Human genetics of male infertility

Elias Elinati

UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE

des Sciences de la Vie et de la Santé

IGBMC - CNRS UMR 7104 - Inserm U 964

THÈSE présentée par :

Elias ELINATI

soutenue le : **10 Septembre 2012**

pour obtenir le grade de : **Docteur de l'Université de Strasbourg**

Discipline/ Spécialité : Aspects moléculaires et cellulaires de la biologie

TITRE de la thèse

Génétique de l'infertilité masculine

THÈSE dirigée par :

M. VIVILLE Stéphane — Professeur, Université de Strasbourg

RAPPORTEURS :

Mme CHABOISSIER Marie-Christine — Docteur, Institut de Biologoie Valrose

M. TURNER James — Docteur, National institute of medical research

---

AUTRES MEMBRES DU JURY :

M KOENIG Michel — Professeur, Université de Strasbourg

The learning and knowledge that we have, is, at the most, but little compared with that of which we are ignorant.

Plato

Acknowledgments

Finally, it’s time for me to graduate … Yeah, I did it!

While I am enthousiastic to finally bring a chapter of my life to a close, I have to say it has changed me on all levels and allowed me to meet many outstanding persons.
The three years I spent in Strasbourg for my thesis have been a challenging trip, with both ups and downs. Fortunately, I was not alone on this road, but accompanied by an extended team of experts, and by thoughtful and generous friends always willing to coach, help, and motivate me. For this, I would like to kindly thank all of them.

First and foremost, I wish to thank my supervisor Professor Stéphane Viville. You are full of knowledge and ideas; you are eager to share them and were always ready to find time for me disregarding your busy schedule. Thank you so much for always being there for me, in times of when the research was going on, but also in stressful periods and for your understanding during my PhD. I will never forget your support especially at the beginning of my journey when I was homesick and I felt like a fish out of water. It was not obvious to a person who never been abroad to adapt quickly and integrate in a society completely strange. The discussion we had that day in your office and all the advices you gave to me motivated and reassured me.

Furthermore, I would like to thank the jury members: Mr. Michel Koenig, Ms. Marie-Christine Chaboissier, Mr. James Turner, as well as Mr. Jean Muller for accepting to evaluate my thesis.

Special thanks go to all people who have worked with me over the years, in particular our collaborators who provided us with DNA samples. Without your efforts, this work couldn’t be done. I would like to thank my colleagues at the department for many useful discussions, comments and suggestions as well as the IGBMC common facilities for the great job they did.

I will not forget the priceless collaboration with Jean Muller and Claire Redin that lead to two peer-reviewed articles. Jean, your remarks, rewarding discussions and your tremendous efforts are so valuable. It was a pleasure to work with someone as organized as you and who showed a full motivation for the project. Claire, thank you for all the bioinformatics studies you did. I appreciated working with you. You are such a good friend too; I enjoyed the parties you organized even if I missed some.

For Julie Thompson, an enormous thanks. It is so kind that you accept to read and correct my English.

I would like to express my sincere appreciations to all the members of Stéphane Viville's group. This thesis would not have been possible without your support.

Yara Tarabay: I couldn't imagine the lab without you. Thank you for being here each time I needed someone to share my joys and my worries with.

Laura Jung: The "Junge Sterne". Laura you are a friend that I highly appreciate. You taught me to how to be perseverant and patient. I will miss our late night discussions...

Cecile Andre: I can say you are the one who introduced me to the French culture. From all the songs you used to sing my preferred one will stay "Avec Cécile je ne sais pas…"

For the three of you, I will never forget these precious moments we spent together. Without you, I am sure that I couldn't make it.

Marius Teletin: Thank you for your good mood and for the funny stories you used to tell us. This made the days more pleasant.

Philippe Tropel: Your knowledge was an asset and an idol for me.

Valerie Skory: The women who has whatever you need. Thanks for providing me with the protocols and the solutions when needed.

Mayada Achour: Your presence in the lab was encouraging. Thanks for all your advices and especially for taking over my duty and training Ozum while I was writing this thesis.

I would like to extend my appreciation to all those individuals who were involved directly in this work:

Camille Fossard: I will never forget the enjoyable time we had especially at 4:15 p.m.

Paul Kuentz: We worked together in a total harmony. Many thanks for your contribution to this work.

Sara Jaber: The three months you spent in our team were so agreeable and have contributed to the rapid and effective progress of this project. I liked what we called "the genetics team" and all the discussions we had.

Joelle Makarian: Thank you jojo for all the time we spent in/outside the lab. Good luck for your future.

Ozum Gunel: I will never forget your joyful presence in the lab and I appreciated the way you managed the experiments. I apologize that I wasn't available all the time for the discussions.

I am also grateful for the following members of our team that I met during my three years in the lab: Isabelle Koscinski, Emmanuelle Kieffer, Fiona Bello, Baptiste Maupoil, Rosy El Ramy, Adeline Tosh, Catherine Celebi, Laetitia Furhrman, Emese Gazdag and Aafke von Montfoort.

Thanks for the neighbors (Gerard Gradwohl's team members) mainly:

Aline Meunier: for your distinct mood and for the tasty cakes you used to prepare.

Perine Strasser: for your direct involvement in the cultural dinner we prepared in 2011. Without your motivation this dinner wouldn't be done.

Julie Piccand: thanks for sharing the latest and the most stressful moments of the Ph.D.

My stay in Strasbourg wouldn't be more pleasant without you my dear fellows:

Mirella Gemayel: your distinguished support and care were remarkable. Without your advices, your technical knowledge and criticism this thesis wouldn't be done easily.

Miled Chamoun: thank you bro for your support. You are a true friend.

Mirna Assoum: I will forever be grateful for your hospitality and your scientific advices.

I am indebted to Anastasia Salloum, Andre Eid, Maher El Mashriki, Ranine Tarabay, Mona Karout, Firas Fadel, Hassan Hammoud, Andrea Costan, Elie Feghali, Hussein Hamade, Christine Ayoub and to all my friends in the Greek parish, for the unforgettable moments we spent together.

Even if I left Lebanon, my home country, I will always be grateful for all my professors who strengthened my passion for biology.
For my friends there, an exceptional thanks. You have been great amigos.

And beyond friends, there is family. My parents, Abdallah and Aida, and my sisters, Pamela and Patricia, have been there for me all along. I can't find the words to thank you. I couldn't make it if you didn't believe in me and without the sacrifices you made so far. Thanks for supporting me. I love you so much.

Preface

This thesis represents a culmination of work and learning that has taken place over a period of almost three years (2009 - 2012). It has been written on the basis of experiments conducted in the scientific group of Professor Stéphane Viville at the IGBMC in Strasbourg in France. The overall objective is to identify autosomal genes implicated in idiopathic human male infertility which will contribute to a better understanding of human spermatogenesis. The project has three main goals: 1) to strengthen the patient recruitment network, 2) to analyse the patients by whole genome scanning techniques in order to identify genes responsible for infertility, 3) to perform a functional analysis in order to confirm the pathogenicity of the identified mutations and to investigate the role of these genes in the physiopathology of gametogenesis. Two phenotypes are studied: globozoospermia and non-obstructive azoospermia with maturation arrest.

The first five chapters describe the background, methods and literature, with the remaining four chapters each presenting an analysis of data. The first chapter gives a broad overview of testicular histology and organization, with the second chapter providing descriptions of spermatogenesis. In chapter three, some mechanisms of DNA double strand break (DSB) repair are discussed. Chapter four presents a description of genomic rearrangements that can occur during DNA DSB repair and which are responsible for some genomic disorders. In chapter five, I introduce the genetics of human male infertility, describing the established causes. I focus on the autosomal mutations associated with spermatogenic failure and on the techniques used to find them. The following four chapters concern the results of my thesis, which are divided into two sections evoking the two studied aspects of infertility. In total, two consanguineous families, as well as 92 isolated cases, have been analyzed. Chapters six and seven concern the identification of *DPY19L2*: the major gene implicated in human globozoospermia. This study started by performing a genome-wide scan analysis of a Jordanian family, using 10K SNP arrays (Affymetrix Genechip). In chapter eight, pathogenicity of *SPATA16* mutations are confirmed by identifying a second mutation: a deletion of exon2 in an additional patient. Localization of the protein at the acrosome is confirmed by immunofluorescence using a generated polyclonal antibody. Furthermore, interactors of Spata16 were determined by GST pulldown allowing the identification of

putative candidate genes implicated in globozoospermia. In chapter nine, I discuss the preliminary results obtained concerning azoospermia. Since a linkage analysis strategy has been shown to be a powerful technique, a 10K SNP array was performed on a Turkish consanguineous family including two azoospermic brothers, a sister and an aunt who both underwent repetitive hidatidiform moles and two unaffected sisters. We hypothesised that a common mutation perturbs meiosis, but with a different impact on female and male gametogenesis. A unique homozygous region located on chromosome 11 was shared only with the affected siblings. This region of 27 Mb contains almost 477 genes. Thus whole exome sequencing was performed on the genome of the two azoospermic brothers in order to determine the causative gene. Finally, I present conclusions and some perspectives.

Table of contents

Table of figures

Abreviations

ABP	Androgen binding protein
Afaf	Acrosome formation associated factor
AHR	Allelic homologous recombination
AMPA	α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor
ARHGEF17	Rho guanine nucleotide exchange factor (GEF) 17
ASH	Allele specific hybridization
AURKA/B/C	Aurora kinase A/B/C
AZF	Azoospermia factor region
bHLH	Basic helix-loop-helix
BIR	Break-induced replication
B-NHEJ	Back-up NHEJ
CATSPER	Cation channel, sperm associated 1
CBAVD	Congenital absence of the vas deferens
CCDC89	Coiled-coil domain containing 89
CFTR	Cystic fibrosis transmembrane conductance regulator
CNV	Copy number variation
CO	Crosing over
CSNK2A2	Casein kinase 2, alpha prime polypeptide
DAZ	Deleted in azoospermia 1
dbSNP	SNP database
DDE	Dichlorodiphenyldichloroethylene
DDT	Dichlorodiphényltrichloroethane
dHJ	Double Holliday junction
D-loop	Displacement loop
DNA	Deoxyribonucleic acid
DNAJA4	DnaJ (Hsp40) homolog, subfamily A, member 4
Dnmt3l	DNA (cytosine-5-)-methyltransferase 3-like
DPY19L2	dpy-19-like 2
DSB	Double strand break
DSBR	Double strand break repair pathway
FGF4	Fibroblast growth factor 4
FoSTeS	Fork stalling and template switching
FSH	Follicule stimulating hormone
G phase	Gap phase
GATK	Genome analysis toolkit
GEFs	Guanine nucleotide exchange factor

GnRH	Gonadotropin releasing hormone
GOPC	Golgi associated PDZ and coiled-coil motif containing
GWAS	Genome wide association study
H3K4me3	Tri-methylation of histone 3 on lysine 4
HapMap	Haplotype Map
HJ	Holliday junction
HPG	Hypothalamic-Pituitary-Gonadal
HR	Homologous recombination
HRB	HIV-1 Rev binding protein
IAM	Inner acrosomal membrane
ICSI	Intracytoplasmic sperm injection
INCENP	Inner centromere protein antigens
IRGC	Immunity-related GTPase family, cinema
KAL	Kallmann syndrome 1
Kb	Kilobase
LC	Leydig cells
LCR	Low copy repeats
LEP	Leptin
LH	Luteinizing hormone
LINE	Long-interspersed nuclear elements
LOH	Loss of heterozygosity
LTR	Long terminal repeats
Mb	Megabase
MEPS	Minimal efficient processing segments
MMA/B	MismatchA/B
MMBIR	Microhomology-mediated break induced replication
MMEJ	Microhomology-mediated end joining
MMR	Mismatch repair
MSY	Male specific Y
MTL5	Metallothionein-like 5, testis-specific
NAHR	Non allelic homologous recombination
ng	nanogram
NGS	Next generation sequencing
NHEJ	Non-homologous end joining
NOA	Non-obstructive azoospermia
NR5A1	Nuclear receptor subfamily 5, group A, member 1
NROB1	Nuclear receptor subfamily 0, group B, member 1
NSF	N-ethylmaleimide sensitive factor
OA	Obstructive azoospermia
OAM	Outer acrosomal membrane
PAR	Pseudoautosomal region
PCR	Polymerase chain reaction
PCSK1	Proprotein convertase subtilisin/kexin type 1

PEX10	Peroxisomal biogenesis factor 10
pH	power of hydrogen
PICK1	protein interacting with C kinase 1
PIWIL4	Piwi like 4
PLCζ	Phospholipase C zeta
PMA/B	Perfect matchA/B
POLA2	Polymerase (DNA directed), alpha 2, accessory subunit
PolyPhen	Ploymorphism phenotyping
PRDM9	PR domain containing 9
PRM	Protamines
PRMT6	Protein arginine methyltransferase 6
PTM	Peritubular myoid cell
RBMY	RNA binding motif protein, Y-linked, family 1, member A1
RNA	Ribonucleic acid
S phase	Synthesis phase
S. cerevisiaie	*Saccharomyces cerevisiae*
SCO	Sertoli cell only syndrome
SD	Segmental duplications
SDSA	Synthesis-dependant strand annealing
SET	Su(var)3-9, E(z)h, Tritorax
SF1	Splicing factor 1
SINE	Short-interspersed nuclear elements
SMS	Smith-Magenis syndrome
SNAPs	NSF attachment proteins
SNAREs	SNAP receptors
SNP	Single nucleotide polymorphism
SOHLH1	Spermatogenesis and oogenesis specific basic helix-loop-helix
SOX9	Sex determining region Y-box 9
SPATA16	Spermatogenesis associated 16
SRY	Sex determing region of Y chromosome
SSA	Single strand annealing
SSC	Spermatogonial stem cells
ssDNA	Single strand DNA
TCP1	t-complex protein 1
TP	Transition proteins
TPR	Tetratricopeptide
TSGA10IP	Testis specific 10 interacting protein
VCF	Variant cell format

Introduction

Chapter I

Testicular histology

1- An overview

For all sexual organisms, spermatogenesis is one of the essential biological events without which, life as we know it would perish. It is defined as a developmental process by which spermatozoa are generated from male germ cells within the seminiferous tubule of the testes. In mammals, testes are paired organs that perform two functions: production of spermatozoa and synthesis of steroids, especially the androgen testosterone. Each testis is encased by a connective tissue, the tunica albuginea, from which the mediastinum testis extends (Clermont 1972). A radial fibrous, called the septa testis, diffuses from the mediastinum testis. The radial fibrous radiates towards the tunica albuginea and divides the parenchyma of the human testis into about 300 lobuli which communicate peripherally. Each lobule contains 2 to 3 convoluted seminiferous tubules i.e. coiled structures forming hairpin loops that fuse with straight tubules, which continue into the rete testis, a labyrinthine system of cavities (Huckins and Clermont 1968). The head of each testis is covered by an epididymis, a tightly-coiled tube connecting the efferent ducts to the vas deferens which in turn connects the epididymis to the ejaculatory ducts **(Figure 1)**. The human epididymis can be divided into three main regions (i) the head (Caput), (ii) the body (Corpus) and (iii) the tail (Cauda). The head receives spermatozoa via the efferent ducts of the mediastinum testis, while the tail absorbs fluid to make the sperm more concentrated (Setchell, Sanchez-Partida et al. 1993). In the following sections, gonad organization and meiosis concerns the human, in other cases, I will define what species is concerned.

2- Cellular components of the seminiferous tubule

The seminiferous tubules, composed of a peritibular tissue and a stratified epithelium, are enclosed by the tunica propria. The peritibular tissue is made of myoid cells interposed between several layers of collagens and elastic fibers and separated from the seminiferous epithelium by the basement membrane **(Figure 1)**. This membrane is one of the main components of the tubular tissue in addition to the tunica propria. Peristaltic contractions of the seminiferous tubule are caused by the myofibroblasts triggering the transport of the immotile spermatozoa to the rete testis (Dym and Fawcett 1970; Hermo and Clermont 1976).

The epithelium consists of Sertoli and germ cells that are intimately associated and form an epithelium bordering a lumen. Sertoli cells separate germ cells from the basement membrane and are arranged in a specific way to form concentric layers with immature germ cells (Fritz 1994).

The germinal epithelium consists of cells that comprise germ cells at different developmental stages, namely spermatogonia, primary and secondary spermatocytes and spermatids. These are located within invaginations of Sertoli cells (mitotically inactive in adults) which are known to provide nutritional and structural support to developing germ cells **(Figure 1)** (Fritz 1994). At the terminal part of seminiferous tubules, where they connect into the rete testis, modified Sertoli cells predominate with the presence of occasional germ cells. The tall elongated modified Sertoli cells are self-oriented in a downstream direction and their apices converge together distally in the direction of the rete testis (Hermo and Dworkin 1988). Here, they form a plug-like structure with a narrow lumen acting as a valve, preventing the reflux of substances from the rete testis back into the tubular lumen (Hermo and Dworkin 1988). The interstitial tissue between the convoluted seminiferous tubules contains Leydig cells, macrophages, blood and lymphatic vessels as well as nerves and constitutes the endocrine part of the testis.

Germ cells develop in cellular associations known as stages of the seminiferous epithelium cycle which are indicated by Roman numerals by Lebond and Clermont (Leblond and Clermont 1952). Six possible stages were described in human (Heller and Clermont 1963), but twelve in mice and fourteen in rat (Oakberg 1956; Oakberg 1957). In a round cross section of human seminiferous tubule, two to four cellular associations are observed (Roosen-Runge 1952). Each association contains many generations of germ cells **(Figure 2)**.

A generation is defined as numerous germ cells at the same developmental step. The

annotation of the stage depends on the content of different generations of germ cells. In human, cellular associations I and II contain 2 generations of spermatids whereas cellular associations III–V contain 2 generations of primary spermatocytes (Amann 2008). Cellular association VI is defined by the presence of metaphase plates of the first or the second meiotic divisions, or both (Amann 2008). The precise sequential occurrence of the six cellular associations that happens over time in a given area of the tubule is called as a cycle of the seminiferous epithelium (Leblond and Clermont 1952). The duration of a cycle is constant and is about 16 days in humans, 8.65 days in mice and 13 days in rats (Leblond and Clermont 1952; Oakberg 1956; Oakberg 1957; Amann 2008). Spermatogenesis in human requires 4.6 cycles over 74 days.

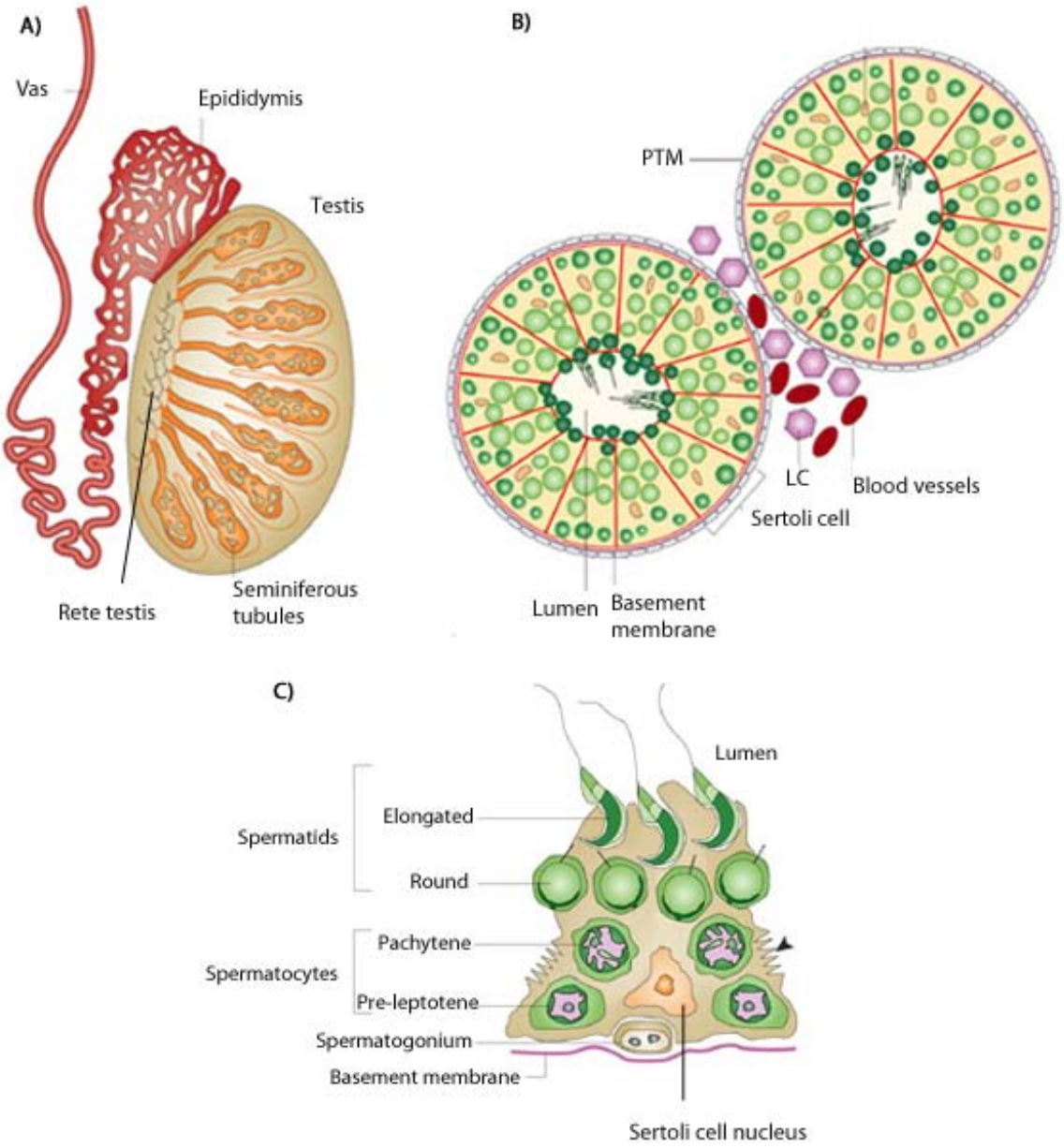

Figure 1: Organization of the testis. A) A cross-section through a testis, showing the location of the seminiferous tubules, the vas deferens and the epididymis. B) A diagrammatic cross-section through a testicular tubule, showing the germ cells (green) at different stages of maturation developing embedded in somatic Sertoli cells (each Sertoli cell is outlined in red). Leydig cells (LC) — where testosterone is synthesized — are present in the interstitium. Maturing sperm are shown in the lumen of the tubules. PTM, peritubular myoid cell. C) A single Sertoli cell with its associated germ cells. Note that tight junctions between Sertoli cells (arrowhead) define two compartments: the stem cells and the pre-meiotic cells (spermatogonia) are found on one side of the junction, whereas the meiotic (spermatocytes) and the post-meiotic (round and elongating spermatids) cells are found organized in strict order of maturation towards the lumen (cytoplasm is shown in dark green, DNA is shown in pink, Sertoli cell nucleus is shown in orange).

