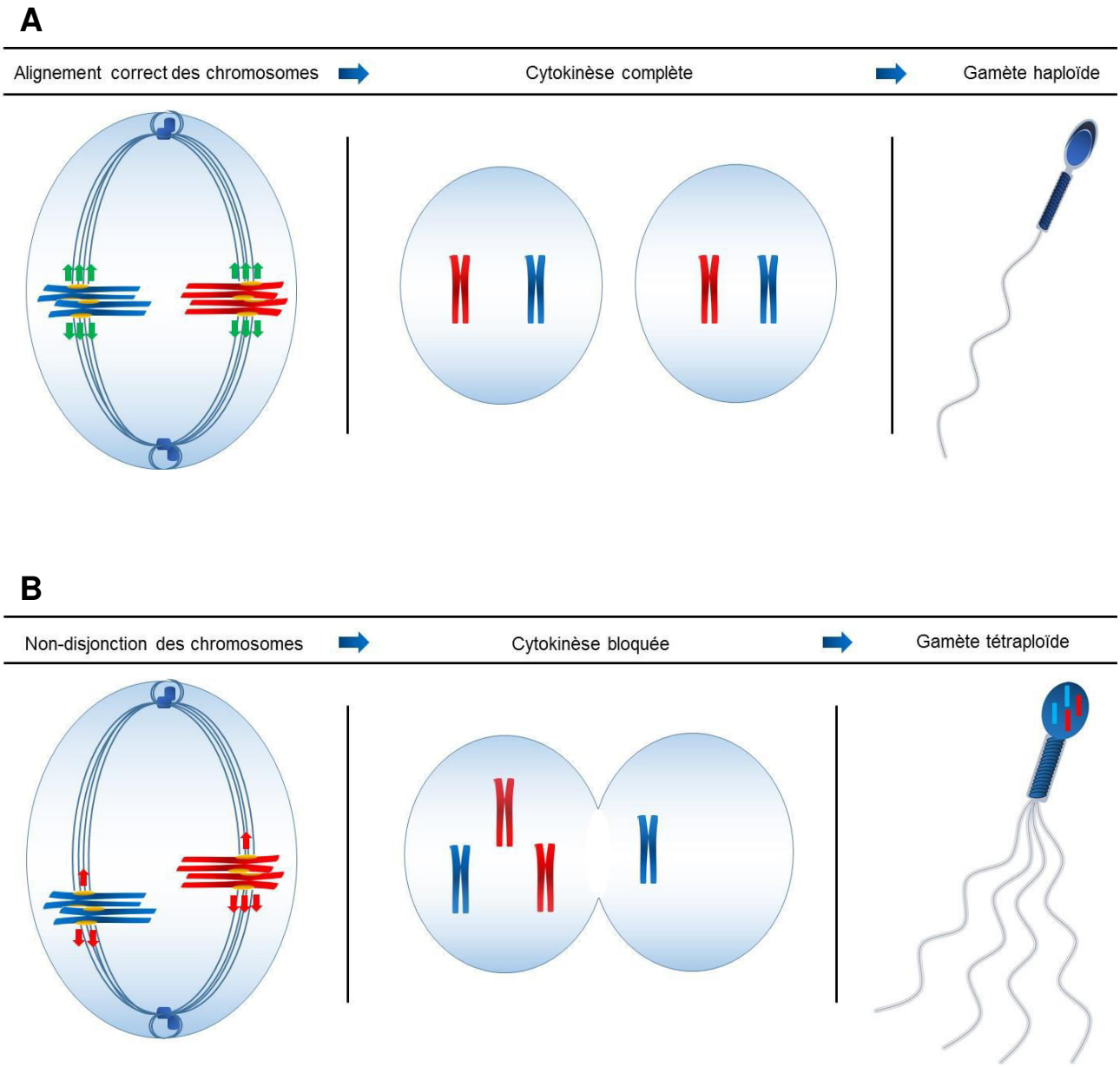

A
Alignement correct des chromosomes
Cytokinèse complète
Gamète haploïde
B
Non-disjonction des chromosomes
Cytokinèse bloquée
Gamète tétraploïde