(According to Howard J. Cooke *et al.* ; 2002)

Chapter II

Spermatogenesis

Spermatogenesis is the development of the male gametes that occurs in the germinal epithelium of the testes. It begins at puberty and is divided into three major stages (i) proliferation, (ii) maturation of spermatocytes and (iii) spermiogenesis.

1- Proliferation phase

The male gamete or spermatozoa results from the differentiation of spermatogonial stem cells (SSC) characterized by two properties: their self-renewal capability and their ability to divide into differentiating daughter cells (Robey 2000). The latter do not derive directly from the adult stem cell but through progenitor cells, intermediate populations inserted between stem and differentiated cells that cannot generate new stem cells but can self-renew for a limited number of cycles.

In men, several types of spermatogonia have been identified **A-dark**, **A-pale** and **B-spermatogonia** (Clermont 1966; Aponte, van Bragt et al. 2005). They can be distinguished by their morphologies. For instance, **A-dark** cells are recognized by their dark staining in tissue sections, **A-pale** by their less dense staining and **B-spermatogonia** by their large spherical shape with less dense cytoplasm and a round nucleus containing one or two irregular nucleoli (Ehmcke and Schlatt 2006).

According to the Clermont model, A-spermatogonia are usually found in pairs. One member of an **A-dark** pair splits to give two **A-dark** cells, while the other member gives two **A-pale** cells (Clermont 1966; Clermont 1966). Each **A-pale** undergoes just one division to produce two **B-spermatogonia**. In this model, **A-pale** spermatogonia do not self-renew; their pool is replenished by proliferating **A-dark** spermatogonia which according to Clermont must

undergo regular mitotic divisions.

Later stage, Ehmcke and Schatt proposed a new scheme of human spermatogonial proliferation taking into account their findings in the monkey **(Figure 2)**. In their model, spermatogenesis starts from a pair or quadruplet of **A-pale** spermatogonia that enter a first division (Ehmcke, Luetjens et al. 2005; Ehmcke, Simorangkir et al. 2005). These cell clones split and enter a second division which gives rise not only to pairs, quadruplets or eight cells of **B-spermatogonia**, but also a pair or quadruplet of **A-pale** spermatogonia. Hence, **A-pale** show cyclic proliferation leading to a high number of differentiating daughters, while renewing themselves to maintain a constant supply, whereas **A-dark** show low mitotic activity throughout their lifetime to maintain genome integrity (Ehmcke, Wistuba et al. 2006). In contrast, **A-dark** show high proliferative activity during pre-pubertal testicular development, when the pool of both A-types is expanding or when most **A-pale** and **B-spermatogonia** have been abolished and need to be restored (Ehmcke and Schlatt 2006).

To summarize, **A-dark**, known as testicular stem cells, are considered as a reserve that maintains spermatogonia population. They do not directly participate in producing sperm and simply ensure a supply of stem cells. However, **A-pale** show typical characteristics of progenitors. Thus, spermatogonial stem cells maintain a pool of diploid germ cells and guarantee a constant high sperm output. This combination is the ideal system in male germ line in order to preserve genome integrity and to produce millions of gametes in adult primate males. In fact, the presence of low mitotic cells (**A-dark)** with the progenitors (**A-pale**) decrease the risk for germline mutations and vulnerability to cytotoxic events and correlate with the lifespan of the species (Ehmcke, Wistuba et al. 2006). Indeed, primates, contrary to rodents, have a long lifespan with a low number of offspring per individual leading to intense environmental exposure and thus need to protect and preserve their reproductive capability. Therefore, the presence of progenitors is primordial in primates, but not in rodents.

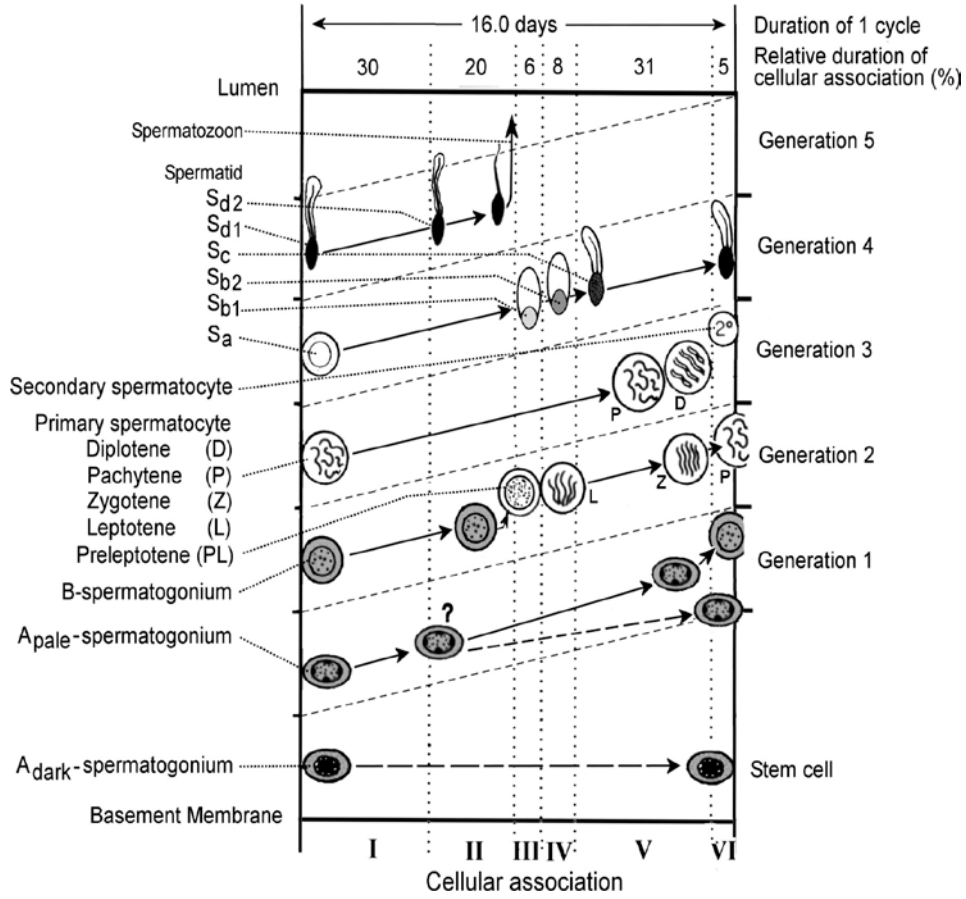

Figure 2 : The cycle of Human seminiferous epithelium. Examination of cross sections of human testes reveals 6 cellular associations (Clermont, 1963; designated I–VI on the x axis). The types of germ cells that should be present in each cellular association are illustrated in the 6 columns (upward from basement membrane to tubule lumen) as generations of progressively more differentiated cells. Reinterpretations (Ehmcke and Schlatt, 2006; Ehmcke et al, 2006) of Clermont's (1966a,b) data led to conclusions that Adark-spermatogonia indeed are stem cells, to allow repopulation of a testis, and that Apale-spermatogonia might undergo an unreported division (designated by ?) near the start of cell association II. If it is established that Apale-spermatogonia produced in cell association II (designated ?) are the first cells committed to differentiate, as suggested by Ehmcke and Schlatt (2006) then the duration of spermatogenesis would be close to 4.2 cycles or ,68 days.

(According to RUPERT P. AMANN, 2008)

The difference between the Clermont and Ehmcke models is due to the fact that Clermont missed a second division of **A-pale** spermatogonia that occurs in stage I-II of the spermatogenic cycle. His model is based on the numeric ratio between cell types present at different spermatogenic stages (Ehmcke and Schlatt 2006). He combined stages I-II to calculate the ratio (1:2) of **A-** and **B-spermatogonia**. This combination accounts for 50% of the spermatogenic cycle which lasts 16 days in human. This long period of time, corresponding to 8 days (16/2), allows more than one cell division to occur which is missed in this approach. Actually, Clermont's model is based on cell counts, which will always be

proportional values in any model. If the two divisions of **A-pale** which give rise to **B-spermatogonia** and **A-pale** (according to Ehmcke) occur with the same kinetics, a temporary increase of A-spermatogonia in stages I-II will be provoked. Consequently, the Clermont counts resulted in a 1:2 ratio of **A-pale** with respect to B-spermatogonia, which in fact might be a 1:3 ratio (Ehmcke and Schlatt 2006).

The types and numbers of spermatogonial stem cells are not the same in all mammals **(Figure 3)**. They vary widely in different species and in some, the self-renewing progenitors are not detected. In mice and rats, seven types of A-spermatogonia are described (A-single, A-pair, A-aligned, A1, A2, A3 and A4) with Intermediate and B-spermatogonia with no progenitors (Dettin, Ravindranath et al. 2003). In non-human primates, the situation is different. Two types of A spermatogonia are present (A-dark and A-pale) associated with four types of differentiated B-spermatogonia (B1, B2, B3, and B4) (Simorangkir, Marshall et al. 2005).

In fact, rodents, in contrast to primates, have no need for progenitors since A-single acts as a regenerative and functional reserve. A-single self-renews (low mitotic division) and gives rise to A-pair via clonal expansion, meaning that all germ cells derive from an initial stem cell division (Ehmcke, Wistuba et al. 2006).

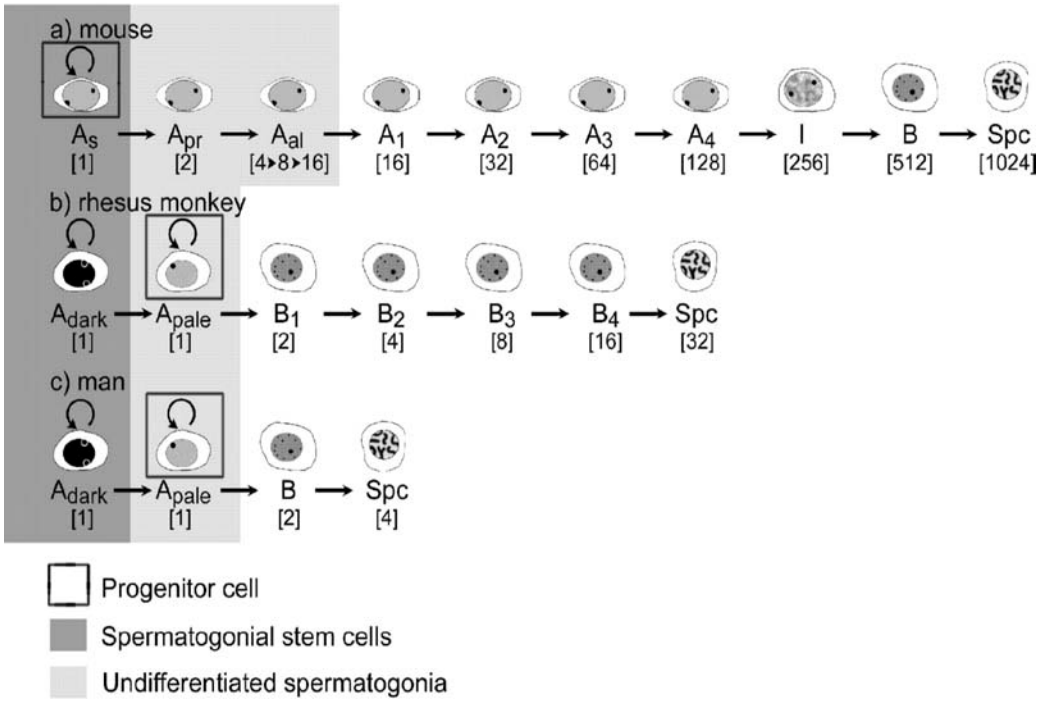

Figure 3: Schematic overview of the premeiotic steps of spermatogenesis in different species of mammals. The number given in brackets underneath the cells indicates the total number of daughter cells derived from any one progenitor cell that enters differentiation.

(According to Jens Ehmcke *et al.* ; 2006)

2-Spermatocyte's maturation

Proliferation of differentiated B-spermatogonia gives rise to primary spermatocytes that enter the meiotic division. Meiosis consists of two successive divisions preceded by a single phase of DNA replication. The first is a reductional division, involving the separation of the duplicated homologs from each other and the second is an equational division that concerns the segregation of sister chromatids. As a result, haploid gametes are produced from a diploid cell. The first meiotic cell division involves tremendous changes and comprises prophase I, metaphase I, anaphase I and telophase I.

90% of the time for meiosis is spent in prophase I, traditionally divided into five stages - leptotene, zygotene, pachytene, diplotene and diakinesis- that combines complex events classified in three processes: chromosome pairing, synapsis and meiotic recombination **(Figure 6)** (Cobb and Handel 1998; Pawlowski and Cande 2005).

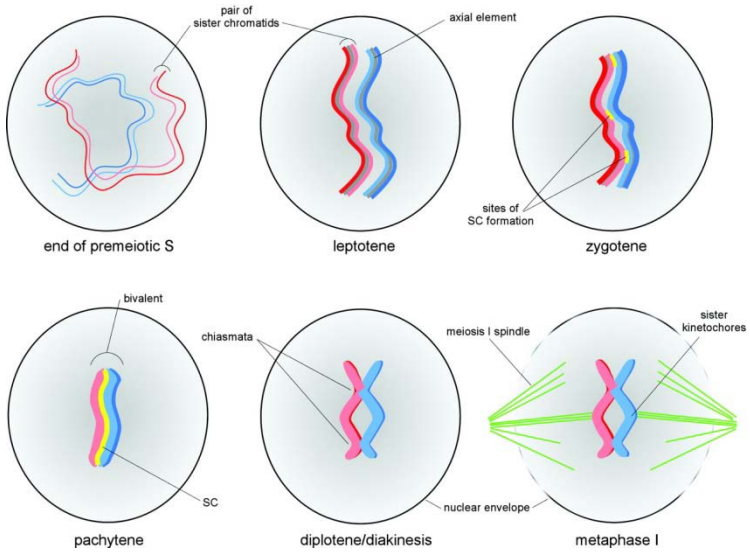

Figure 4 : Prophase I stages. During prophase of meiosis I, chromosomes accomplish the three basic steps of pairing, synapsis, and recombination. Interactions between one pair of homologous chromosomes (red and blue) within a nucleus during prophase of meiosis are schematically represented. Sister chromatids produced during premeiotic S phase are shown as different shades of red or blue. Meiotic prophase is classically subdivided into five stages: leptotene, zygotene, pachytene, diplotene, and diakinesis. Chromosomes begin to condense, homologs become aligned along their lengths, and axial elements form between sister chromatids during leptotene. Zygotene is marked by the initiation of synapsis and building the Synaptonemal Complex (yellow) between the paired homologs. By pachytene, synapsis is completed to produce a mature bivalent. Chiasmata resulting from interhomolog recombination that occurs during the zygotene and pachytene stages are evident during the diplotene and diakinesis stages and serve to connect the homologs. Breakdown of the nuclear envelope signals the end of prophase and is followed by formation of the meiosis I spindle (green) at metaphase I.

(According to Scott L. Page *et al.*; 2003)

A/ Chromosome pairing

The diploid nucleus of primary spermatocytes contains two versions of each autosomal chromosome called homologs and a pair of sex chromosomes. Before entering meiosis, DNA replication occurs. Thus, each chromosome is duplicated and is formed by two sister chromatids linked by a centromere. At the leptotene stage, sister chromatids attach to the nuclear envelope, begin to condense their chromatin and simultaneously start to recognize their homolog partners **(Figure 4)** (Bhalla and Dernburg 2008). Chromosomes are anchored to the nuclear envelope by the telomeres and are polarized within the nucleus as parallel arms leading to the description of a meiotic bouquet (Zickler and Kleckner 1998). This process known as alignment is followed by a pairing mechanism. It involves the stabilization of homolog interactions that results in the juxtaposition of the homolog chromosomes physically connected side by side along their entire length (Peoples-Holst and Burgess 2005). They form a structure called the bivalent, composed of two chromosomes **(Figure 4)**. Concerning sex chromosomes, pairing is also primordial. In female mammals, pairing can occur as for other homologs since they have two X chromosomes. Nevertheless, in males that have one X and one Y chromosome which are not homologous, pairing occurs because of the small region of homology between the X and the Y at the two ends of these chromosomes.

B/ Synapsis

At the zygotene stage, the synaptonemal complex begins to develop between the two sets of sister chromatids in each bivalent (Cobb and Handel 1998). The synaptonemal complex consists of two axial elements, a central element and transverse filaments (Zickler and Kleckner 1999). The sister chromatids of each homolog are close together by the axial element. Once the pair chromatids of each bivalent are well assembled, a central element begins to form between the two homologs and connects to the two axial elements, now referred to as lateral elements by transverse filaments **(Figure 5)** (Cohen, Pollack et al. 2006). The assembly of the synaptonemal complex is called synapsis (Bhalla and Dernburg 2008). Pachytene, the long stage of prophase I, begins when synapsis is complete. It involves thickening and shortening of the chromosomes and results in a zipper-like structure.

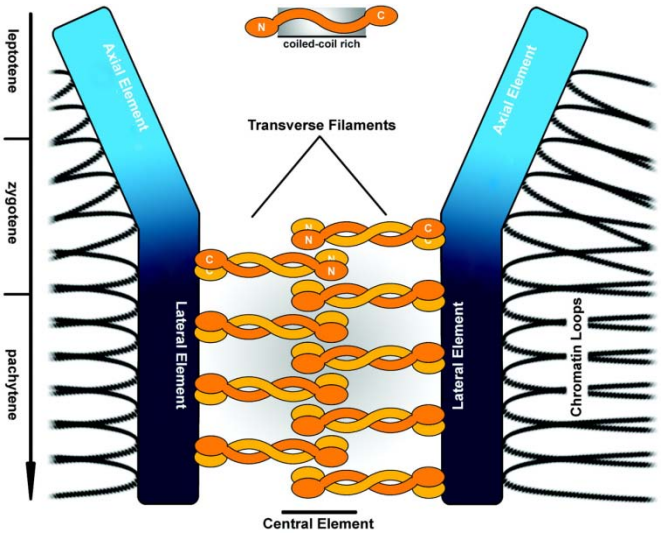

Figure 5 : Synaptonemal complex structure. During zygotene, lateral elements are formed from axial elements located along homologous chromosomes. Transverse filaments interact with lateral elements and with each other, forming the central element as well.

(Adapted from Scott L. Page *et al.*; 2003)

C/ Homologous recombination

Homologous recombination occurs at the end of the pachytene stage, after which desynapsis begins. Thus, spermatocytes enter the diplotene stage, in which chromosomes are partially separated but still connected by chiasma (i.e. connection between two non-sister chromatids). More than one connection (chiasmata) can be detected in each bivalent, indicating that several crossing overs (i.e. DNA exchange) can occur between homologs. At least, one crossing over must occur in order to ensure accurate chromosome segregation. Crossing overs are preceded by double strand breaks (DSBs) in the genome in order to initiate homologous recombination between the non-sister chromatids of the homologs (Richardson, Horikoshi et al. 2004). This is catalyzed by two types of recombination nodules: (i) early nodules (present before pachytene) are numerous and mark the sites of the initial DNA strand exchange and (ii) late nodules (arising later in meiosis) are few and mark the sites of crossing overs (Moens, Kolas et al. 2002) **(Figure 6)**. DSBs can occur anywhere along a chromosome; however they are not distributed uniformly but preferentially induced by recombination hot spots. These mechanisms are detailed in a separate chapter.

In the last stage of prophase I, diakenesis continues and involves shortening of chromosomes

and is marked by the disappearance of the synaptonemal complex with the dispersion of the nuclear membrane and the organization of fibers of the meiotic spindle.

Prophase I is followed by metaphase I which is marked by chiasmata alignment at the equator of the spindle. All bivalents are lined up at the metaphase plate in such a way that kinetochores of each homolog attach microtubules in the opposite direction. At anaphase I, the two duplicated homologs separate from each other due to cohesin degradation along the chromosome arms except at the centromere (to ensure sister partnership), move to opposite poles of the spindle caused by microtubule degradation, and then the cell divides in spermatocyte II. Each daughter cell produced contains one of the two homologous chromosomes, consisting of two sister chromatids which are identical except where recombination has occurred. During Anaphase I, sister chromatids remain attached by meiosis-specific cohesins introduced during DNA replication. At this stage, kinetochores on both sister chromatids act as a single unit, as they attach microtubules pointed in the same direction, allowing them to stay together during the separation of the homologs (Dej and Orr-Weaver 2000).

At the end of this reductional division, daughter cells are haploid since each contains half of the chromosomes, but in diploid amounts of DNA. Thus, in order to produce haploid gametes, daughter cells undergo a second division without further DNA replication. The interphase period between the two meiotic divisions is short and the events of the second equational division take place rapidly. The duplicated chromosomes align to a second spindle during metaphase II. At anaphase II, kinetochore microtubules function separately on each sister chromatids, and point in opposite directions (Dej and Orr-Weaver 2000). As a result, sister chromatids separate and move to opposite poles to produce cells with a haploid DNA content. Four haploid cells called spermatids are therefore produced from each spermatocyte that entered meiosis **(Figure 6)**.

Meiosis produces progeny that differ from the parent in their genetic make-up. A first factor of diversity comes from the random segregation of each homolog chromosome. At the time of metaphase I of meiosis, chromosomes arrange themselves randomly on either side of the equatorial plate. Each chromosome of a pair then migrates to a pole (anaphase I), without influencing the direction of migration of the chromosomes of the other pairs. Each daughter cell thus possesses a different set of chromosomes from that of the mother cell. Diversity is increased by the homologous recombination that took place during prophase I thus generating recombined chromatids that differ from parental chromatids.

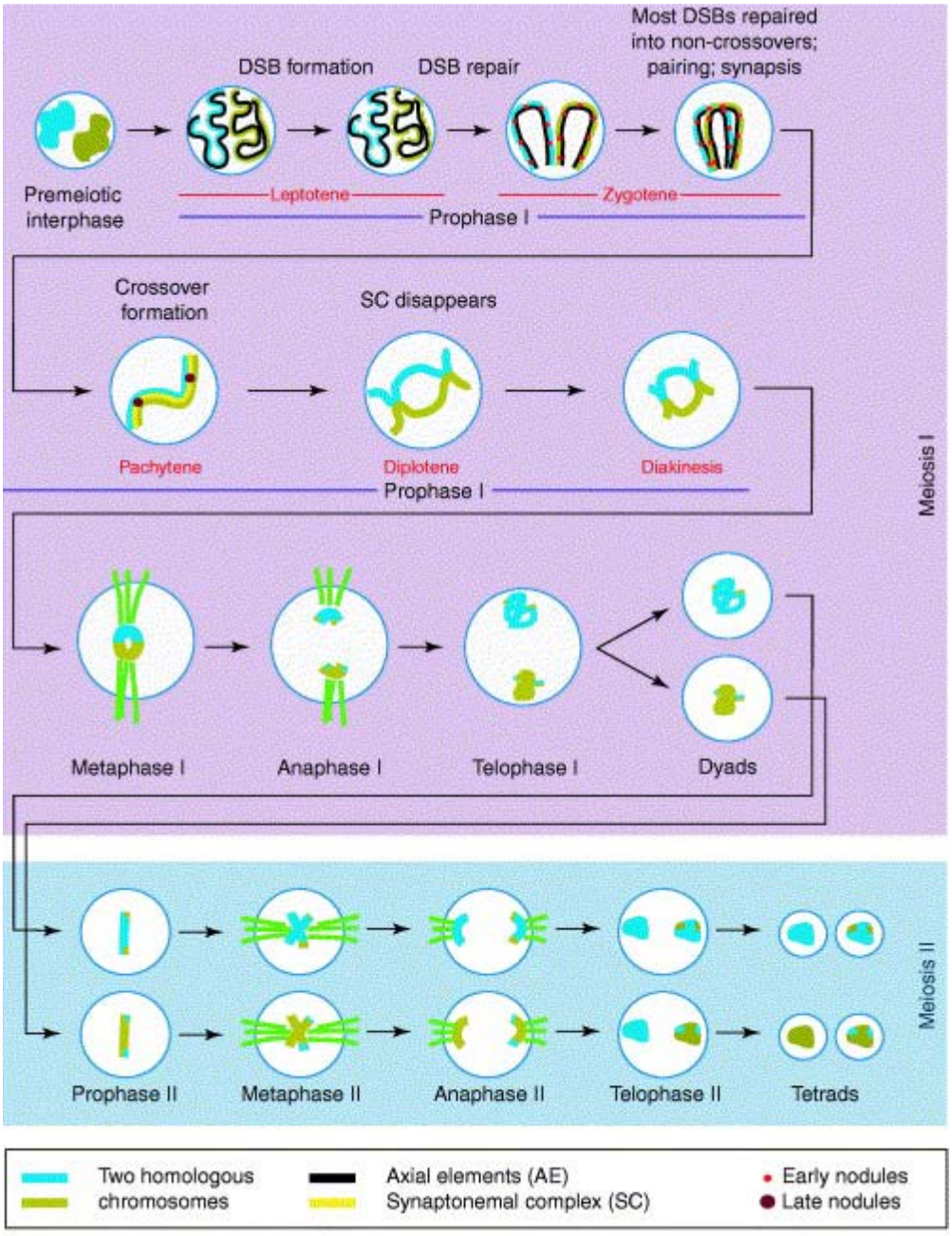

Figure 6 : Diagram showing the stages and main events in meiosis. Only one pair of chromosomes is shown, and each homolog is depicted in a different color. Early and late recombination nodules are depicted as dots of different size. The two upper panels show the extended meiotic prophase I. Major events during prophase I are depicted, including DNA double-strand break (DSB) formation and repair, crossover formation, homologous chromosome pairing and synapsis. Interactions between homologous chromosomes during prophase I lead to formation of homolog pairs (bivalents) and reciprocal exchanges of chromatid arms, as a result of crossing-over. Homologous chromosomes segregate in anaphase I. During the second division of meiosis (meiosis II), the chromatids segregate, like during a normal mitotic division.

(According to Sabrina Z. Jane *et al.* ; 2012)

3-Spermiogenesis

Spermiogenesis involves a complex series of functional and structural changes that spermatids undergo to form spermatozoa. They do not divide further but they go through a differentiation process that involves cytoplasmic and nucleic modifications. Early spermatids are round cells with spherical nucleus associated with a prominent golgi and with mitochondria arranged peripherically near the plasma membrane (Abou-Haila and Tulsiani 2000). They also contain an endoplasmic reticulum, two centrioles and dispersed ribosomes.

During differentiation, the size and the shape of the nucleus change from a sphere to a fusiform body and the nucleus shifts to the cell surface, while the chromatin condenses. This nuclear condensation is accomplished by histone modification, leading to a high degree of DNA packaging inside the sperm head. Somatic and testis-specific histones are replaced by transition proteins (TP1 and TP2) and protamines (PRM1 and PRM2) (Carrell 2012). As a result, the sperm DNA is stable and completely inactive and does not replicate until after the sperm enters the egg.

Changes in the cytoplasm concern the formation of the midpiece and the flagellum, accompanied by reorganization of several cytoplasmic organelles and even creation of novel ones. A significant cytological feature is the formation of the acrosome and the flagellum (Yan 2009). Concerning the flagellum synthesis, organelles undergo several modifications. Centrioles migrate to the opposite end of the acrosomal pole and lodge in a small depression of the nuclear envelope, where it gives rise to a thin flagellum during nucleus elongation. While one of the centrioles does not change, the other gives rise to microtubules responsible for the formation of the flagellar axoneme and the fibrous sheath of the sperm (Hermo, Pelletier et al. 2010). The central core of the axoneme, consists of nine doublet microtubules surrounding two single central microtubules, which represents a common pattern found in cilia. This basic structure is modified at the region of its contact with the nucleus through the formation of a complex structure known as the connecting piece (Hermo, Pelletier et al. 2010). Mitochondria are organized at the connecting piece, generating ATP responsible for flagellum movement. The cytoplasm moves behind the nucleus and forms the flagellum that extends very quickly. Mitochondria migrate and take position in a helical arrangement at the midpiece to provide energy to the flagellum. A thin cytoplasm surrounds the nucleus, whereas the excess cytoplasm and organelles form the residual body, which is eliminated and phagocytosed by Sertoli cells later, leaving behind a small mass called the cytoplasmic

droplet (Hermo, Pelletier et al. 2010).

The acrosome is a lysosome-like structure located over the anterior part of the spermatozoa nucleus, allowing the sperm to penetrate the zona pellucida in order to reach the cytoplasmic membrane of the oocyte (Moreno, Ramalho-Santos et al. 2000). It contains the oocyte activation factor *PLCζ* responsible for repeated oscillation of free cytosolic calcium leading to meiosis resumption (Heytens, Parrington et al. 2009). It originates from the golgi apparatus with an acidic pH that contains hydrolytic enzymes responsible of the interaction with the zona pellucida. It is comparable to lysosome, since both are derived from the golgi apparatus with an acidic pH and contain several common enzymes such as Acid glycohydrolases, Proteases, Esterases, Acid phosphatases and Aryl sulfatases, although; the acrosome presents distinctive characteristics (Berruti and Paiardi 2011). Its membrane is divided into (i) the outer acrosomal membrane (OAM), lying directly underneath the plasma membrane and (ii) the inner acrosomal membrane (IAM), overlying the nuclear envelop (Toshimori and Ito 2003). The acrosome is also distinguished by unique enzymes, such as Acrosin, Acrin1, Acrogranin and others (Moreno and Alvarado 2006). The synthesis of many proteins involved in acrosome biogenesis starts at the pachytene stage and continues until the last steps of spermiogenesis throughout the elongated spermatids (Anakwe and Gerton 1990). These proteins are packed in pro-acrosomiales vesicule that remain at the Golgi apparatus until their transportation to the acroplaxome, where they fuse. The acroplaxome is a medullary region located between the acrosomal membrane and nucleus membrane, allowing the binding of the acrosome to the top of the nucleus (Toshimori and Ito 2003).

According to Moreno & al, the formation of the acrosome can be divided into four phases: Golgi, Cap, Acrosome and Maturation (Abou-Haila and Tulsiani 2000). During the golgi phase, the proacrosomic granules formed from trans-golgi bind to the surface of the nuclear envelope at the acroplaxome and then fuse with each other to form a single sack covering the anterior part of the nuclear membrane (Moreno and Alvarado 2006). The dense microtubule array present at the cortical part of spermatids disappears. During the cap phase, the acrosomic granule increases in volume due to the merging of new acrosomic vesicules arising from the golgi apparatus, and flattens over the surface of the nucleus **(Figure 7)** (Moreno, Ramalho-Santos et al. 2000). The centriole giving rise to the flagellum moves to the opposite site of the acrosome, near the nucleus and the complex acrosome-nucleus turns upside down so that the acrosome faces the plasma membrane **(Figure 7)** (Ramalho-Santos, Schatten et al. 2002). Microtubules start to assemble on the nuclear surface on the opposite side of the

acrosome. The acrosomic sack diffuses over two-thirds of the nucleus during the acrosome phase. Microtubules are now oriented parallel to the main axis of the spermatid around the nucleus. This array of microtubules is known as the manchette, which remains attached to the nucleus until the maturation phase (Moreno, Palomino et al. 2006). Finally, at the maturation phase, the acrosome acquires functional modifications corresponding to an increase in fertilizing capacity of spermatozoa. At the end of the acrosome biogenesis, the golgi apparatus migrates to the caudal portion of elongating spermatids and is eliminated in the cytoplasmic droplet (Clermont and Leblond 1955).

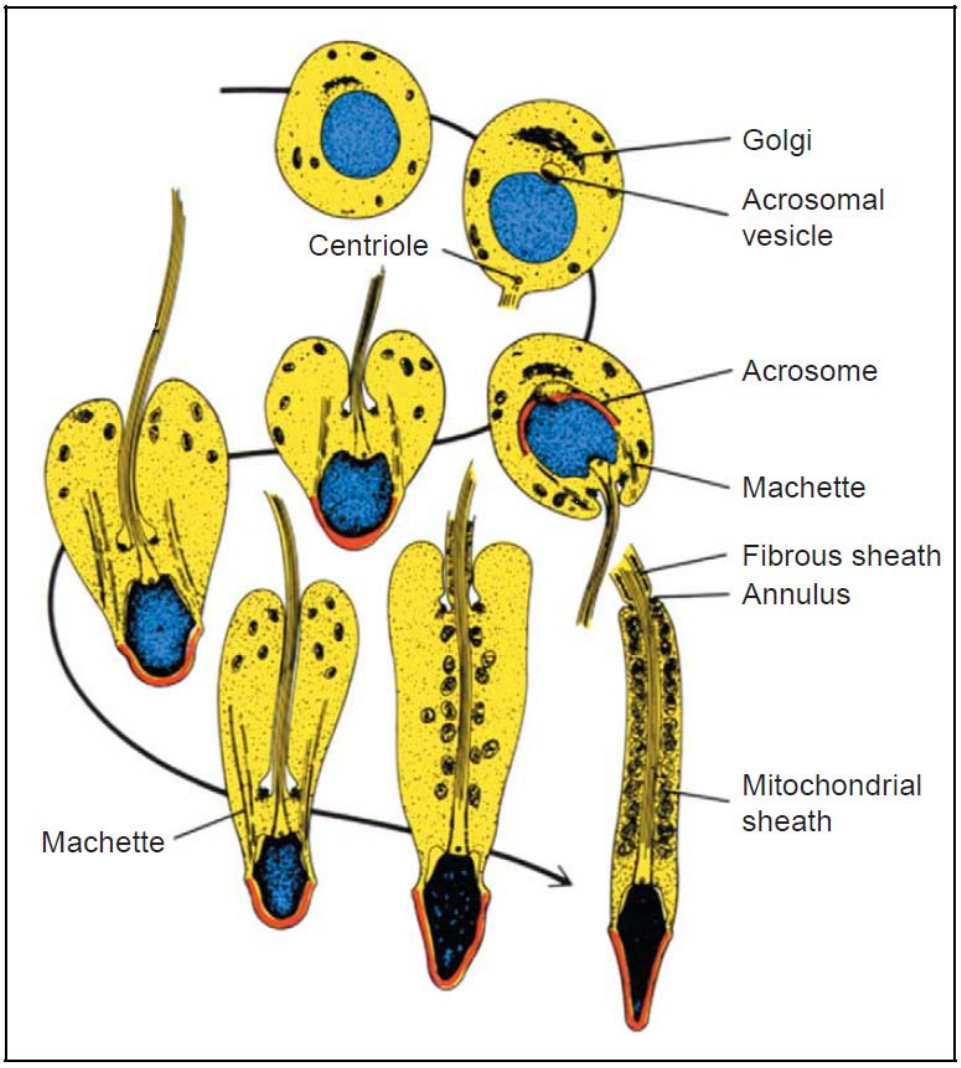

Figure 7: Steps of spermatid differentiation. (1) Immature spermatid with round shaped nucleus. The acrosome vesicle is attached to the nucleus; the tail anlage fails contact to the nucleus. (2) The acrosome vesicle is increased and flattened over the nucleus. The tail contacted the nucleus. (3–8) Acrosome formation, nuclear condensation and development of tail structures take place. The mature spermatid (8) is delivered from the germinal epithelium.

(According to Faycal Boussouar *et al.;* 2004)

The transport of acrosomial vesicles across the cytoplasm depends on two major intramanchette pathways: one via microtubules and the second via F-actin (Kierszenbaum and Tres 2004). The microtubule-based pathway requires motor proteins (kinesin/dynein) attached to vesicules via a linker, while F-actin-based traffic requires motor proteins (myosin V/myosinVII) that are attached to vesicles by a vesicle receptor (Rab27a/b) via motor recruiter proteins (MyRIP) **(Figure 8)**. Acrosomal vesicles are transported by these two pathways to the acroplaxome, where they fuse to form the acrosome. Also non acrosomal vesicles are transported to the centrosome region and the developing spermatid tail using the same traffic.

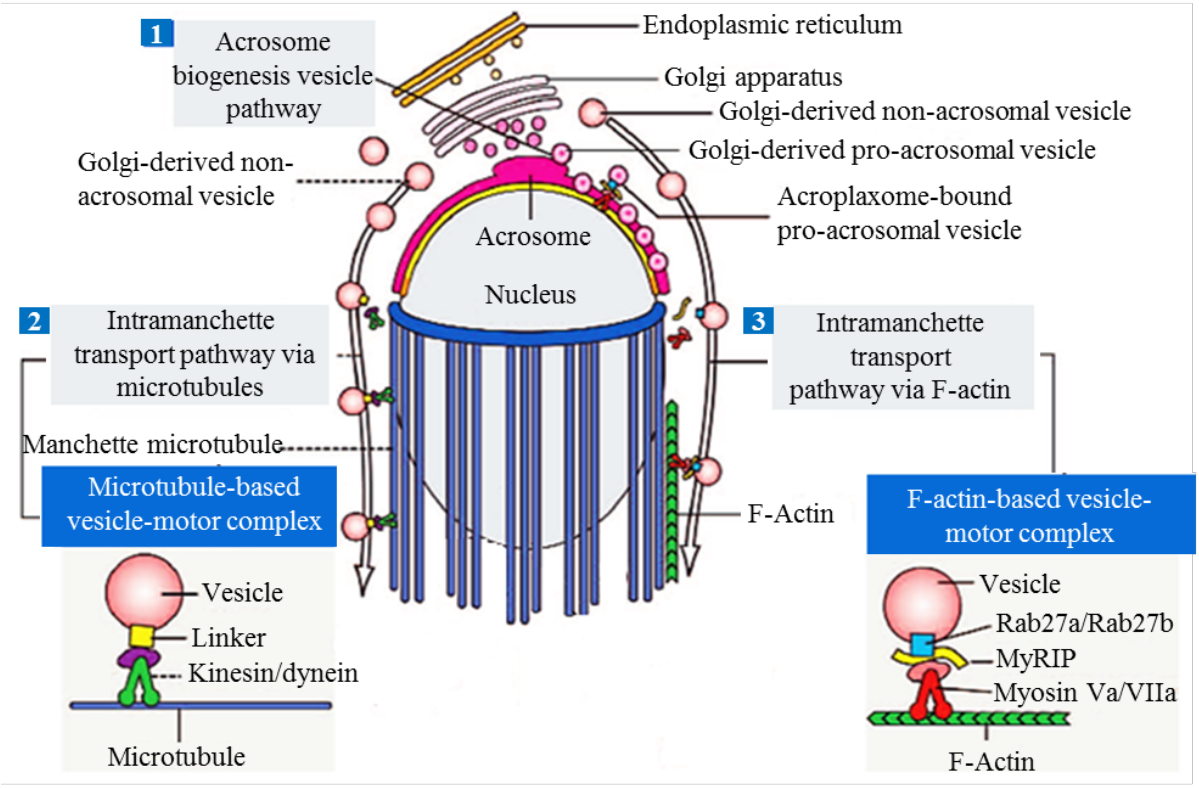

Figure 8 : Representation of the two intramanchette vesicule pathways. 1: The acrosome biogenesis vesicle pathway includes transporting vesicles form the endoplasmic reticulum to the Golgi. The Golgi generates two types of vesicles: pro-acrosomal and non-acrosomal vesicles. Proacrosomal vesicles are transported to the acroplaxome by actin-based and microtubule-based molecular motors, where they fuse and organize the acrosome. 2: non-acrosomal vesicles associate with microtubules of the manchette and are transported by microtubule-based molecular motors to the centrosome region and the developing spermatid tail (not shown). 3: Non acrosomal vesicles can also associate with F-actin present in the manchette and transported by F-actin-based molecular motors.

(Adapted from Abraham L. Kierszenbaum *et al.* ; 2004)

Membrane trafficking also involves the N-ethylmaleimide sensitive factor (NSF) that plays an important role in intracellular fusion via the recruitment of soluble NSF attachment proteins (SNAPs). A large family of membrane-bound SNAP receptors (SNAREs) ensure the docking and fusion of vesicles with the appropriate target (Duman and Forte 2003). The model involves the pairing of two complementary SNAREs: i) the v-SNARE at the vesicle membrane ii) and the t-SNARE at the target membrane, thus avoiding vesicle fusion before the completion of meiosis. However, this distribution is not exclusive, as some v-SNARE may exist at the target membrane as well as the t-SNARE at the vesicle membrane (Duman and Forte 2003). The association of complementary SNAREs brings the two heterotypic membranes into close contact leading to fusion.

Recent work showed that in addition to the Golgi apparatus, the plasma membrane and endocytic trafficking are also involved in acrosome formation, since a disruption of the Golgi by brefeldin A treatment does not affect the acrosome biogenesis. The acrosome formation associated factor (Afaf) is an example that illustrates the extra-Golgi trafficking in acrosome biogenesis. Afaf is located in the plasma membrane and is an early endosome vesicle membrane protein that localises in the inner and outer membrane of the acrosome (Li, Hu et al. 2006).

Finally, an opposite trafficking pathway may also exist: the acrosome-to-Golgi pathway. This hypothesis is strengthened by the fact that several types of Golgi proteins present in developing acrosome are no longer detected in the mature spermatozoa. This suggests that recycling occurs during the final stages of spermatid maturation, given that the Golgi is still active until its elimination in the cytoplasmic droplet (Moreno, Ramalho-Santos et al. 2000; Ramalho-Santos, Moreno et al. 2001).

At the end of spermiogenesis, spermatozoa are released from the seminiferous epithelium into the lumen of the tubule. This process is known as spermiation. The resulting mature spermatozoa are transported to the epididymis in testicular fluid secreted by the Sertoli cells. There, they gain motility and fertilization capacity (Kopera, Bilinska et al. 2010). The long process of spermatid metamorphosis is subdivided into 8 distinct steps in human and extends over 21.6 days **(Figure 7)**. In mice and rats, 16 and 19 steps are defined respectively that extend over 22.7 days in both rodents (Hermo, Pelletier et al. 2010).

Note that during mitotic and meiotic phases and even during early spermiogenesis, germ cells remain connected to each other by cytoplasmic bridges due to incomplete cytokinesis (Jan, Hamer et al. 2012). This connection supplies the essential products of a complete diploid

genome to the developing haploid cell. At the end of the first meiotic division, resulting cells contain one version of each chromosome. Thus developing sperm carrying a Y chromosome, for example, can be supplied with essential proteins encoded by genes on the X chromosomes. Cytoplasmic bridges remain till spermiation (Ehmcke and Schlatt 2006). Moreover, at all stages of spermatogenesis, germ cells are in close contact with Sertoli cells until the release of gametes in the lumen of the tubules. This connection provides structural and metabolic support to the developing sperm cells (Kopera, Bilinska et al. 2010).

Chapter III

DNA double strand breaks

As was mentioned in the first chapter, crossing over (CO) occurs at the pachytene stage during the first meiotic division, allowing the exchange between the paternal and maternal chromosomes. This is crucial for the first meiotic division. COs ensure an accurate chromosome segregation, since they are the basis of the physical linkages that facilitate the correct orientation of the homologs on the first meiotic spindle (Page and Hawley 2003). This central event during meiosis is mediated by homologous recombination and is initiated by double strand breaks (DSBs) induced by SPO11 topoisomerase (conserved from yeast to human). Although, DSBs are crucial during meiosis, they are a threat to the genomic integrity if they are not repaired. Reparation takes place, either through homologous recombination (HR), non-homologous end-joining (NHEJ) or by single strand annealing (SSA) **(Figure 9)**.

1- Mechanisms of repair

A/ Homologous recombination (HR)

Homologous Recombination relies on extensive sequence homology to copy the genetic information for the damaged DNA strand from the undamaged homologous DNA molecule. Homologous recombination can be divided into three stages: i) presynapsis, ii) synapsis and iii) postsynapsis (Heyer, Ehmsen et al. 2010). In presynapsis, the DNA damage that occurred is processed by nucleolytic resection to result in 3'-OH single stranded tails. During synapsis, a recombinase filament is formed on the ssDNA ends, which starts a strand invasion into an intact DNA duplex of a homologous sequence to form a displacement loop (D-loop).