Figure 2

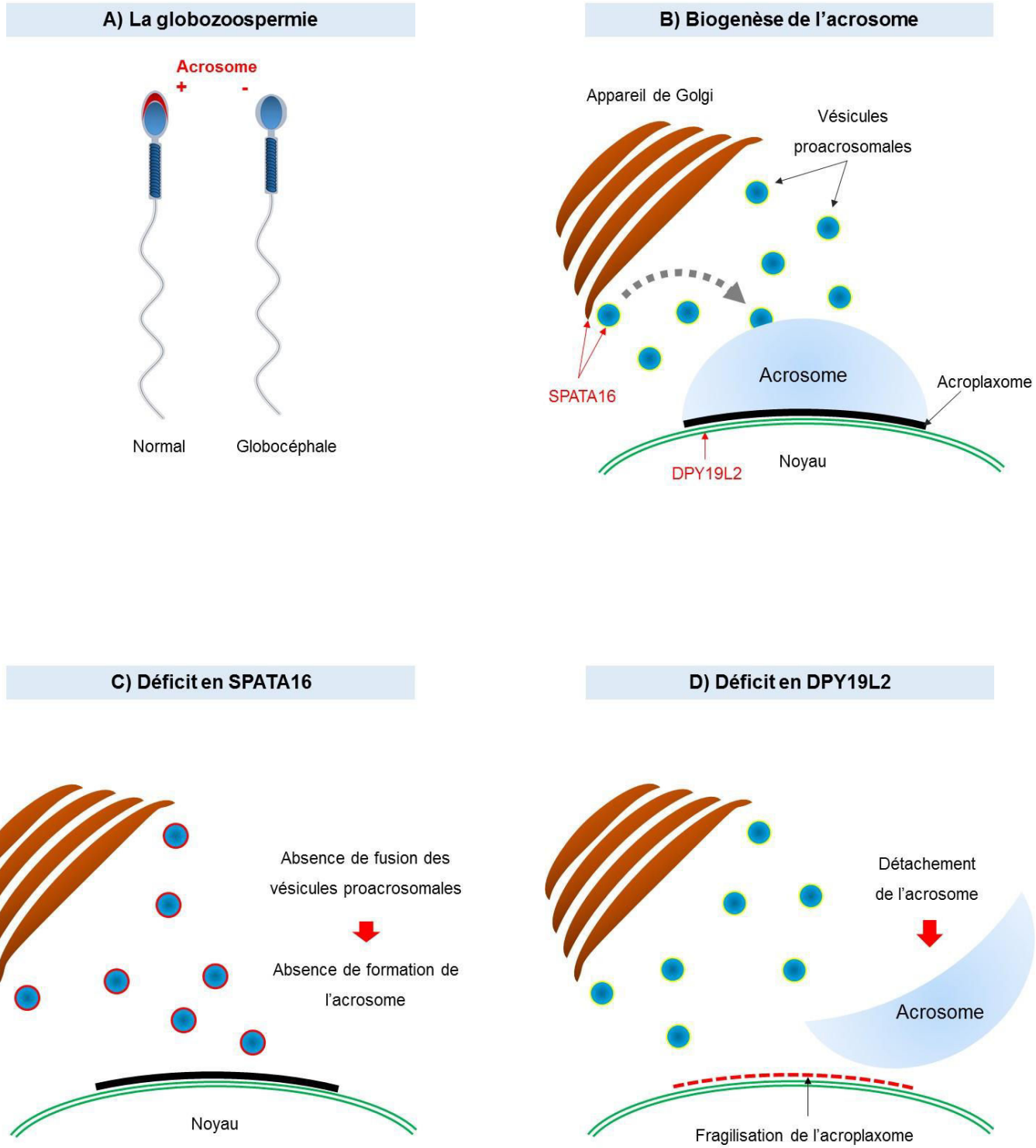
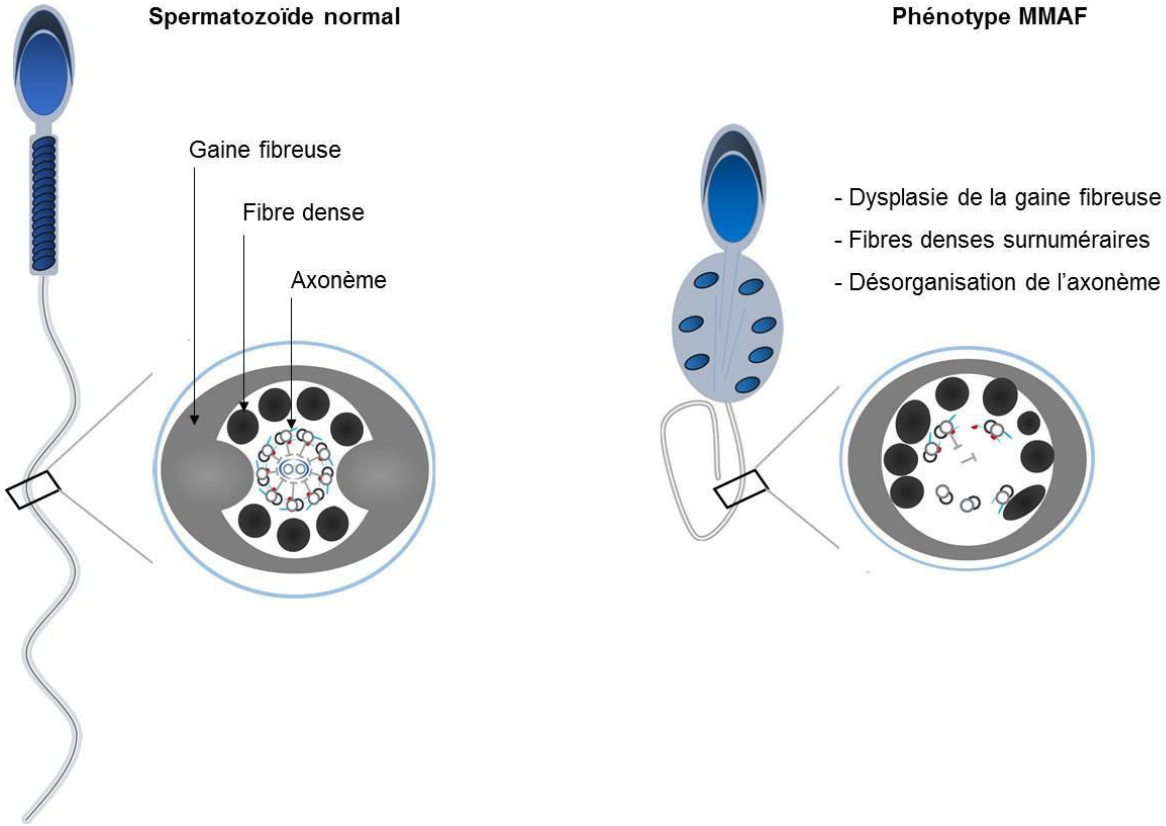

A) La globozoospermie
Acrosome
+
-
Normal
Globocéphale
B) Biogenèse de l'acrosome
Appareil de Golgi
Vésicules
proacrosomales
SPATA16
Acrosome
Acroplaxome
DPY19L2
Noyau
C) Déficit en SPATA16
Absence de fusion des
vésicules proacrosomales
Absence de formation de
l'acrosome
Noyau
D) Déficit en DPY19L2
Détachement
de l'acrosome
Acrosome
Fragilisation de l'acroplaxome

Figure 3


Spermatozoïde normal
Gaine fibreuse
Fibre dense
Axonème
Phénotype MMAF
- Dysplasie de la gaine fibreuse
- Fibres denses surnuméraires
- Désorganisation de l'axonème