The D-loop is a DNA structure where the two strands of a double-stranded DNA molecule are separated for a stretch and held apart by a third strand of DNA (Martinez-Perez and Colaiacovo 2009). The third strand has a base sequence which is complementary to one of the main strands and pairs with it, thus displacing the other main strand in the region. Finally, in postsynapsis, three subpathways of HR are distinguished: 1- break-induced replication (BIR), 2- synthesis-dependant strand annealing (SDSA), 3 – double Holliday junction (dHJ), and repair is achieved by DNA synthesis and ligation **(Figure 9)**.

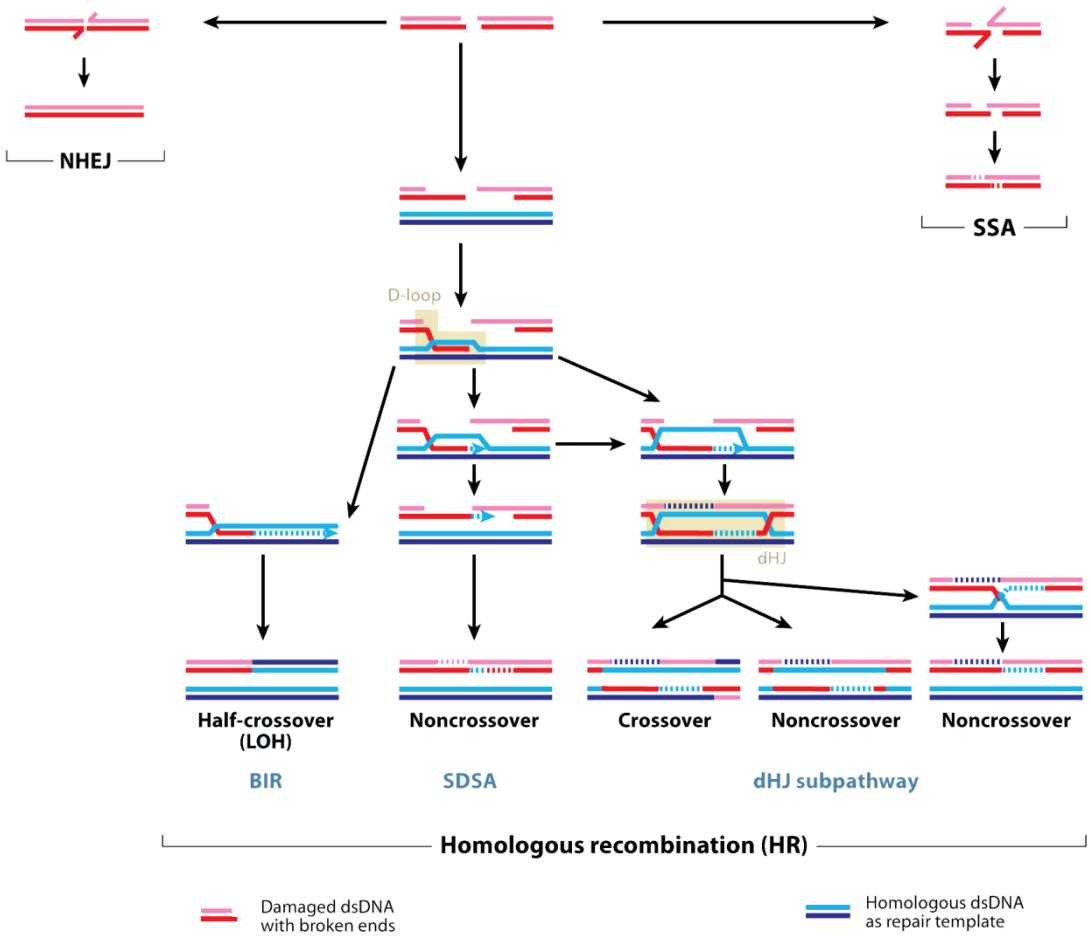

Figure 9 : Summary of different pathways of double-strand break repair. Broken lines indicate new DNA synthesis and stretches of heteroduplex DNA that upon mismatch repair (MMR) can lead to gene conversion. Abbreviations: BIR, break-induced replication; dHJ, double Holliday junction; NHEJ, non-homologous end joining; LOH, loss of heterozygosity; SDSA, synthesis-dependent strand annealing; SSA, single-strand annealing.

(Adapted from Wolf-Dietrich Heyer *et al.* ; 2010)

a. Break-induced replication (BIR)

BIR begins in the presence of one free DNA end, or because only one of two ends of the DSB succeeds in strand invasion of a homologous sequence.

Three mechanisms are proposed (Kraus, Leung et al. 2001), which differ in the modeling of the migration of the D-loop and in later phases of recombination. In a first scenario, the D-loop migrates down the template and fills the single strand, which subsequently can be made double-stranded (Kraus, Leung et al. 2001) **(Figure 10 A)**. Another proposition is that the D-loop becomes a replication fork, then migrates down the template molecule. The single strand extends by processivity polymerization, using a unique strand of the intact molecule as a template. At the same time the synthetized DNA become double-stranded by replication of the template strand of the homolog **(Figure 10 B)**. This will result in semiconservative synthesized molecules. Alternatively, in the third mechanism, replication occurs simultaneously during invasion and extension of the single strand using the intact homolog as a template **(Figure 10 C)**.

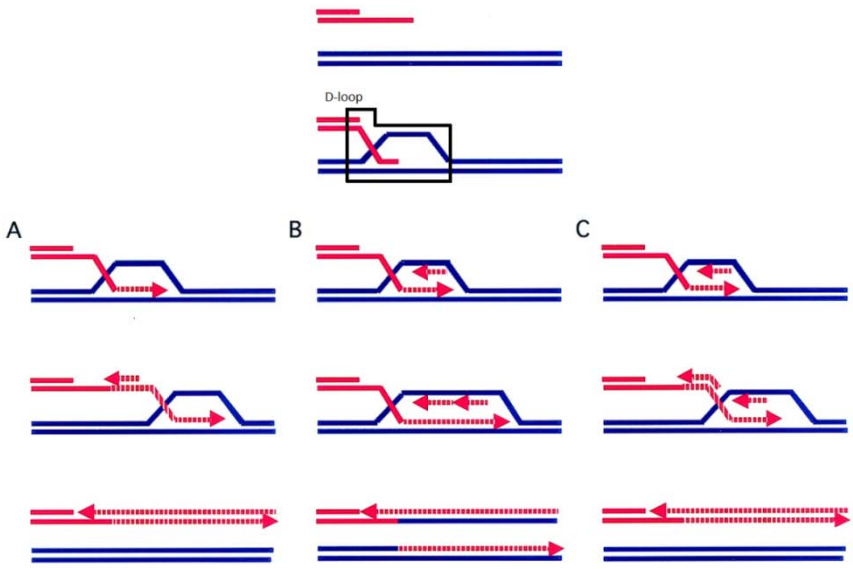

Figure 10 : Alternative BIR mechanisms. A broken chromosome end will be resected by 5′ to 3′ exonucleases, allowing the 3′ end to interact with various recombination proteins to carry out strand invasion. (A) The 3′ end of the invading strand initiates DNA replication, leading to a migrating D-loop "bubble" as described by Formosa and Alberts (5). The displaced newly synthesized DNA strand can then be made double-stranded. (B) Strand invasion sets up a replication fork that will result in semiconservatively synthesized molecules. A Holliday junction will be resolved at some point. (C) Strand invasion sets up a replication fork in which branch migration enzymes displace both newly synthesized DNA strands as the replication structure migrates down the template.

(According to Eliyahu Kraus *et al.* ; 2001)

b. Synthesis-dependant strand annealing (SDSA)

In this model, after the formation and the recognition of DSB, the migrating D-loop never leads to a capture of the second DSB end (San Filippo, Sung et al. 2008; Sasaki, Lange et al. 2010). Alternatively, following the extension of the invading 3' strand along the recipient DNA duplex by a DNA polymerase, and its reparation via DNA synthesis, the invading strand is released and anneals with the second resected DSB end **(Figure 11)**. Because no Holliday junctions are formed, SDSA results in non-crossover products, which reduce genomic rearrangements (San Filippo, Sung et al. 2008; Sasaki, Lange et al. 2010).

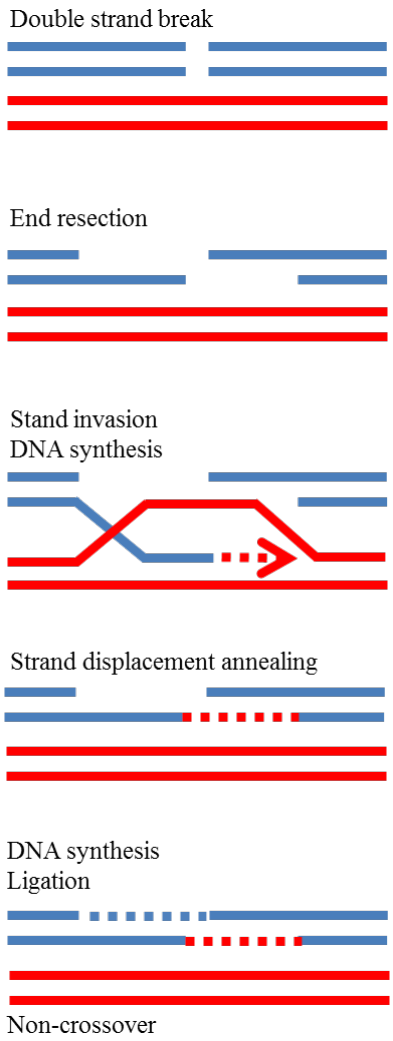

Figure 11 : DSB repair via SDSA. Repair is initiated by resection of a DSB to provide 3' single-stranded DNA (ssDNA) overhangs. Strand invasion by these 3' ssDNA overhangs into a homologous sequence is followed by DNA synthesis at the invading end. The reaction proceeds to SDSA by strand displacement, annealing of the extended single-strand end to the ssDNA on the other break end, followed by gap-filling DNA synthesis and ligation. The repair product from SDSA is always non-crossover.

(Adapted from Sung, P. et al.; 2006)

c. Double Holliday junction (dHJ)

The double Holliday junction is also known as the double strand break repair pathway (DSBR). It involves the formation of Holliday junctions (mobile junctions between four strands of DNA) that result in crossover or non-crossover depending on how the junctions are resolved. After a DSB occurs and is catalysed by SPO11, the 5' ends are resected to give 3'-OH ends. These ends are coated with DNA recombinases (RAD51/DMC1 in eukaryotes, RecA in prokaryotes) in order to catalyse the invasion of one 3'end into the homologous sequence, forming a D-loop. Then, a DNA extension starts from the 3'end of the invading strand, followed by the capture of the second DSB end **(Figure 12)**.

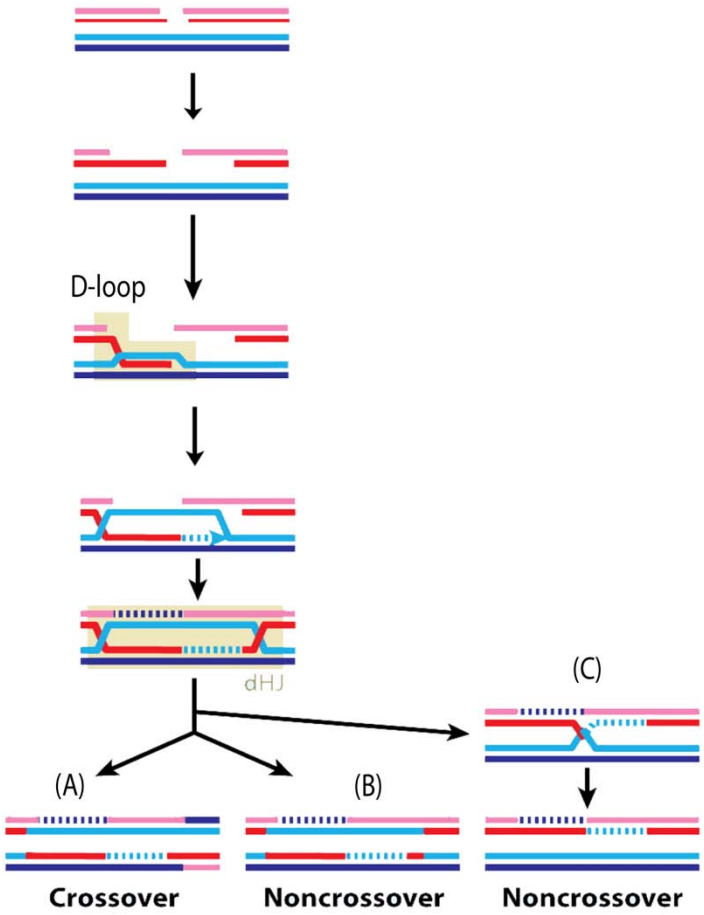

Figure 12 : DSB repair via double holliday juncton. After DSB formation, the DNA ends are resected and execute strand invasion of a partner chromosome to form a nascent D-loop structure. The second DSB end is captured to form an intermediate that harbors two Holliday junctions (HJ)s, accompanied by a gap filling DNA synthesis and ligation. A) The resolution of HJs results in crossover. B) The resolution of HJs results in noncrossover products. C) When the two HJs migrate to each other only noncrossover products are obtained.

(Adapted from Wolf-Dietrich Heyer *et al.* ; 2010)

The latter anneals to the D loop and also extends. Subsequent to ligation, a double Holliday junction is formed (Bhalla and Dernburg 2008). Lastly, resolution of dHj by endonucleases is done in two ways. If the two junctions are resolved in the same orientation, this will lead to non-crossover products and if they are resolved in different orientation this will lead to crossover products (Phadnis, Hyppa et al. 2011). Alternatively, dHJ resolution can be mediated by a complex mechanism involving helicase and topoisomerase, allowing the migration of the two junctions toward each other **(Figure 12)**. The resulting hemicatenane is then eliminated, generating only a non-crossover outcome (Heyer, Ehmsen et al. 2010).

B/ Non-homologous end-joining (NHEJ)

In contrast to homologous recombination, non-homologous end joining (NHEJ) is a pathway that repairs DSB without the need for a homologous template. It proceeds in four steps that involve the recognition of the DSB, followed by bridging of both broken DNA ends and tracked by modification of the ends by adding or removing some nucleotides in order to make them compatible for ligation in a final step **(Figure 13)**. Thus, NHEJ is termed error-prone since it results in imprecision, leaving an information scare at the repaired site (McVey and Lee 2008; Heyer, Ehmsen et al. 2010). Therefore, NHEJ restores molecular integrity without ensuring sequence information in the DNA.

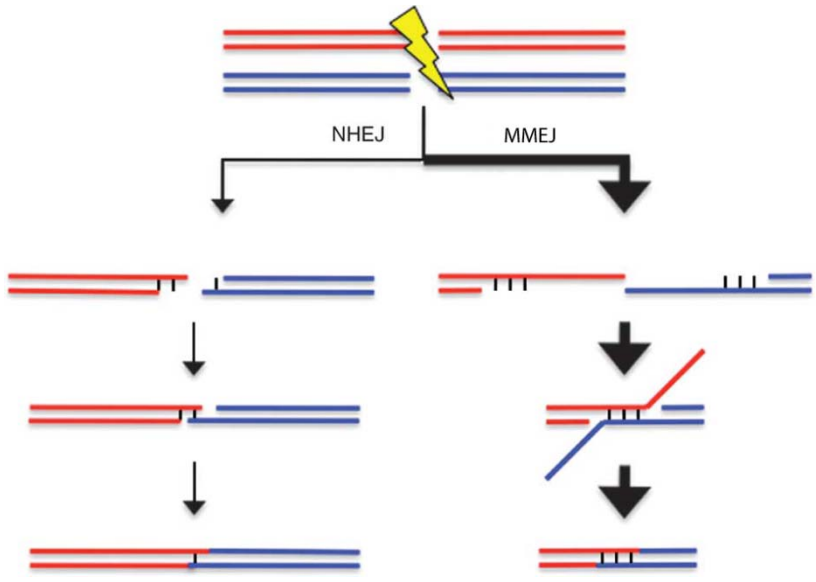

Figure 13 : DSB repair via Non-homologous end-joining (NHEJ) and microhomolgy mediated end joining (MMEJ).

(Adapted from Zhang *et al*. 2011)

In addition to the NHEJ repair pathway, a second distinct form of end joining has been proposed, termed microhomology-mediated end joining (MMEJ). This process involves the presence on both strands of 5 to 25 bp long homologies, which are identified when DNA break occurs. Microhomologies are first exposed in order to anneal, then remaining non complementary DNA are removed before ligation (McVey and Lee 2008; Hastings, Lupski et al. 2009). Hence, MMEJ results in small deletion and insertion products flanking the original break. This new form of NHEJ is also known as B-NHEJ for Back-up NHEJ. Currently, evidence exists that shows that this form of end joining occurs only when the classical NHEJ is defective or fails to engage at certain DSBs (Iliakis, Wang et al. 2004).

C/ Single strand annealing (SSA)

SSA is considered to be a transitional pathway between the NHEJ and HR. It repairs double-strand breaks between two repeated sequences oriented in the same direction. It does not require a homologous DNA molecule, like the dHJ or SDSA pathways of homologous recombination (Hastings, Lupski et al. 2009). In this mechanism, after the detection of DSB, 5' to 3' nucleolytic strand resection can extened for multiple kilobases before a homologous template is found. One repeated sequence on each strand is used to allow the annealing of the resected ends **(Figure 14)**.

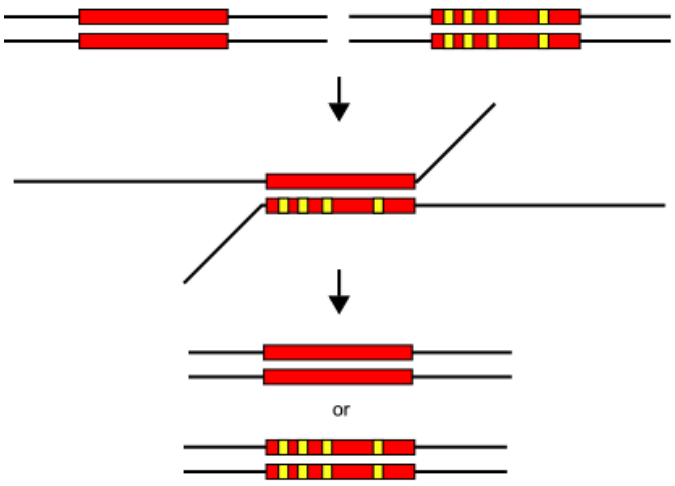

Figure 14 : DSB repair via Single Strand Annealing (SSA). Following DNA breakage at a position flanked by short stretches of homologous sequence shown as boxes, 3′ 5′ -end resection exposes single-stranded ′3 -overhangs with embedded complementary DNA sequence Annealing of complementary sequences produces unpaired "gaps" of noncomplementary single-stranded DNA These are trimmed by a flap endonuclease, and ligation of the DNA double-strand break (DSB) or staggered single-strand nicks restores the DNA double strand, accompanied by loss of a single repeat and flanking DNA sequence

After annealing, non-homologous 3' flaps are digested, DNA gaps are filled and the process is finished by ligation, which restores the DNA duplex as two continuous strands. The DNA sequence between the repeats is always lost, as is one of the two repeats (San Filippo, Sung et al. 2008).

2- Competition between HR, SSA and NHEJ

The balance between the three pathways depends on the species, cell type, cell cycle stage and the type of DNA damage. Homologous recombination is a template dependent process that requires two DNA molecules (sister chromatids, chromosomes), whereas NHEJ and SSA can occur within the context of single DNA molecules. In mitotic cells, repair involves the three pathways depending on the placement of the DSB (Lieber, Ma et al. 2003). NHEJ comes into play when DSB is located at the end of DNA molecules (near the telomeres) when in fact, SSA promotes recombination of tandem repeat DNA sequences. When HR is recruited in mitotic cells, the sister chromatid is used as a template. Studies in *S.cervisiaie* showed that during the G1 phase, the homologous chromosome is used as a template, whereas in S and G2 phase, the sister chromatid is preferred over the homolog (Pradillo and Santos 2011). The three subpathways of homologous recombination also compete when the latter is initiated. During mitotic DSBs repair, SDSA is the preferred subpathway (Heyer, Ehmsen et al. 2010). In fact, BIR can result in two identical alleles carrying a deleterious mutation due to the loss of heterozygosity. It is also a slow process by reason of its repression at the DNA synthesis step for five hours after DSB formation. In addition, dHJs can cause genomic rearrangements if they occur in non-allelic sites and are formed at low levels which correlate with crossing over outcomes during mitosis.

In meiotic cells, the preferred mechanism is homologous recombination, where invasion of the homolog is stimulated (Lieber, Ma et al. 2003). SDSA used in meiotic DSBs repair results in noncrossover products, while the dHJ subpathway is established only when strand exchanges occur (Heyer, Ehmsen et al. 2010). Obviously, at meiosis, DNA repair processes are necessary for accurate segregation of chromosomes and for generating new combinations of linked alleles. However, recombination is also potentially dangerous as it can lead to gross chromosomal rearrangements (Kolodner, Putnam et al. 2002).

Chapter IV

Genomic rearrangements

Genomic rearrangements are DNA changes in the genome, ranging from hundred base pairs to megabases, that include deletion, duplication, translocation, insertion and inversion. They can cause variation in the number of copies of chromosomal segments that may include genes. This is known as copy number variation (CNV), which can be either neutral in function where they represent polymorphisms, or can also cause genomic disorders due to non-tolerated gene alterations. Hence, genomic disorders are associated with changes in chromosome structure, thus differing from the traditional Watson-Crick base pair modifications (Lupski and Stankiewicz 2005), where monogenic point mutations reflect errors of DNA replication and/or DNA repair (Lupski 1998). The architectural structure of the genome can occur via homologous recombination and non-homologous recombination (the two mechanisms of DSB repair), leading to recurrent and non-recurrent rearrangements respectively. The former is characterized by a common size and fixed breakpoints in all events, while the latter is represented by different sizes and distinct breakpoints for each event (Gu, Zhang et al. 2008).

1- Recurent rearrangements

The three subpathways of homologous recombination described previously can generate CNV during DSB repair. In dHJ and SDSA processes, structural changes are caused by a non-allelic homologous recombination (NAHR) (Hastings, Lupski et al. 2009). This results when homologous sequences in different chromosomal positions are used as templates, rather than the homologous sequence at the same chromosomal position in the sister chromatid or homolog (Stankiewicz and Lupski 2002). In recurrent rearrangement, NAHR is mediated by low copy repeats (LCR) (also known as segmental duplication (SD)) with recombination

hotspots (Gu, Zhang et al. 2008). LCRs are defined as continuous portions of DNA that map to two or more genomic locations (Bailey and Eichler 2006). They are 10 to 300 kilobases in size and share more than 95% identity (Stankiewicz and Lupski 2002).

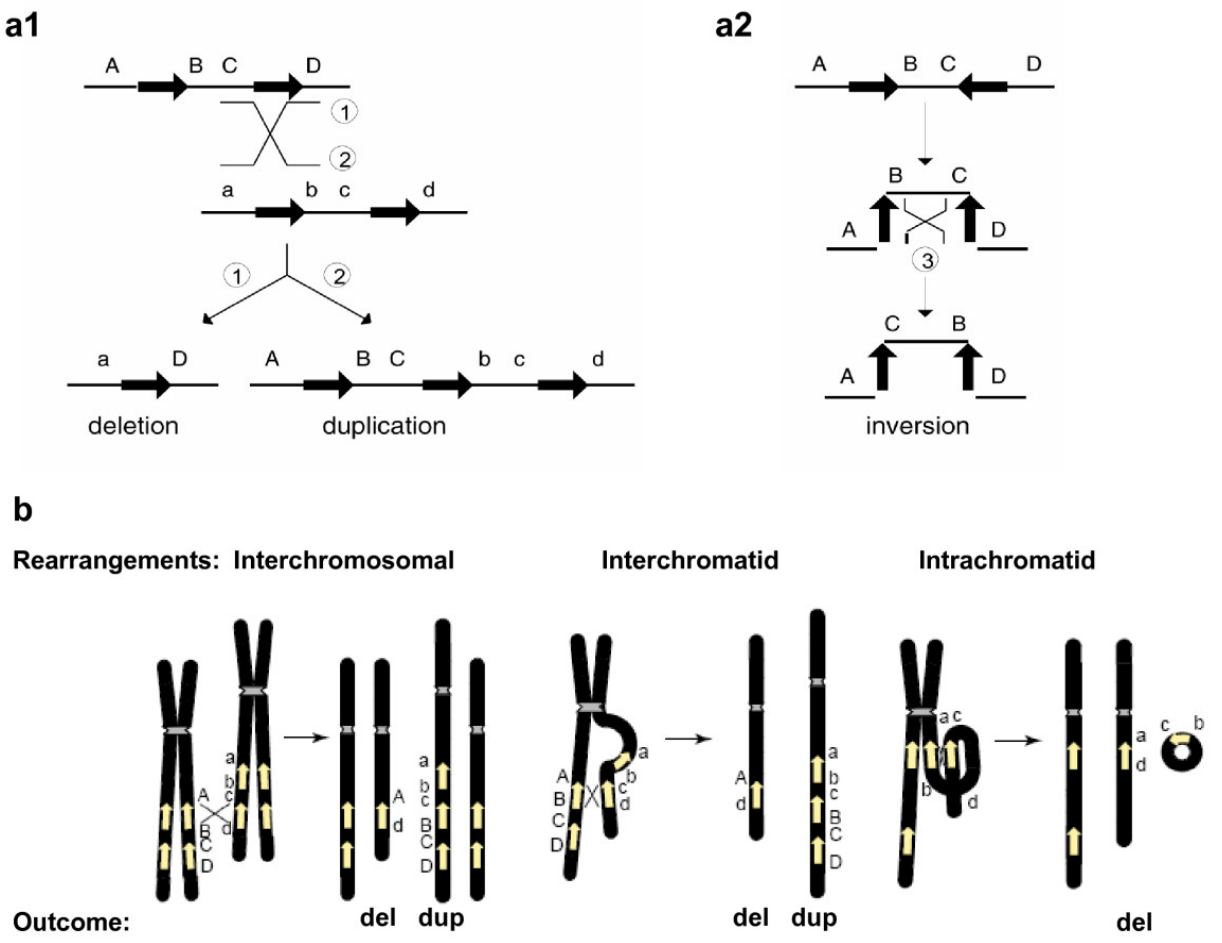

Figure 15 : Genomic rearrangements. **a1** and **a2** Genomic rearrangements resulting from recombination between low-copy repeats (LCRs). LCRs are depicted as black arrows with the orientation indicated by the direction of the arrowhead. Capital letters above the thin horizontal lines refer to the flanking unique sequences (for example, A). Homologues on the other strand (can be another chromatid or the homologous chromosome) are also shown (for example, a). Thin diagonal lines refer to a recombination event with the results shown by numbers 1, 2 and 3. **a1** Recombination between direct repeats results in deletion and/or duplication. **a2** Recombination between inverted repeats results in an inversion. **b**. Schematic representation of reciprocal duplications and deletions mediated by interchromosomal (left), interchromatid (middle) and intrachromatid (right) non-allelic homologous recombination (NAHR) using LCR pairs in direct orientation. Chromosomes are shown in black, with the centromere depicted by hashed lines. Yellow arrows depict LCRs. Letters adjacent to the chromatids refer to the flanking unique sequence (for example, A, a). Interchromosomal and interchromatid NAHR between LCRs in direct orientation result in reciprocal duplication and deletion, whereas intrachromatid NAHR only creates deletion. Signatures of homologous recombination include the sequence identity of the substrates (LCRs) used for NAHR, recombination hotspots within the LCRs, and evidence for gene conversion at the crossovers within the LCRs.

(According to Gu et al.; 2008)

In terms of sequence content, segmental duplications can contain a variety of highly homologous repetitive sequences and normal constituents of genes: exons, introns, promoters and enhancers. NAHR during interchromosomal, intrachromosomal or intrachromatid recombination leads to deletion or duplication if it is occurs between two LCR in direct orientation, while recombination between two inverted LCR leads to inversion **(Figure 15)** (Lupski 1998).

When NAHR takes place, it happens in a region of the LCRs that shares a maximum of identity, named minimal efficient processing segments (MEPS), that appear to be in the range of 300 to 500 bp in length in humam meiosis (Reiter, Hastings et al. 1998). Reducing the MEPS by 2 mismatches results in a reduction of recombination (Inoue and Lupski 2002).

Besides dHJ and SDSA, BIR also contributes to genomic rearrangements. It can lead to loss of heterozygosity distal to the DSB if the broken end invades a homolog instead of a sister molecule. Occasionally, BIR results in translocation, duplication or deletion when the repair involves a homologous sequence in a different chromosomal position (Smith, Llorente et al. 2007). Thus, it is considered to be an alternative mechanism for NAHR.

2- Non-recurrent rearrangements

A/ Via DSBs repair

a. Homologous repair

Despite the fact that NAHR accounts for most recurrent rearrangements, it also accounts for some of the non-recurrent rearrangements. In this case, LCRs do not mediate but stimulate the recombination, which is mediated by retrotransposons (SINEs, LINEs) within the LCRs (Stankiewicz, Shaw et al. 2003; Shaw and Lupski 2004).

Retrotransposons are genetic elements that can amplify themselves in a genome via RNA intermediates carrying with them the reverse transcriptase (agent of their mobility). They constitute ~ 40% of the human genome and can be broken down into two groups: long terminal repeat retrotransposons (LTR retrotransposons) and non-LTR retrotransposons (Lander, Linton et al. 2001). LTR refer to the presence of flanking repeat sequences. The first class uses RNA as a template and the reverse transcriptase to synthesize a double-stranded

DNA, which is then inserted into the host chromosome via a recombination event (Beauregard, Curcio et al. 2008). The second class lacks the terminal repeats. They encode endonucleases and reverse transcribe a copy of their RNA template directly into the chromosome (Luan, Korman et al. 1993). Long-interspersed nuclear elements (LINEs) and short-interspersed nuclear elements (SINEs) are members of the mammalian non-LTR retrotransposons. LINEs make up ~20% of the genome and are categorized into major groups of LINE1, LINE2, and LINE3 (Ostertag and Kazazian 2001). LINE1 composes the majority of the repeats (~17%) in the human genome and is the only active one (Lander, Linton et al. 2001).

Alu sequences, which are SINE family members, are the most frequent interspersed elements in the human genome (~11%) (Rowold and Herrera 2000; Bailey, Yavor et al. 2001). They are approximately 300bp in size, GC rich and have a very low pairwise divergence (<1%). Hence, many sites of near-perfect identity are created throughout the human genome for NAHR to occur (Batzer and Deininger 2002; Bailey, Liu et al. 2003). NAHR mediated by Alu is the main example of non-recurrent rearrangement. In fact, it has been shown that some non-recurrent deletions of Smith-Magenis syndrome (SMS) patients can be mediated by NAHR between *Alu* sequences (Shaw and Lupski 2005).

In addition, sequence analysis of several studies indicated that a portion of NAHR events utilize repetitive elements (SINEs, LINEs, LTRs), rather than LCRs as homology substrates (Korbel, Urban et al. 2007; Kidd, Cooper et al. 2008).

b. Non homologous repair

Other non-recurrent rearrangements can be explained by non-homologous DSB repair mechanisms. Indeed, NHEJ was proposed as a mechanism of non-recurrent rearrangement when analyses of junctions in a genomic disorder showed microhomology in some cases, short insertions in other cases and even some sequences of unknown origin. This evidence fits well with the features of NHEJ mechanism. Data also showed that the resulting breakpoints do not cluster within LCRs, but, they were dispersed throughout the causative gene and fell within repetitive elements such as LINE, SINE or in proximity of sequences motifs (TTTAAA) known to cause DSB. Furthermore, some other findings suggest that NHEJ may still be stimulated by certain genomic elements like LCRs, but do not depend directly on these genomic architectures. Indeed, Shaw and Lupski reported two non-recurrent SMS caused by NHEJ and identified the two breakpoints located in the LCR (Shaw and Lupski 2005).

Another example is the study of Stankiewicz *et al.* in which they identified many breakpoints of 17p translocations located within LCRs (Stankiewicz, Shaw et al. 2003). On the other hand, NHEJ was proposed by Woodward *et al.* and Lee *et al.* to explain duplications. They suggest that when DSB occurs, one of the broken ends invades and copies the sister chromatid causing the duplication before rejoining the ends via NHEJ (Woodward, Cundall et al. 2005; Lee, Inoue et al. 2006).

B/ Via DNA replication

Although the presence of microhomology was assigned as a signature of NHEJ, the formation of microhomology junctions is linked in some cases to DNA replication which may lead to structural chromosomal changes (Hastings, Ira et al. 2009).

Replication or DNA synthesis starts at specific points, called origins of replication after unwinding the parental DNA strands by helicase generating Y-shaped replication forks **(Figure 16)**.

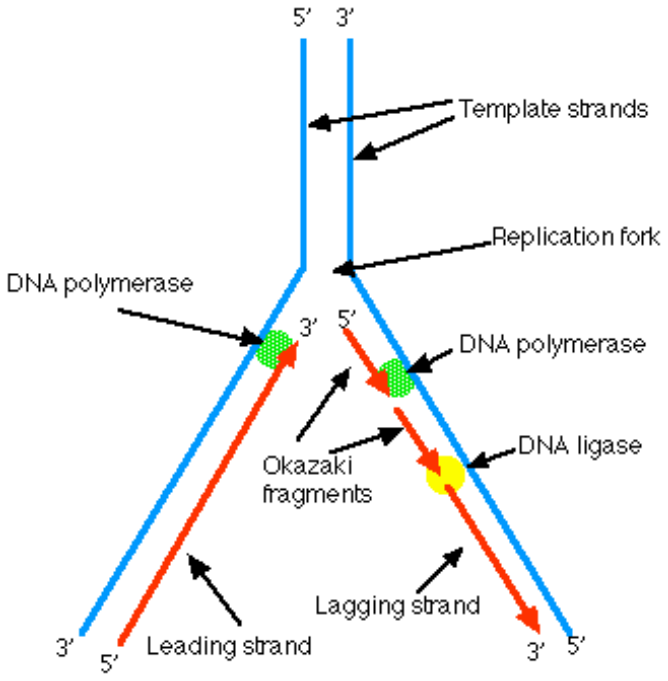

Figure 16 : DNA replication. The enzyme helicase opens up a replication fork, where synthesis of new daughter DNA strands can begin. The overall direction of movment of the replication fork matches that of the continuous 5'→3' synthesis of the leading daughter DNA strand. The lagging strand, which is synthesized in the opposite direction, is built up in pieces (Okazaki fragments).

The synthesis of complementary daughter strands occurs in opposite directions of the anti-parallel parental DNA strand that serves as a template for the DNA polymerase. Two directions of chain growth are observed: a 5'→3'for the leading strand and a 3'→ 5' for the lagging strand. The overall direction of movement of the replication fork matches that of the continuous 5'→3'synthesis of the leading daughter DNA strand. The continuous elongation of the leading strand is due to the presence of a free 3'hydroxyl group, always allowing the addition of a deoxynucleoside monophosphate residue. However, the direction of the lagging strand is opposite to the movement of the replication fork and thus generates DNA segments that are typically 100-1000 nucleotides long known as Okazaki fragments **(Figure 16)**. Synthesized fragments are then stitched together covalently by DNA ligase to ensure the creation of two complete daughter DNA duplexes. Hence, DNA replication is a semi-conservative and semi-discontinuous process.

Three replicative mechanisms have been proposed to be responsible for CNV: 1- replication slippage, 2- Fork stalling and template switching (FoSTeS), 3- Microhomology-mediated break induced replication (MMBIR).

a. Replication slippage

Replication slippage occurs when a length of the lagging strand becomes exposed as a single strand, causing the movement of the 3' end to another sequence of the exposed strand that shares a short length of homology **(Figure 17)**. Then, DNA synthesis continues after having missed a part of the template (Hastings, Lupski et al. 2009).

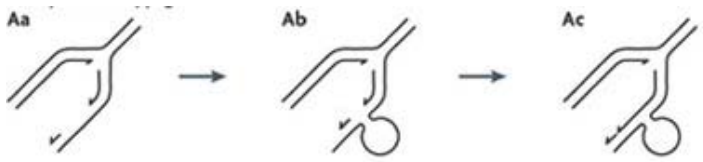

Figure 17 : Replication slippage. During replication, a length of lagging-strand template becomes exposed as a single strand (Aa). The 3' primer end can move to another sequence showing a short length of homology on the exposed template (Ab); this move might occur owing to the formation of secondary structures in the lagging-strand template. Lagging strand synthesis can continue after having failed to copy part of the template (Ac). As shown, this will produce a deletion. Several variations on this mechanism can also produce a duplication of a length of DNA sequence with or without sister chromatid exchange.

(According to P. J. Hastings *et al.* ; 2009)

b. Fork stalling and template switching (FoSTeS)

This model was proposed by Lee et al inspired by findings in *Escherichia coli* (Slack, Thornton et al. 2006). When the DNA replication fork stalls (due to secondary structures in the lagging template that block the progression of the replication fork), the 3' end moves from the original template, transfers and anneals to microhomology at the 3' end of another replication fork in physical proximity and restarts DNA synthesis **(Figure 18)**. Deletion results if switching occurs to another fork located downstream (forward invasion), while switching to an upstream replication fork results in duplication (backward invasion). Serial replication fork disengaging and lagging strand invasion could occur multiple times in series, before resumption of replication on the original template (Gu, Zhang et al. 2008).

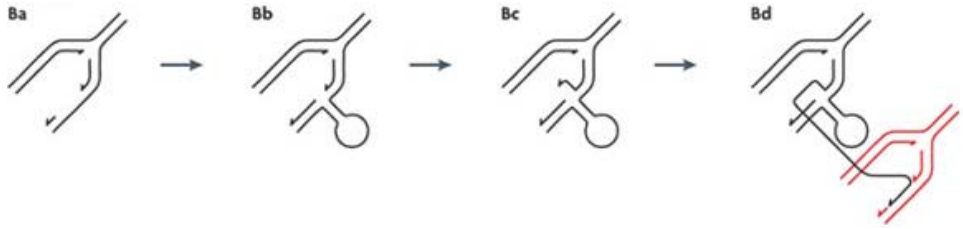

Figure 18 : Fork stalling and template switching (FoSTeS). An exposed single-stranded lagging strand template (Ba) might acquire a secondary structure (Bb), which can block the progress of the replication fork. The 3' end then becomes free from its template (Bc), and might then alight on another exposed single-stranded template sequence on another replication fork that shares microhomology (Bd), thus causing duplication, deletion, inversion or translocation, depending on the relative position of the other replication fork.

(According to P. J. Hastings *et al.*; 2009)

c. Microhomology-mediated break induced replication (MMBIR)

This model is based on the mechanism of BIR repair. It was proposed as an alternative model to FoSTeS, since the latter does not propose mechanistic molecular details and does not involve DNA double strand ends (Hastings, Lupski et al. 2009). MMBIR occurs when the replication fork collapses due to the presence of a nick on a template strand. When a replication fork collapses due to a single break, rad51 is down regulated, preventing the invasion of the broken end and thus the classical BIR repair cannot occur. Instead, when the 5' end of the broken strand is recessed from the break to expose a 3' tail, the latter in the absence of Rad51, will anneal to any exposed end that shares microhomology in another replication fork **(Figure 19)**. DNA synthesis is then initiated with a low processivity resulting in multiple fork collapses and template switching (until a fully replication is established at the

end of a chromosome) leading to complex rearrangements (Hastings, Ira et al. 2009). Depending on the position of the other replication fork, the resulting product can be accompanied by a deletion or by duplication, if annealing occurs in front or behind the position of the fork collapse respectively. Translocation occurs if microhomology is found in a different chromosome.

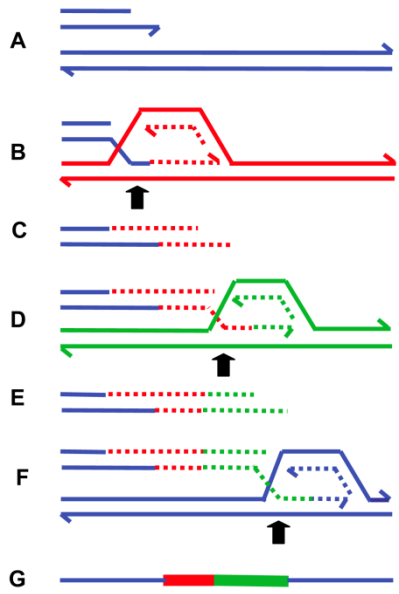

Figure 19 : MMBIR. The figure shows successive switches to different genomic positions (distinguished by color) forming microhomology junctions (arrows). (A) shows the broken arm of a collapsed replication fork, which forms a new low-processivity fork as shown at (B). The extended end dissociates repeatedly ((C and E) shown with 5′-ends resected) and reforms the fork on different templates (D and F). In (F), the switch returns to the original sister chromatid (blue), forming a processive replication fork that completes replication. (G) shows the final product containing sequence from different genomic regions. Each line represents a DNA nucleotide chain (strand). Polarity is indicated by half arrows on 3′ end. Whether the return to the sister chromatid occurs in front of or behind the position of the original collapse determines whether there is a deletion or duplication.

(According to P. J. Hastings *et al.*; 2009)

3-Hotspots of homologous recombination

Homologous recombination between alleles or non-allelic paralogous sequences does not occur only by shared sequence identity among substrates, but is concentrated in “hotspots”; i.e., regions of the genome where recombinations occur with a high frequency (Kauppi, Jeffreys et al. 2004). Jeffreys *et al.* indicated that the distribution of human recombination hotspots is very similar to the meiotic DSB in yeast (Baudat and Nicolas 1997) implying that

hotspots indicate where recombination is initiated by DSB. Since the latter may lead to allelic homologous recombination (AHR) or to NAHR, this suggests that both allelic and non-allelic recombination probably takes place at the same hotspots throughout the genome (Shaw and Lupski 2004).

AHR and NHAR hotspots reveal many features in common: they cluster in small regions of 1-2kb, they are located in different genomic environments, they share no obvious sequence similarity with one another and they lie at or near genes (Kauppi, Jeffreys et al. 2004).

DNA structures shown to induce DSB have been associated with the hotspots, which makes the region apt to recombination. Many studies pointed out the presence of AT-rich palindromes near several hotspots increasing the frequency of spontaneous DSBs (Bi, Park et al. 2003; Kurahashi, Shaikh et al. 2003). Other sequence features expose the DNA to an increased frequency of DSBs, such as triplet repeats which are able to form hairpin structures as seen in some syndromes (de Graaff, Rouillard et al. 1995; Meservy, Sargent et al. 2003). Repeated elements and minisatellite-like sequences have also been identified near hotspots in some genomic disorders (peripheral neuropathies and neurofibromatosis type I) indicating that their presence may increase the DSBs (Reiter, Murakami et al. 1996; Lopez-Correa, Dorschner et al. 2001). Finally, epigenetic modifications may also be implicated. Indeed, tri-methylation of histone H3 on lysine 4, H3K4me3 is an important mark of open chromatin and thus of recombination in yeast and mice (Buard, Barthes et al. 2009; Smagulova, Gregoretti et al. 2011). An open chromatin structure exposes DNA to DSBs or other damage that is then repaired in an aberrant fashion, yielding rearrangements (Shaw and Lupski 2004).

Using a genome-wide map of recombination hotspots estimated from genetic variation data, Myers *et al.* identified hotspot motifs that, in certain sequence contexts, play a role in hotspot activity. First, they found that in humans, recombination hotspots preferentially occur near genes outside the transcribed domain and that there are differences in the frequency of certain sequence features between hotspots and coldspots (i.e. regions that showed no evidence for being a hotspot) (Myers, Bottolo et al. 2005). Indeed the presence of long terminal repeats like THE1A and THE1B, as well as the presence of CT- rich and GA-rich repeats, are over represented in hotspots. Contrary to what has been shown previously, TA-rich, GC-rich and some LINE elements are underrepresented in hotspots (Myers, Bottolo et al. 2005).

Second, when comparing the presence of THE1A/B in hotspots with their presence in coldspots, several sequence differences are observed. The main difference is the presence of the consensus CCTCCCT, which is more frequent in THE1A/B hotspots than elsewhere in the

genome. This motif showed the greatest enrichment in hotspots. Actually, when CCTCCCT is located within THE1A/B elements, it results in a hotspot more frequently than if it is located outside of the repeats. Another motif, the 9mer CCCCACCCC, was found to promote hotspot recombination outside repeat elements. These motifs do not match with any sequences previously linked to recombination activity. Direct evidence for a role of these motifs in hotspot activity came from studies of polymorphic hotspots, where single nucleotide variation reduced recombination at hotspots (Myers, Freeman et al. 2008). An extended family of motifs was later identified by Myers *et al.* based around the 13-mer CCNCCNTNNCCNC, which was described as a driver of genome instability since it clustered in breakpoint regions of NAHR (Myers, Freeman et al. 2008). The 13-mer sequence motif was found to be enriched in human hotspots representing 41% of identified hotspots (Myers, Freeman et al. 2008). Once again, the genetic background influences the penetrance of the motif, since a hotspot is detected 73% of the time in the presence of THE1A, while a hotspot is detected only 10% of the time in the presence of unique DNA.

Recently, it has been shown that tremendous variation exists in the placement and intensity of crossing overs among humans (Coop, Wen et al. 2008), among mice strains and between human and primates (Ptak, Hinds et al. 2005; Segurel, Leffler et al. 2011). The identification of the role of the PR domain containing 9 (PRDM9), a DNA binding protein, led to an understanding of how breaks and hotspots are specified (Baudat, Buard et al. 2010; Parvanov, Petkov et al. 2010). This gene is expressed only in testis and ovaries. It contains a SET domain that tri-methylates H3K4 and a zinc finger domain able to bind DNA (Segurel, Leffler et al. 2011). The knock-out of *Prdm9* in mice showed an arrest of gametogenesis at the pachythene stage for both sexes due to impairment of DSB repair pathways (Hayashi, Yoshida et al. 2005). The knock-out model also showed that the PRDM9 protein is not required for the initiation of genetic recombination, since in the absence of the protein the vast majority of DSB hotspots still coincides with H3K4me3 marks (Brick, Smagulova et al. 2012). However, locations of hotspots in *Prdm9* null mice do not overlap with the hotspots detected in the wild type littermate. Indeed, in the absence of PRDM9, most recombination is initiated at promoters and at other sites of PRDM9-independent H3K4 trimethylation (Brick, Smagulova et al. 2012). Brick *et al.* showed that practically all recombination hotspots are PRDM9 dependent (Brick, Smagulova et al. 2012). In fact, PRDM9 determines the preferred recombination sites through sequence-specific binding of its multi-Zn-finger domain. Many

variants of the protein exist (20 in human) that bind to different motifs (Berg, Neumann et al. 2010).

In men, the PRDM9 A variant was associated with the 13-mer motif, whereas the PRDM9 C variant was found to recognize a 17-mer motif "CCCCaGTGAGCGTtgCc", enriched in hotspots that tend to be used specifically in the African population. Similary in mice, direct binding of PRDM9 to a consensus motif overrepresented in hotspots has also been reported (Segurel, Leffler et al. 2011). The diversity of the identified hotspots (25 000 in human) (Myers, Bottolo et al. 2005) is due to the variation in PRDM9 zinc fingers, which are the only diverse domains in the protein. Indeed, the identity of amino acids present at three positions on each zinc finger plays a major role in determining the DNA sequences to which the protein binds. Remarkably, in the case of PRDM9, these three amino acids show a high polymorphism rate between subspecies (in mice) and individuals (in humans), while the rest of the protein sequence is highly conserved (Baudat, Buard et al. 2010). The common PRDM9 variant found in human is the protein with the zing finger that recognizes the 13-mer motif associated with hotspots. The high rate of changes in the zinc finger residues in contact with DNA might be to avoid the self-destruction of hotspots. In fact, the DNA molecule that undergoes a double-strand break is replaced with DNA from the homologous chromosome. In heterozygous individuals, the allele that increases the frequency of double-strand breaks will be systematically replaced by the less favorable allele in the formation of double-strand breaks (colder allele). The appearance of new PRDM9 alleles can potentially resolve this paradox: the new PRDM9 isoforms will recognize new sites, creating a new set of crossovers hotspots that will replace those recognized by the original isoform (Segurel, Leffler et al. 2011). Consistent with this model, recombination hotspots are not conserved between human and chimpanzee, despite the large similarity between their genome sequences (less than 1.2% divergence).

Chapter V

Human male infertility

Infertility, defined as the inability to conceive after one year of unprotected intercourse (Krausz 2011), is a healthcare problem that has an worldwide impact. Based on a recent report, the prevalence of infertility was estimated to be 9% of couples (Boivin, Bunting et al. 2007) where male factors are involved in at least half of these cases. Infertility in men is initially diagnosed by semen analysis that may reveal several sperm abnormalities which can be due to the male partner not producing any spermatozoa at all (azoospermia) or in insufficient numbers (oligozoospermia), without adequate motility (asthenozoospermia), without normal morphology (teratozoospermia) or a combination of these defects. Thus, the clinical features of male infertility vary from azoospermia to oligoasthenoteratozoospermia. Teratozoospermia can either be mild, moderate or severe and concerns different parts of the spermatozoa. Abnormality can affect the head, resulting in different phenotypes such as round head lacking the acrosome (globozoospermia), enlarged head (macrocephaly), pinhead and double headed spermatozoa. Sperm may also exist with abnormal middle-piece or flagellum. Teratozoospermia is classified into monomorphic (the sperm shows only one of these abnormalities) and polymorphic teratozoospermia (association of more than one abnormality). On the other hand, azoospermia is classified into two main groups: obstructive azoospermia (OA) and non-obstructive azoospermia (NOA). In the OA, spermatozoa are produced normally but they are not ejaculated due to an obstruction of the reproductive tracts, while in the NOA, spermatozoa are not produced. The NOA is characterized by two main phenotypes: Sertoli cell only syndrome (SCO) defined by a total lack of germ cells and maturation arrest due a blockage of spermatogenesis at a certain stage.

1- Causes of male infertility

Causes affecting male infertility are numerous and can be grouped into a number of major categories.

A/ Environment and lifestyle

A general decline in fertility may result from environmental factors. Exposure to high concentrations of environmental endocrine disruptors, such as environmental oestrogen-like molecules may disrupt spermatogenesis and result in reduced sperm volume and sperm count (Sharpe and Skakkebaek 1993; Skakkebaek, Rajpert-De Meyts et al. 2001; Medicine 2008).
Pesticides and some chemicals also interfere with spermatogenesis, causing degeneration in sperm production and a reduction of Leydig cells (Balabanic, Rupnik et al. 2011). There is a wealth of data showing that male animals exposed in utero or perinatally to exogenous oestrogens (diethylstilboestrol, ethinyl oestradiol, bisphenol A) and anti-androgens [flutamide, vinclozolin, 1,1-dichloro-2,2-bis(*p*-chlorophenyl) ethylene (DDE), 1,1,1-trichloro-2,2-bis(4-chlorophenyl)ethane (DDT)] develop undescended testis, low sperm counts or, in the worst case, teratomas and Leydig cell tumours (Skakkebaek, Rajpert-De Meyts et al. 2001). Exposure to heavy metals (e.g. copper) has detrimental effects on male infertility and can be linked to oligoteratoasthenozoospermia. Moreover, lifestyle factors, primarily smoking and consumption of alcohol, are associated with oligoteratozoospermia and can affect sperm DNA integrity (Vine, Margolin et al. 1994; Zenzes, Puy et al. 1999). Finally, it has also been shown that nutrition, some medicinal plants and some drugs may impair male fertility in a reversible or irreversible way (Kamal, Gupta et al. 2003). For example, the powdered seeds of *Abrus precatorius* (Linn) when taken orally inhibit conception in humans. Dose dependent reduction in testicular weight, sperm count and degeneration in later stages of spermatogenesis were found in the testis of rats treated with a steroidal fraction of *A. precatorius* seeds (Kamal, Gupta et al. 2003). Kasturi *et al.* have reported antiandrogenic properties of *A.indica* leaves in male rats. The ethanol extract of neem bark and flowers induced reversible infertility in male rats. The extract interfered with spermiogenesis at Stage XII of late spermatids (Kasturi, Manivannan et al. 1995).
The effects of environmental and lifestyle factors on male infertility are becoming more important given the increased trends in smoking, sedentation and alcohol consumption.

B/ Hypothalamic-Pituitary-Gonadal axis dysfunction

The Hypothalamic-Pituitary-Gonadal (HPG) axis controls spermatogenesis. Regulation of spermatogenesis involves the secretion of Gonadotropin Releasing Hormone (GnRH) by the hypothalamus. GnRH then stimulates the secretion of both follicle stimulating hormone (FSH) and luteinizing hormone (LH) both released by the anterior pituitary (de Kretser, Loveland et al. 1998). The latter acting on Leydig cells, will lead to the production of testosterone that will act on germ cells **(Figure 20)**. Since the germ cells do not possess receptors for FSH and testosterone, the hormonal signals are transduced through the Sertoli cells and peritubular cells by the production of androgen-binding protein (ABP), which will allow the internalization of hormones by spermatogonia (Sofikitis, Giotitsas et al. 2008). Sertoli cells will also produce activine that stimulates the production of FSH. Spermatogenesis is also regulated by a retro negative control. Indeed, testosterone directly inhibits the production of hormones by the hypothalamus and the pituitary gland. It can also stimulate the production of inhibin by sertoli cells, leading to a blockage of FSH production **(Figure 20)**.

However, any disturbance in hormonal secretion leads to infertility. For example, elevated concentration of LH and FSH accompanied by low concentrations of testosterone cause hypogonadotropic hypogonadism, resulting in deficient spermatogenesis in the testis (Seminara, Oliveira et al. 2000). Moreover, mutations of the FSH receptor have been associated with variably severe reduction in sperm count (Sofikitis, Giotitsas et al. 2008). Finally, hormonal deregulation may be a cause or a consequence of the absence of testicular cells.

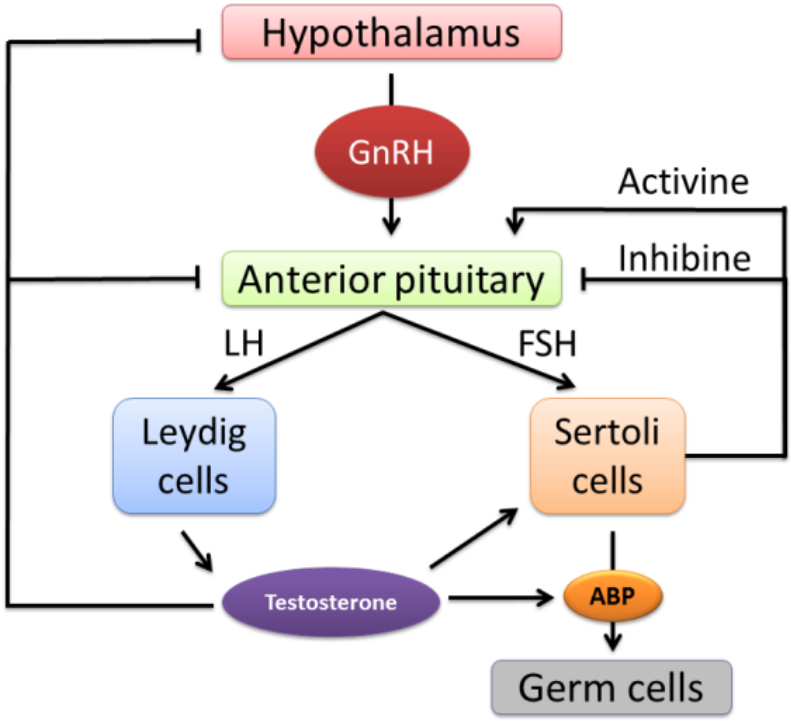

Figure 20 : Hormonal regulation of spermatogenesis. The anterior pituitary secretes LH and FSH in the general circulation. LH stimulates the Leydig cells whereas the FSH stimulates the Sertoli and Germline cells. In response to LH stimuli, Leydig cells secrete testosterone that will affect Sertoli and Germline cells, as well as create a negatif feedback loop affecting the hypothalamus and anterior pituitary. Im function of the state of activation of the Sertoli cells, they cab secrete activine or inhibine that can either activate or repress respectively the anterior pituitary.

C/ Sexual disorders

Erectile dysfunction and ejaculatory failure are the most common examples of male sexual disorders. Erectile dysfunction affects 10-15% of all males and is induced principally in 80% of cases by medications, blood flow abnormalities or hormonal abnormalities (Poongothai, Gopenath et al. 2009). Furthermore, it has been shown that smoking also causes erectile dysfunction by inhibiting steroidgenesis in mouse Leydig cells, resulting in reduction of testosterone production (Patterson, Stringham et al. 1990; Feldman, Goldstein et al. 1994). Psychological factors account for the remaining cases and may be attributed to stress, performance anxiety or misinformation about sexuality. Concerning ejaculation disorders, a spectrum exists, ranging from mild premature to severely retarded or absent ejaculation (Singh and Jaiswal 2011). The latter can be caused by an obstruction of reproductive tracts. Another aspect of ejaculation disorder is the retrograde ejaculation observed in diabetic patients. It occurs when semen, which would normally be ejaculated via the urethra, is

redirected to the urinary bladder due to a failure of a closure of the bladder neck (internal urethral sphincter) (Jefferys, Siassakos et al. 2012).

D/ Chromosomal aberrations

It has been shown that genetic contributions in infertility are increasing. Indeed, genetic infertility can be associated with chromosomal aberrations, with pathologic syndromes and with monogenic autosomal recessive mode of transmission.
Chromosomal abnormalities interfere with spermatogenesis and account for 5% of oligozoospermia, to 15% of non-obstructive azoospermia (Ferlin, Raicu et al. 2007). They include numerical (aneuploidy) and structural aberrations and involve sex chromosomes as well as autosomes.

a. Aneuploidies

For instance, Klinfelter syndrome is the most widely described abnormality found in NOA (Huynh, Mollard et al. 2002), with a prevalence of 1/500 and accounts for 14% of cases of azoospermia (Singh and Jaiswal 2011). It results from an incorrect sex chromosome number in infertile men, due to a non-disjunction of sex chromosomes during maternal (60%) or paternal meiosis (40%) (Thielemans, Spiessens et al. 1998).

In 90% of cases, the syndrome is characterized by a 47,XXY karyotype where men show a maturation arrest of spermatogenesis at the primary spermatocyte stage or even a total absence of spermatogenesis due to a hyalinization of seminiferous tubules (Huynh, Mollard et al. 2002). However, mosaicisms such as 47,XXY /46,XY, have been reported in 10% of cases, where the affected patients show a less severe phenotype manifested by an oligozoospermia (Huynh, Mollard et al. 2002). Finally, in some rare cases multiple copies of X chromosomes may be present.

b. Structural aberrations

Structural abnormalities are frequently observed in infertile men and include translocations, inversions and Y micro-deletions.

- Translocations

Chromosomal translocations are defined as rearrangements between non-homologous chromosomes. It can be balanced without any consequences or unbalanced (loss or gain of genetic material), causing a corruption in the genetic message. Autosomal translocations were found in a higher rate in infertile males showing a variety of sperm phenotypes from oligozoospermia to azoospermia (Elliott and Cooke 1997). Two types of translocations have been described: reciprocal and Robertsonian translocations.

Reciprocal translocations take place when exchanges between two non-homologous chromosomes occur during meiosis. They are found from four to ten times more often in infertile men compared to the general population (Elliott and Cooke 1997). Several sperm karyotyping studies performed on human carriers of balanced reciprocal translocations have shown a large range in frequency of unbalanced sperm compared to normal or balanced sperm (18.6–80.7%) (Morel, Douet-Guilbert et al. 2004; Benet, Oliver-Bonet et al. 2005). These frequencies fluctuations showed that the production of unbalanced spermatozoa is determined by the characteristics of the chromosomes involved and the breakpoint positions.

However, Robertsonian translocations involve the fusion of two acrocentric chromosomes near the centromere region with loss of the short arms resulting in a karyotype of 45 chromosomes. The phenotype of Robertsonian translocations varies from a normal spermatogenesis to the inability to produce spermatogonia and has been found in oligozoospermic and azoospermic men with rates of 1.6% and 0.09% respectively (O'Flynn O'Brien, Varghese et al. 2010). Carriers of Robertsonian translocations can also lead to the formation of balanced/unbalanced gametes depending on chromosome segregation.

Both translocations can be balanced and lead to the formation of normal gametes carrying the translocation if the two rearranged chromosomes segregate during anaphase I to the same cell pole. However, if homologs migrate in the same direction, this will result in gametes with unbalanced combinations leading to reproductive failures, spontaneous abortions or children with abnormalities (Morel, Douet-Guilbert et al. 2004).

- Inversions

Inversions happen when two breaks occur in the same chromosome, followed by a rearrangement in an inverted order. If the breakpoints are located outside coding sequences, the reorientation of a chromosomal region in the opposite direction does not alter gene

functions. However, problems occur during pairing of homologous chromosomes at meiosis, which may lead to offspring with duplications or deficiencies if crossover occurs in the inverted region of the paired chromosomes. There are two types of inversions: pericentric (chromosome breaks occur in both chromosome arms and include the centromere in the inversion) and paracentric (both break points are in one chromosome arm). Inversions have been associated with male infertility involving severe oligoasthenoteratozoospermia and azoospermia 13 times more often than in the population (Krausz and Forti 2000).

- Y microdeletions

The human Y chromosome that comprises 60 MB with a short arm (Yp) and a long arm (Yq) is an area of interest in the field of male infertility. It contains the sex determining region of the Y chromosome (SRY), located on the short p arm and several spermatogenesis- related genes located on the long q arm. The Y chromosome is composed of repetitive sequences including direct repeats, inverted repeats and palindromes and contains pseudoautosomal, euchromatic and heterochromatic regions. Sequence analysis revealed 78 protein-coding genes located on the Y chromosome within palindromes that comprise the two arms with a high nucleotide identity (Skaletsky, Kuroda-Kawaguchi et al. 2003). Mutations affecting genes on the Y chromosome are deleterious, since the chromosome has no homologous partner that can be used during repair mechanisms. Therefore, spermatogenesis genes contained within palindromes are present in multiple copies. When a mutation occurs in a gene, correction can be performed using the non-mutated copy.

The pseudoautosomal regions (PARs) are located at both ends and allow pairing with the X chromosome during meiosis. The euchromatic region occupies more than 95% of the whole Y chromosome and is called male specific Y (MSY). A particular region of interest on the long arm is the Azoopermia factor region (AZF), which is divided into three subregions: AZFa or proximal, AZFb or central and AZFc or distal, and contains specific genes implicated in spermatogenesis and male gonad development **(Figure 21)**. Microdeletions of these regions are associated with male infertility, causing a failure of spermatogenesis, and are found in azoospermic and oligozoospermic men with normal karyotype. Y microdeletions, caused by intrachromosomal recombination between large homologous sequence blocks, represent the most frequent genetic cause in infertile men. In fact, 13% of azoospermic men and 1-7% of severe oligozoospermic men (< 5 million spermatozoa per ml semen) showed Y

microdeletion.

These deletions are detected by polymerase chain reaction (PCR) methods, since they are small in size and cannot be diagnosed cytogenetically. Deletions of AZFa remove the two main genes of this region (USPY9 and DBY) and cause Sertoli cells only syndrome. Deletion of the AZFb region leads to a maturation arrest of spermatogenesis at the primary spermatocyte stage, caused by the absence of the RBMY gene. Finally, deletion of the main gene DAZ at the AZFc locus is associated with a wide range of phenotypes from oligozoospermia to azoospermia. Recently, a fourth type of micro-deletion detected in 8% of azoospermic men was found that removes the Yq sequences, resulting in a isodicentric Y chromosome with a duplication of Yp and centromeric sequences (Heard and Turner 2011). The isodicentric Y chromosome can be lost during anaphase as a result of the chromosome breakage that contains two centromeres that form a bipolar spindle attachment.

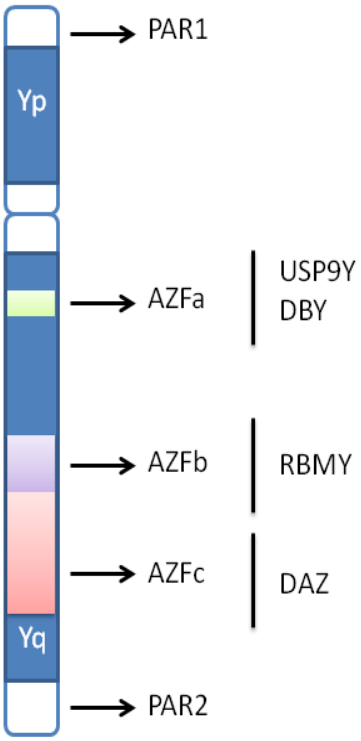

Figure 21 : Schematic representation of Y chromosome

(Adapted fromPoongothai *etal.*)

E/ Infertility associated with pathological syndromes

Infertility can be associated with other complex phenotypes such as sex reversal syndrome (involving genes such as *SOX9*, *SRY*, and *NROB1*), hypogonadotropic-hypogonadism defects (involving genes such as *GNRH*, *KAL*, *PC1*, *GNRHR*, *LEP*, and *PCSK1*), myotonic dystrophy (gene *DMPK*) or a mild form of cystic fibrosis and many others(Matzuk and Lamb 2008). For example, mutations in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene are detected in infertile men exhibiting azoospermia with congenital bilateral absence of vas deferens (CBAVD) (Yu, Chen et al. 2012). Men with CBAVD undergo normal or slightly reduced spermatogenesis and present clinical cystic fibrosis symptoms suggesting that CBAVD is a mild form of this syndrome (Yu, Chen et al. 2012).

2- Monogenic autosomal infertility

Despite the fact that infertility in men can be explained by multiple reasons, a considerable number of cases (25 to 30%) remain idiopathic and at least some of them can be attributed to genetic reasons (O'Flynn O'Brien, Varghese et al. 2010; Krausz 2011). In this section, I will focus on non-syndromic genetic defects affecting human male gametogenesis presenting a monogenic autosomal recessive mode of transmission. First, I will present what are the requirements to perform informative genetics studies in the field of infertility, what are the techniques and then the results obtained so far.

A/ Requirements for the identification of mutated genes in human infertility

The first major requirements when considering identifying genes mutated in patients with infertility problem is the quality of the clinical diagnosis. Indeed, any improper diagnosis will bias the selection of patients to be included in the study (as in all genetics studies) and introduce confusion in the analysis, rendering the genetic analysis difficult or even impossible. It may also limit the number of patients included in the study. Since the bigger the patient cohort is, the easiest the analysis will be.

The second major requirement in the case of autosomal recessive mutations is the ability to study informative patients. Indeed, although geneticists need well diagnosed patients, it is

important to study large families or large cohorts whenever possible. The analysed patients will preferentially be part of i) three or more affected patients as well as two or more unaffected patients from families with some degree of consanguinity, or ii) a larger group of patients coming from a restricted area or living in a limited socio-cultural enclave where the abnormality seems to have a quite high occurrence, which should correspond to a founder effect. Patients studied in the case of autosomal recessive infertility, must present a non-syndromic infertility with a homogeneous abnormality of the spermatogenesis. The non-syndromic character implies that patients are healthy and do not present any other pathology.

B/ Strategies to identify causative genes

Two approaches are commonly used to identify genetic causes of pathologies, and these can also be applied to the genetics of infertility: reverse and forward genetics (Jamsai and O'Bryan 2011).

a. Candidate genes approach

The reverse approach, also known as the candidate gene approach, is initiated by selecting genes from infertile animals, mainly mice models, assuming that the gene function is conserved through evolution. Such an approach needs many well characterized patients and a large amount of time. To date, the vast majority of mouse models of altered gene function have been knockouts. As soon as a convergence of phenotypes is observed between a specific mutated mouse gene and a human phenotype, mutations can be screened in the human orthologue gene (Jamsai and O'Bryan 2011). So far, the discoveries resulting from this approach have been disappointing (Miyamoto, Hasuike et al. 2003; Gianotten, Schimmel et al. 2004; Pirrello, Machev et al. 2005; Stouffs, Lissens et al. 2005; Westerveld, Repping et al. 2005; Westerveld, Repping et al. 2005; Aoki, Christensen et al. 2006). This is probably due to the great genetic heterogeneity and the limited number of patients tested.

b. Linkage analysis

Forward genetics, also known as positional cloning or descent studies; seek to identify the chromosomal localization of the gene mutated in hereditary cases. The locus of a gene can be determined by genotyping polymorphic markers via linkage analysis studies. Linkage analysis

methods rest on the biological phenomenon of recombination of homologous chromosomes during meiosis. The probability of recombination event occurring between loci far apart on a single chromosome is larger than loci closer together. Hence, alleles at loci near each other are generally inherited together, making it possible to follow transmission from generation to generation by using polymorphic markers distributed over all human chromosomes at known chromosomal positions. This approach benefits from the recent and ongoing evolutions in molecular biology. Indeed, new powerful tools are continuously being developed and have reached the scale of whole genome analysis, facilitating the identification of recessive autosomal mutations. The main approach, homozygosity mapping is based on genotyping studies and aims to investigate, for a single patient, thousands of Single Nucleotide Polymorphisms (SNPs) simultaneously on a microarray. The success of homozygosity mapping studies strictly depends on the availability of patients and their rigorous selection. This approach allows the identification of homozygous regions shared by affected siblings of the same family and absent in non-affected individuals including the parents (Bier, von Nickisch-Rosenegk et al. 2008; Grant and Hakonarson 2008) .

Genotypes can be organized according to the chromosomal location of each SNP and haplotypes of several individuals can be compared. In order to easily visualize and compare homozygous regions between different families, software called homoSNP has been developed at the IGBMC. This software includes a graphical interface for viewing areas for which a defined number of consecutive SNPs are homozygous. The homozygous regions are represented in purple, light blue and blue depending on the number of consecutive homozygous SNPs. For a 10K chip, these colors correspond respectively to a number of SNPs [15-20 [[20-25 [[25- ... [.

The candidate regions, selected on the basis of consecutive homozygous SNPs, are then screened using a genome browser in order to list the candidate genes (**Figure 22**). The genes are ranked according to their expression, which in this case must be predominant in the testes, and their potential function. Once genes are chosen, PCR and sequencing of all exons and exon/intron junctions are performed in order to highlight the mutation involved in the patient infertility.

SNP microarrays have tremendous advantages over previous genotyping methods as they are very dense and cover the whole genome (Bier, von Nickisch-Rosenegk et al. 2008; Grant and Hakonarson 2008). It is a fast technique that requires only a small amount of genomic DNA (250ng). However, genomic regions generated by homozygosity mapping are generally

multiple megabases in size and can contain multiple genes. Therefore, the identification of the mutated gene from a large number of candidates is expensive and time consuming. This is where the number of patients or the size of the family is of great importance. Indeed, the higher the number of siblings in the family or the larger the patient cohort, the smaller the homozygous region should be, limiting the number of candidate genes to be screened.

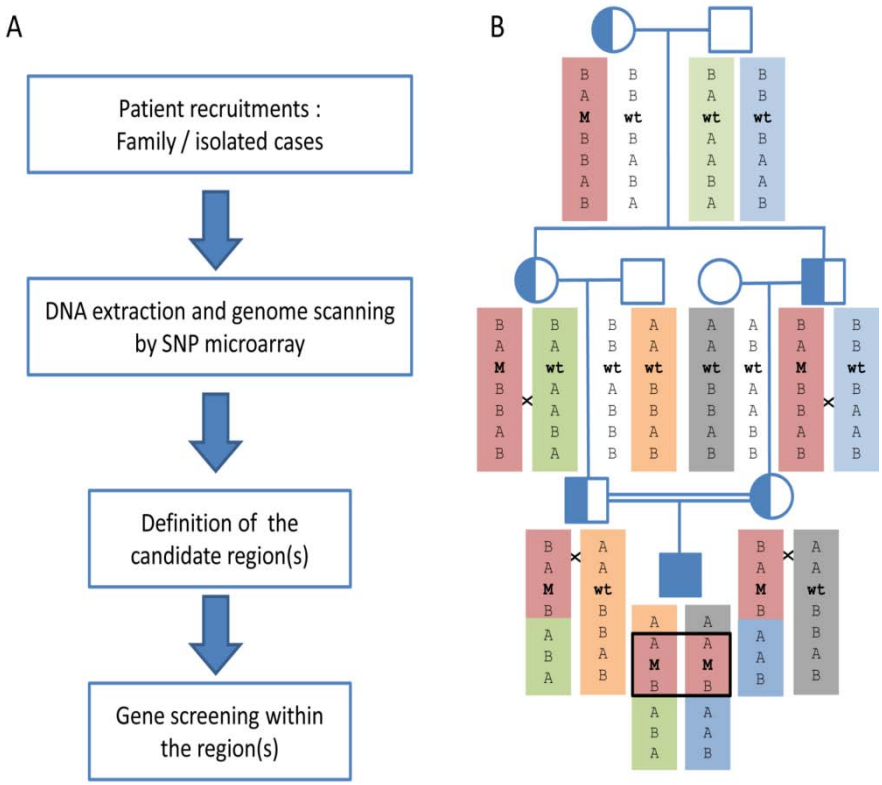

Figure 22 : Linkage analysis. (A) Positional cloning procedure. The positional cloning procedure used to identify candidate genes. A first critical step is the recruitment of the patients. Two groups of patients are of interest: 1) patients presenting a familial infertility, where the families contain at least two infertile members, 2) patients from a small isolated geographic area or tribes, where the abnormality has a relatively high occurrence. In both cases, patients should present a non syndromic infertility. A whole genome scan by SNP microarray is performed in order to select a locus shared only by the patients and absent in non-affected ones. Once the region is defined, a bioinformatics screen is performed to select candidate genes within the region. (B) Principle of homozygosity mapping A patient with a rare genetic disorder resulting from consanguineous parents inherited a mutation (M) on both chromosomes, which was present at the heterozygous state in a common ancestor (great-grandmother in this diagram). The polymorphic markers located in the chromosomal region around the gene are represented by different colors and letters depending on the alleles. Because of various events of meiotic recombination (represented by crosses), the alleles of markers far from the gene change and are therefore heterozygous in the patient, while markers close to the mutated gene are transmitted in a block (called haplotype) from the great-grandmother and are therefore in the homozygous state in the patient. They define a region of homozygosity by descent (framed region).

- Principle of the technique

SNP microarray genotypes more than 10,000 human SNPs on a single array, using a single polymerase chain reaction (PCR) primer. As its name implies, it is a mapping tool designed to identify regions of the genome that are linked to, or associated with, a particular trait or phenotype. The Affymetrix experimental protocol follows the same steps for the four categories of chips (10K, 50K, 250K, 6.0G). Genotyping chips are based on the principle of Allele-specific hybridization (ASH). ASH is a method of allele discrimination in genotyping (Wang, Fan et al. 1998). By synthesizing probes on the array corresponding to both of the two possible alleles at each SNP and hybridizing the target to the array, a SNP can be determined as AA, AB, or BB by analysing the resulting signals from the allele-specific probes. 25-mer probes are synthesized, corresponding to a perfect match of the A allele sequence (PMA) and to a perfect match of the B allele sequence (PMB). To determine binding specificity, a single base pair mismatch is included at the centre position of each 25-mer, for each allele (MMA and MMB). For each SNP, 40 different 25 bp oligonucleotides are tiled, each with a slight variation in perfect matches, mismatches, and flanking sequence around the SNP.

Genomic DNA is digested with one of several restriction enzymes and ligated to adaptors recognizing the cohesive four base overhangs. All fragments resulting from restriction enzyme digestion, regardless of size, are substrates for adaptor ligation **(Figure 23)**. A generic primer, which recognizes the adaptor sequence, is used to amplify ligated DNA fragments and PCR conditions are optimized to preferentially amplify fragments in the 250-1000 bp size range. The amplified DNA is labeled by a biotinylated anti-streptavidin antibody and hybridized to GeneChip arrays. Hybridization between two DNA strands forms hydrogen bonds between complementary nucleotide base pairs. A large number of complementary base pairs in a nucleotide sequence means tighter non-covalent bonding between the two strands. The arrays are washed and stained on a GeneChip fluidics station and scanned on a GeneChip Scanner 3000. After washing off of non-specific bonding sequences, only strongly paired strands will remain hybridized. So fluorescently labeled target sequences that bind to a probe sequence generate a signal that depends on the strength of the hybridization determined by the number of paired bases, the hybridization conditions, and washing after hybridization. The total strength of the signal from a spot, depends on the amount of target sample binding to the probes present on that spot. The determination of genotype is carried out from the hybridization intensity at each allele specific oligonucleotide.

In an individual homozygous for the A allele, only probes with this specific allele will be

hybridized, and vice versa in an individual homozygous for the B allele, while a hybridization will be found on both types of probes in an AB heterozygous individual. The genotype of each SNP is determined by the software provided by Affymetrix (GeneChip DNA Analysis Software, GDAS, or more recently the GeneChip Genotyping Analysis Software, GTYPE).

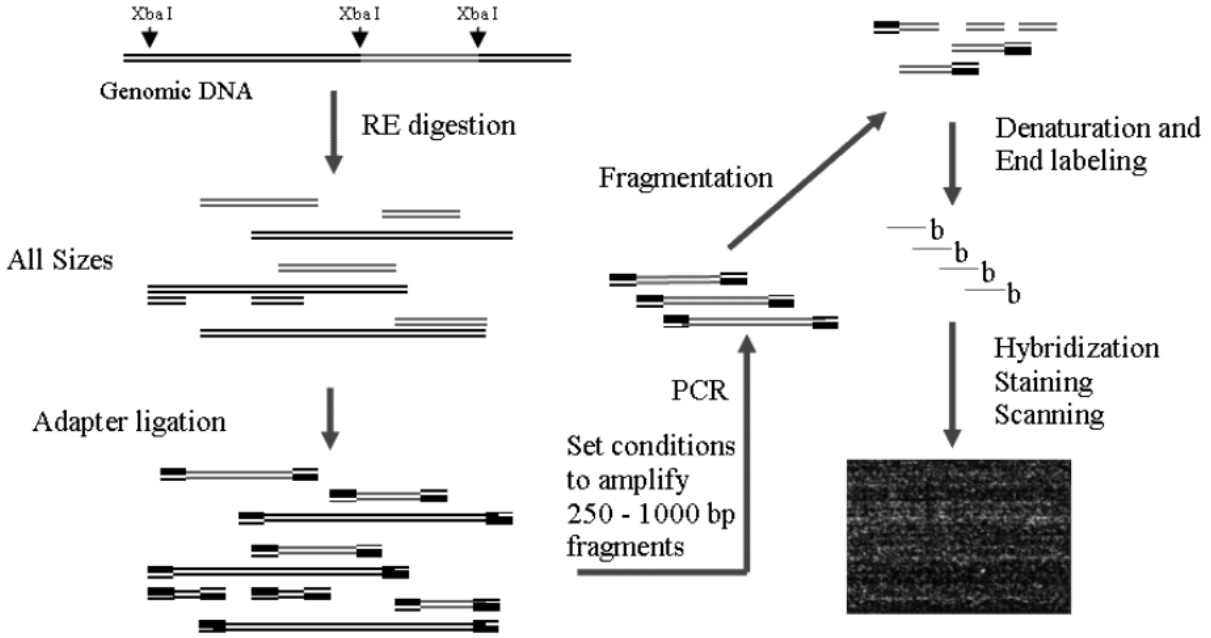

Figure 23 : GeneChip Mapping Assay. The genomic DNA is digested with a restriction enzyme (XbaI in the case of 10K chips). Adapters are ligated to the fragments. PCR with a primer complementary to the adapters preferably amplifies the XbaI fragments of about 250 to 1000 bp that were targeted by the Affymetrix to contain SNPs analyzed. The amplified DNA is fragmented and labeled and hybridized on the chip.

(From GeneChip® Mapping 10K 2.0 Assay Manual, Affymetrix)

c. Exome high throuput sequencing

Recently, a new method has promised to speed up discovery of the genetic causes of disease by allowing the parallel sequencing of millions of sequences at high throughput: next generation sequencing (NGS). NGS combined with enrichment technologies is used to sequence the exome (the set of all coding exons representing 1% of the human genome) in order to discover most of the disease-related variations in exons **(Figure 24)** (Metzker 2010). In other words, this technique allows the identification of protein coding mutations, including missense, non-sense, splice site, and small deletion or insertion mutations.

Exome sequencing has become a powerful and efficient strategy for identifying the genes responsible for Mendelian disorders and complex diseases (Gilissen, Arts et al. 2010; Krawitz, Schweiger et al. 2010; Musunuru, Pirruccello et al. 2010; Ng, Buckingham et al. 2010; Singleton 2011) and may be applicable in the case of infertility.

The challenge in this technique lies in highlighting the particular mutation responsible for the disease, since the genome of a single individual will reveal about 25 000 variants (Singleton 2011). The majority of these variants are known variants in human populations. In fact, sequencing highlights variations that can have different origins. According to the type of variation they can be easily categorized as mutations. Filtering steps are performed giving a priority to frameshifts, stop codons, disruptions of splice sites than to variants that are predicted *in silico* not to affect the protein structure or function. Synonymous variants and variants which do not fit with the mode of disease inheritance and do not segregate with the disease (in families) are also excluded.

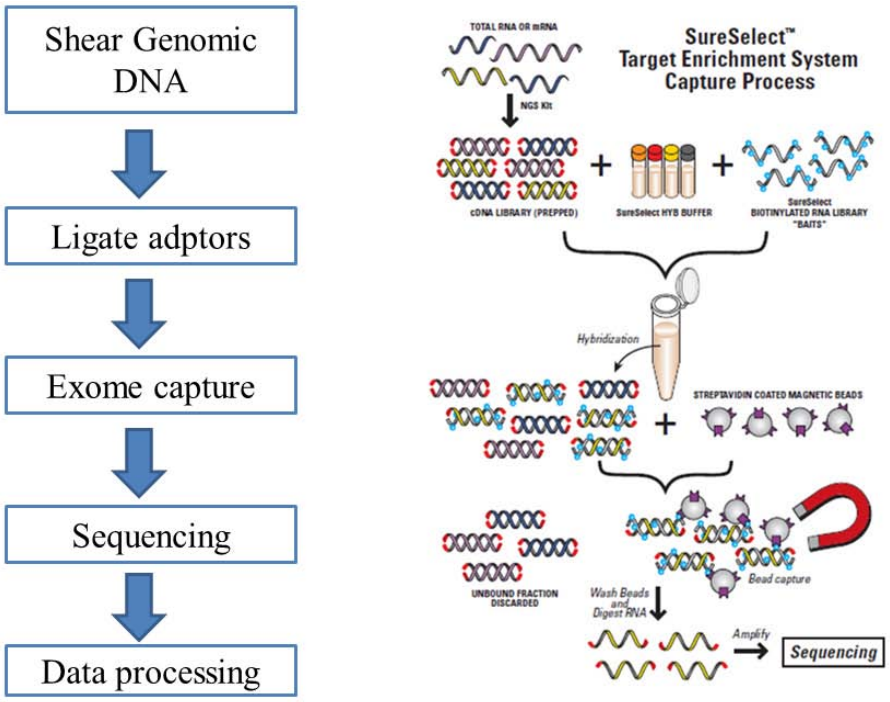

Figure 24 : Exome sequencing. The basic steps required for exome sequencing are five. First, genomic DNA is sheared into random fragments of about 300bp which are then flanked by adapters to allow sequencing. This is followed by exome capture. Sequences that correspond to exons are captured by hybridization to biotinylated DNA or RNA baits and then pulled down by biotin-streptavidin coated magnetic beads. Finally this step is followed by amplification and massive parallel sequencing. Barcodes to allow sample indexing can potentially be introduced before capturing or during post-capture amplification.

Once a variation is maintained and selected as responsible of the disease rigorous controls are needed. For this, different possibilities are now available. One classical method is to check whether the variation/mutation is found in at least 200 sequenced healthy controls matching the same geographic region and/or same ethnicity of patients. For this purpose, controls must have proven fertility or normal spermatogenesis or, in other words, they should have conceived without any medical assistance. It is now also possible to check databases that collect human sequences and repertory polymorphisms among the human population (dbSNP (Sayers, Barrett et al. 2011), the HapMap project (Consortium 2003), and 1000 genomes project (Clarke, Zheng-Bradley et al.)). This powerful strategy reduces the number of candidate genes but sometimes it can be problematic since dbSNP might be contaminated with small number of pathogenic alleles. More recently, software systems have been developed that are able to predict whether alterations will or will not have functional consequences on the produced protein (SIFT and PolyPhen v2 (Kumar, Henikoff et al. 2009; Adzhubei, Schmidt et al. 2010)). Whatever the method used, the best way to prove the effect of a mutation on a protein is to set up a biological functional test. In the case of genetics of infertility, this is very challenging because genes involved in male infertility are often mainly, if not exclusively, expressed by the testes from which biopsies are not easily collected. There is also a dramatic lack of cellular models, for example, carrying on a meiosis that can be explored in vitro.

The inheritance mode of monogenic disorder affects the study in many ways. In fact, filtering is more efficient for recessive diseases than for dominant, because the genome of any given individual has around 50-fold fewer genes with two, rather than one, novel protein-altering alleles per gene (Bamshad, Ng et al. 2011). In consanguineous studies in which a recessive mode of transmission is suspected, the most efficient strategy is to sequence a pair of affected individuals. Contrariwise, to identify *de novo* coding mutations or sporadic mutations sequencing of parents-child trios is highly effective (Hoischen, van Bon et al. 2010). This approach is also efficient when a dominant transmission is suspected. In the case of male infertility, autosomal dominant mutations can be maternal transmitted without showing any phenotype in females. However, any record in this case has been listed so far.

A combination of homozygosity mapping with genome sequencing may lead to rapid identification of the causative mutation for a disease (Nielsen, Paul et al. 2011). For example, in a recessive mode of inheritance, instead of sequencing two affected members, sequencing

can be limited to one person who shows the smallest region of homozygosity shared between the affected individuals.

Although, exome sequencing has proven its worth in genetics, this technique has some limitations related to the coverage of the entire exons, which will never reach 100% (Bloch-Zupan, Jamet et al. 2011), the poor capture and sequencing of a causative gene, the mismapped reads or alignment errors that can generate false variants, the difficulty in detecting genomic rearrangements such as large deletions and duplications and finally the causative gene is not in the target definition because is not yet reffered in the refseq database. Analytical failure mainly concerns the false positive and flase negative calls. False positive calls result in detection of candidate genes that cannot be logically eliminated by filtering and are observed in segmental duplications and pseudogenes. False negative calls result from the presence of pathogenic variants in the data used as a reference to eliminate the known variants. Finally the power of this technique is reduced in the case of genetic heterogeneity where no single gene is found to be responsible of all cases of a disease making impossible to distinguish between causative and non-causative variants.

The technical locks of whole genome genotyping and sequencing have now been broken and we believe that the key to successfully identifying causative mutations by descent studies depends on the adequate recruitment of patients. The major limitation of NGS is the price. Indeed, the technology remains expensive and the analysis requires a multi-disciplinary approach, but the cost is falling very rapidly and soon it will be the reference technology for investigating genetic mutations and the field will be moving from exome to whole genome sequencing making sense for non-coding variation.

B/ Outcome of this quest for genes responsible of infertility

Recent identification of mutations in well-known sperm defects provides proof of their genetic origins. Indeed, gene mutations have been identified for two types of teratozoospermia, namely globozoospermia and macrocephaly, for asthenozoospermia and for NOA.

The first example of a non syndromic male infertility in human caused by an autosomal gene defect was a mutation in *SPATA16* (spermatogenesis associated 16). This gene was identified by homozygosity mapping while studying a consanguineous Ashkenazi Jew family where 3 brothers were globozoospermic (Dam, Koscinski et al. 2007).

SPATA16 is a testes expressed gene located on 3q26.32. It contains 11 exons encoding for a 569 amino acid protein, which is highly conserved across mammals. The most conserved domain of the protein is a tetratricopeptide repeat (TPR) domain, a protein-protein interaction domain commonly but not exclusively found in cochaperone proteins (Blatch and Lassle 1999). The affected brothers showed a homozygous variation in exon 4 (c.848G→A) altering the last nucleotide of the exon and leading to an inappropriate splicing. The mutation leads to the deletion of exon 4 and disrupts the TPR domain since 36 amino acids are deleted (Dam, Koscinski et al. 2007).

PICK1 was identified by a gene candidate strategy. Liu et al screened four candidate genes (*GOPC*, *HRB*, *CSNK2A2* and *PICK1*) in 3 globozoospermic type I patients and found a homozygous missense mutation (c.G198A) in exon 13 of the PICK1 gene for one of these patients (Liu, Shi et al. 2010). This homozygous mutation altered the amino acid G393R in the C-terminal part of the protein. PICK1 is a peripheral membrane protein involved in protein trafficking and highly expressed in the brain, testes and pancreas. In the brain, PICK1 has a role in AMPA receptor trafficking and synaptic plasticity (Gardner, Takamiya et al. 2005). Male *Pick1* knockout mice validate these results (Steinberg, Takamiya et al. 2006) and surprisingly show an infertile phenotype. In testes, PICK1 is a Golgi derived proacrosomal granule protein expressed in round spermatids. Male *Pick1* null mice showed a fragmented acrosome due to a fusion failure of acrosomal vesicules in the Golgi phase and the failure of the cap-like acrosome structures during the cap phase (Xiao, Kam et al. 2009). Since the mutation identified by Liu *et al* does not change a highly conserved amino acid and no functional consequences have been shown, caution should be taken not to over interpret this result without additional data.

In an older work using a candidate approach, we unsuccessfully screened 6 globozoospermic patient for mutations in *CSNK2A2* (Pirrello, Machev et al. 2005).

For one other teratozoospermia, macrocephalia or enlarged headed spermatozoa, a gene has been identified. Indeed, the *AURKC* gene was identified by linkage analysis involving a genome wide microsatellite scan of ten infertile men with a large headed spermatozoa phenotype coming from the same geographic area (Dieterich, Soto Rifo et al. 2007). *AURKC* is located on 19q13.3-qter, has 7 exons and is highly expressed in testes. A homozygous deletion in exon 3 (c.144delC) was detected, which introduces a frameshift leading to a premature stop codon and resulting in the alteration of the conserved kinase domain (Dieterich, Soto Rifo et al. 2007). Aurora Kinase C is the last discovered member of an aurora

kinase family, which includes two other members: AURKA and AURKB. All three are cell cycle regulatory serine/threonine kinases that play an essential role in mitotic cell division (Dieterich, Soto Rifo et al. 2007; Kimmins, Crosio et al. 2007). Little is known about AURKC. In mice spermatogenesis, AURKC is involved in chromosome segregation and cytokenesis, which fits well with the human phenotype (Tang, Lin et al. 2006). Indeed, it was shown that all the spermatozoa of these patients are tetraploid, strongly suggesting the implication of AURKC in the segregation of chromosomes and/or meiotic cytokinesis, explaining the large size of the gametes (Dieterich, Zouari et al. 2009).

The founder origin of the (c.144delC) mutation was further confirmed when studying a larger cohort of North African patients. This study also established a frequency of heterozygotes of over 1 in 50, implying an expected prevalence of 1/10000 men. Such a high frequency is surprising since it makes *AURKC* infertility among the most frequent single gene defect in this population. Interestingly, the authors identified two fertile homozygous females, excluding a fundamental role of AURKC in female meiosis and confirming the differences between male and female meiotic mechanisms (Dieterich, Zouari et al. 2009). Recently, the same group described a second mutation affecting the splicing of *AURKC* (Ben Khelifa, Zouari et al. 2011).

Using a combination of both methods (i.e. candidate gene approach and homozygosity mapping), Avearius *et al* studied two consanguineous Iranian families and found insertion mutations in *CATSPER1* responsible for sperm immotility in the three asthenozoospermic patients and segregating as a non-syndromic autosomal recessive pathology (Avenarius, Hildebrand et al. 2009). They initiated their study based on observations made in the mouse model. Indeed, in mice, the deficiency of any one of the four *Catsper* genes leads to male infertility by affecting spermatozoa motility (Jin, Jin et al. 2007; Qi, Moran et al. 2007). They then performed a homozygosity mapping specific to the *CATSPER* locus. In both families, an insertion mutation was observed. In the first one, the mutation (c.539-540insT) introduces a premature stop codon that could eventually lead to the production of a truncated protein of 188 amino acids. In the second one, the mutation (c.948-949insATGGC) also introduces a premature stop codon with a resulting protein of 335 residues. In both cases, even if the protein is produced, it would lack all six transmembrane domains, abolishing the CATSPER1 channel activity (Avenarius, Hildebrand et al. 2009). Indeed, the CATSPER1 protein is required for the entry of Ca^{2+} into the flagellum and for the hyperactivation of the sperm once they enter the female reproductive tract (Ren, Navarro et al. 2001). Hyperactivation is

necessary for the penetration of the spermatozoa into the oocyte (Carlson, Westenbroek et al. 2003). It could be of interest to investigate whether ICSI could be successfully offered to treat these patients.

Recently, starting with a candidate gene strategy, Choi *et al* identified variations in SOHLH1 that could be responsible of non-obstructive azoospermia (NOA). Indeed, by screening 96 NOA Korean patients and 159 controls, they identified 3 variations that could be responsible for NOA (Choi, Jeon et al. 2010). The functional analysis did not clearly establish the pathogenicity for two of these variations. Thus, even if they were not detected in controls, these two variations (c.91T>C and c.529C>A) should be considered with caution. *In vitro* analyses of the third variation showed that it induces a splicing error, resulting in a truncated protein lacking the last 6 amino acids of the bHLH domain. This truncated protein was shown to be less potent in activating the KIT promoter in an *in vitro* test. (Choi, Jeon et al. 2010).

Despite their interest, these results need confirmation. Indeed, the mutation is found at the heterozygous status and therefore should have a dominant negative effect. In addition, the mouse model (Ballow, Meistrich et al. 2006) does not fully phenocopy what is observed in the only patient for whom a testis anapathology could be done. A much larger cohort of controls should be tested. However, the fact that the same mutation has been found in two independent NOA patients reinforces the hypothesis that the c.346-1G>A variation could be considered as a mutation.

Finally, I would like to mention the striking cases of NR5A1, which encodes the steroidogenic factor 1 (SF1). NR5A1 mutations are linked to a wide range of phenotypes, such as male partial and complete gonadal dysgenesis, associated or not with adrenal failure, penoscrotal hypospadias, micropenis and primary ovarian insufficiency (Lin, Philibert et al. 2007; Lourenco, Brauner et al. 2009). In a recent report, Bashamboo *et al* found 7 healthy patients, suffering only from severe oligozoospermia or azoospermia, out of 315 screened patients to be mutated for NR5A1. All the mutations were found in the well conserved hinge region of the protein (Bashamboo, Ferraz-de-Souza et al. 2010). Thus, we have an interesting situation where, depending on the type of mutations, we can observe a large variety of phenotypes, with one of them affecting only spermatogenesis. We can predict that such a situation, which is somewhat reminiscent of what is seen for CFTR mutations, will be encountered with other genes.

C/ Polymorphisms associated with human male infertility

Apart from these conclusive results concerning the identification of mutations in single genes responsible for male human infertility, in the last few years there has been an (too) abundant literature on variations that are presented as mutations or polymorphisms that could be associated with a spermatogenesis alteration or could constitute a risk factor. Most of these studies are based on a candidate gene approach. These studies have been conducted on a limited number of patients and controls. In some cases, the ethnicity of the patients and their controls did not match. Some of the variations have been described as heterozygous and parents or family members have not been checked for the presence or absence of the variation. Finally, none of these variations has been the object of functional studies to validate their deleterious effect. Until such studies are performed, these variations should be considered as polymorphisms. A list of these variations/polymorphisms can be found in Massart *et al* (Massart, Lissens et al. 2012).

However, two recent studies deserve a more detailed review. The laboratory of Pr. Carrell performed a genome-wide association study (GWAS) on 52 oligozoospermic, 40 non-obstructive azoospermic (NOA) patients and 80 controls (Aston and Carrell 2009) and expanded their study by analyzing identified SNPs on 141 oligospermic and 80 NOA patients and 158 controls (Aston, Krausz et al. 2010). Their results identified several promising SNPs associated with oligo or azoospermia, but larger cohorts are needed to definitively confirm these associations.

In a much more powerful study, Hu *et al* analyzed 2927 NOA patients and 5734 controls from the Han population and confidently identified 3 common variants significantly associated with NOA. One of these variants is close to *PRMT6*, the second one to *PEX10* and the last one to *SOX5*. All these genes are potentially involved in spermatogenesis, but no effect of these variants on the level of expression of the associated genes has been demonstrated so far (Hu, Xia et al. 2011). Thus, further functional studies are needed, but this work strongly suggests that NOA, like other complex pathologies, is multigenic and that no single SNP is responsible for a large proportion of azoospermia.

Compared to multi-organ pathologies, only a few non-syndromic genetic causes of human infertility have been described so far, despite the fact that it is estimated that some infertility cases could be explained by genetic causes (Jamsai and O'Bryan 2011; Krausz 2011) and that over 200 infertile or subfertile genetic mouse models have been described (Matzuk and Lamb

2008). This number gives an idea of the number of genes that could be involved in male and female gametogenesis and consequently, the complexity of the process. So far, very little has been discovered in the field of human male reproductive genetics. This slow progress has two main causes: 1) clinical evaluations are preliminary and often limited to the presence or absence of spermatozoa in the ejaculate. This has been reinforced by the advent of the intracytoplasmic sperm injection (ICSI) technique, which requires a very limited number of spermatozoa, even immotile ones, to fertilize the cohort of oocytes, 2) spermatogenesis is a highly sophisticated process involving complex molecular pathways requiring hundreds of genes. Consequently, genetic tests proposed to infertile couples are limited to karyotype analysis, CFTR mutation screening and Y-chromosome deletion screening (Krausz 2011). The worldwide efforts devoted to the field of human genetics of infertility are expected to provide new genetic tests in the near future. At the present time the field is mainly, but not exclusively, concentrating its effort on male genetics of infertility. Indeed, because of the easy access to male gametes, diagnoses are of better quality, allowing a better categorization of the different phenotypes, which is fundamental for carrying out genetic studies.

Results

DPY19L2 and *SPATA16* two autosomal genes implicated in human globozoospermia

Globozoospermia is a rare (incidence < 0.1%) and severe teratozoospermia characterized by round headed spermatozoa lacking an acrosome. Two types of globozoospermia have been described: type I characterized by the presence of 100% round-headed acrosomeless spermatozoa and type II characterized by the presence of some sperm with acrosomes (Dam, Feenstra et al. 2007). The acrosome plays a crucial role during fertilization by allowing spermatozoa to travel through the zona pellucida in order to reach the oocyte cytoplasmic membrane and by bringing oocyte activation factor(s) (Ikawa, Inoue et al. 2010). Therefore, patients suffering from a complete form of globozoospermia are infertile. Round headed spermatozoa do not present chromosomal abnormalities (Carrell, Emery et al. 1999; Viville, Mollard et al. 2000; Machev, Gosset et al. 2005) and pregnancy can be obtained through Intra cytoplasmic sperm injection (ICSI), albeit at a very low frequency (Coetzee, Windt et al. 2001; Kilani, Ismail et al. 2004; Dirican, Isik et al. 2008; Banker, Patel et al. 2009; Bechoua, Chiron et al. 2009; Sermondade, Hafhouf et al. 2011). Through the study of family cases, we recently identified the major gene implicated in human globozoospermia: *DPY19L2*. *DPY19L2* belongs to the *DPY19L* family, encompassing four genes and six pseudogenes and showing an evolutionary history that involved several duplications and pseudogenizations. In fact, *DPY19L1*, *DPY19L2* and the six pseudogenes were identified while analyzing the physical organization of LCRs on chromosome 7. It started when Andrew *et al*. identified eight regions, each over 5Kb, that contain duplicated DNA with >90% identity (Carson, Cheung et al. 2006). Seven of these regions are located on chromosome 7, while one region is located on chromosome 12 and are called LCR7A to LCR7H. Characterizing the genomic structure and gene content of the eight LCRs, Andrew *et al*. showed that LCR7A on chromosome 7 contains *DPY19L1*, while LCR7C on chromosome 12 contains *DPY19L2*. *DPY19L1* has two pseudogenes, *DPY19L1P1* and *DPY19L1P2*, located within LCR7B while *DPY19L2* has four pseudogenes, *DPY19L2P1*, *DPY19L2P2*, *DPY19L2P3* and *DPY19L2P4* located within LCR7A, LCR7D, LCR7E and LCR7F respectively. *DPY19L1* and *DPY19L2* share homology with the *C. elegans* gene *DPY-19*: 43% and 38% respectively. The presence of these two genes within a set of related LCR, as well as the presence of the six pseudogenes, indicates that *DPY19L1* and *DPY19L2* have undergone duplication followed by pseudogenization. The ancestral copy of *DPY19L2* is located on chromosome 7 and is known as *DPY19L1*. The latter underwent an ancient duplication, creating a duplicate copy (*DPY19L2*) adjacent to the gene on chromosome 7. Recent inter-chromosomal duplication occurred later, creating a second copy of *DPY19L2* on chromosome 12. Sequence evolution

introduced two STOP codons into the ancestral copy (*DPY19L2* on chromosome 7), leading to its pseudogenization. The silenced copy on chromosome 7 is known as *DPY19L2P1*. The hypothesis that *DPY19L2* has undergone genomic relocation is strengthened by the evidence that the functional copy of *Dpy19l2* in mice corresponds to a pseudogene of *DPY19L2* in human on chromosome 7 (Carson, Cheung et al. 2006). As for *DPY19L3* and *DPY19L4*, they are located on chromosome 19 and 8, respectively and do not reside in any segmental duplications.

DPY19L2, located on 12q14.2, has 22 coding exons and is flanked by two low copy repeat (LCR) regions sharing 96.5% identity and encodes for a transmembrane protein. A large deletion of about 200kb, encompassing the whole *DPY19L2*, was characterized through the study of a Jordanian consanguineous family and in three additional unrelated patients (Koscinski, Elinati et al. 2011). The mechanism underlying the deletion is most probably due to a non-allelic homologous recombination (NAHR) between the LCR. Indeed, NAHR mechanisms are mediated by highly homologous repetitive sequences, such as Alu sequences, and account for some of the non-recurrent rearrangements (Gu, Zhang et al. 2008). We excluded an ancient founder effect and suggested that the deletion may rely on the architectural feature of the locus, since two different breakpoints were identified in patients originated from different regions (Gu, Zhang et al. 2008). We excluded an ancient founder effect and we suggested that the deletion may rely on the architectural feature of the locus since two different breakpoints were identified in patients originated from different regions.

In our initial description, we found a rate of 19% of our patients (4 out of 21) presenting such a deletion (Koscinski, Elinati et al. 2011). In contrast, a much higher rate of 75% was described in a cohort of 20 patients, most of them originating from Tunisia (Harbuz, Zouari et al. 2011). The difference is probably due to the broader geographic distribution of our patients originated from seven different countries. Therefore, we pursued our study by screening for all types of *DPY19L2* mutations in the largest cohort of globozoospermic patients to date, including 64 patients (from 13 different countries) and corresponding to 54 genetically independent individuals. Out of a total of 54 individuals analysed, 36 (66.7%) showed a mutation in *DPY19L2*, 24 of them being homozygous for the deletion of 200kb, 8 being heterozygous composite associating the 200kb deletion and a point mutation and 4 being homozygous, for point mutations or small deletion. Out of 36 mutated patients, 69.4% are homozygous deleted, 19.4% heterozygous composite and 11.1% showed a homozygous point mutation. Analysing the products of the PCR across the deletion of *DPY19L2* allowed us to

identify 9 breakpoints inside the LCRs that clustered in two recombination hotspots: both containing direct repeat elements (AluSq2 in hotspot 1, THE1B in hotspot 2). In this study, we showed that *DPY19L2* is the major gene implicated in human globozoospermia and we confirmed that the deletion of the gene is due to NAHR. Therefore we present globozoospermia as a first autosomal genomic disorder known to impair fertility in men.

Concerning the function of DPY19L2, initially we thought that the protein was implicated in the polarization of the spermatozoa. This hypothesis was later rejected by examination of the immature germ cells in the semen of three globozoospermic patients. There was no abnormality in the early steps of spermatogenesis and the spermatids were correctly polarized. Furthermore, Harbuz *et al*, showed that the protein is only expressed in the testis and not in the mature sperm (Harbuz, Zouari et al. 2011). Using *Dpy19l2* knockout mice, Pierre *et al.* demonstrated that the protein is expressed predominantly in spermatids with a very specific localization, restricted to the inner nuclear membrane facing the acrosomal vesicle. *Dpy19l2*$^{-/-}$ animals showed a similar phenotype to *DPY19L2* deleted or/and mutated men. Compared to wild type mice, *Dpy19l2*$^{-/-}$ male mice showed a reduction in testis weight and size with a disorganization of the seminiferous tubules. The absence of Dpy19l2 leads to the destabilization of both the nuclear dense lamina (NDL) and the junction between the acroplaxome and the nuclear envelope. Consequently, the acrosome and the manchette fail to link to the nucleus, leading to the disruption of vesicular trafficking, failure of sperm nuclear shaping and eventually to the elimination of the unbound acrosomal vesicles (Pierre, Martinez et al. 2012).

Chapter VI

DPY19L2 deletion as a major cause of globozoospermia

DPY19L2 Deletion as a Major Cause of Globozoospermia

Isabelle Koscinski,[1,2,11] Elias ElInati,[2,11] Camille Fossard,[2] Claire Redin,[2] Jean Muller,[2,3] Juan Velez de la Calle,[4] Françoise Schmitt,[5] Mariem Ben Khelifa,[6] Pierre Ray,[6,7,8] Zaid Kilani,[9] Christopher L.R. Barratt,[10] and Stéphane Viville[1,2,*]

Globozoospermia, characterized by round-headed spermatozoa, is a rare (< 0.1% in male infertile patients) and severe teratozoospermia consisting primarily of spermatozoa lacking an acrosome. Studying a Jordanian consanguineous family in which five brothers were diagnosed with complete globozoospermia, we showed that the four out of five analyzed infertile brothers carried a homozygous deletion of 200 kb on chromosome 12 encompassing only *DPY19L2*. Very similar deletions were found in three additional unrelated patients, suggesting that *DPY19L2* deletion is a major cause of globozoospermia, given that 19% (4 of 21) of the analyzed patients had such deletion. The deletion is most probably due to a nonallelic homologous recombination (NAHR), because the gene is surrounded by two low copy repeats (LCRs). We found *DPY19L2* deletion in patients from three different origins and two different breakpoints, strongly suggesting that the deletion results from recurrent events linked to the specific architectural feature of this locus rather than from a founder effect, without fully excluding a recent founder effect. *DPY19L2* is associated with a complete form of globozoospermia, as is the case for the first two genes found to be associated with globozoospermia, *SPATA16* or *PICK1*. However, in contrast to *SPATA16*, for which no pregnancy was reported, pregnancies were achieved, via intracytoplasmic sperm injection, for two patients with *DPY19L2* deletion, who then fathered three children.

In a previous work, we described the identification of a mutation in *SPATA16* (MIM 609856) responsible for a complete form of globozoospermia (MIM 102530) in three affected brothers of an Ashkenazi Jewish family.[1] We report here the involvement of *DPY19L2* on chromosome 12 in cases of complete forms of globozoospermia. Our study started with the analysis of a Jordanian family of ten siblings, including two sisters and three brothers who were naturally fertile (all of them having 3–7 children) and five affected brothers with a complete form of globozoospermia (Figure 1A). Intracytoplasmic sperm injection (ICSI) was performed for the five brothers at the Farah Hospital, Amman, Jordan, and despite a total of 20 cycles there was only one single-term pregnancy, as well as two miscarriages.[2] Because of the consanguineous marriage of the grandparents, who were first-degree cousins, we postulated that the genetic abnormality was transmitted as an autosomal-recessive disorder. Four of the five affected individuals, Globo1 to Globo4 (IV-1 to IV-4 in Figure 1A), and all three fertile individuals, brothers B1 to B3 (IV-5 to IV-7 in Figure 1A), consented to give blood samples for research (Figure 1A).

We performed a genome-wide scan analysis of all seven brothers, using 10K SNP arrays (Affymetrix Genechip). Regions of homozygosity were defined by the presence of at least 25 consecutive homozygous SNPs. We identified a unique region of 30 homozygous SNPs shared by the four affected brothers; the fertile brothers were heterozygous for this region (Figure 1B). This region of about 6.4 Mb on chromosome 12 (positions 63060074 to 69409722, on the GRCh37/hg19 version of the human genome, corresponding to SNPs rs345945 and rs2172989; see Table S1 available online) contains 101 genes, 40 of which are expressed in the testis, according to the UCSC Genome Browser[3] (Table S2), and thus are potentially involved in spermatogenesis. We selected *DPY19L2* as the most plausible candidate gene because of its predominant testis expression and its potential involvement in cellular polarization.[4] The spermatozoon is a highly polarized cell, and the initial stages of spermiogenesis involve polarization of the round spermatid. Functional studies of DPY-19 (an ortholog of DPY19L2) in *C. elegans* highlighted its role in the establishment of polarity in the migrating neuroblasts.[4] *DPY19L2* is composed of 22 exons and belongs to the *DPY19L* family[5] coding for proteins harboring 9–11 predicted transmembrane domains. This gene family, encompassing four genes and six pseudogenes, has a complex evolutionary history involving several duplications and pseudogenizations.[5] In particular, *DPY19L2* stems from an initial duplication of *DPY19L1* on chromosome 7, followed by a recent relocalization on chromosome 12 in humans, leaving the initial copy on chromosome 7 as a pseudogene (*DPY19L2P1*). This evolutionary event may rely on the presence of low copy repeats

[1]Service de Biologie de la Reproduction, Centre Hospitalier Universitaire, Strasbourg, F-67000 France; [2]Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Institut National de Santé et de Recherche Médicale (INSERM) U964/Centre National de Recherche Scientifique (CNRS) UMR 1704/Université de Strasbourg, 67404 Illkirch, France; [3]Laboratoire de Diagnostic Génétique, CHU Strasbourg, Nouvel Hôpital Civil, 67000 Strasbourg, France; [4]Unité FIV, Clinique Pasteur, 29200 Brest, France; [5]Laboratoire de Biologie de la Reproduction, Centre hospitalier de Mulhouse 68100 Mulhouse, France; [6]Faculté de Médecine-Pharmacie, Domaine de la Merci, Université Joseph Fourier, Cedex 9, Grenoble, France; [7]TIMC-IMAG, UMR CNRS 5525, Team AGIM, 38710 La Tronche, France; [8]Biochimie et Génétique Moléculaire, CHU Grenoble, 38700 Grenoble, France; [9]Farah Hospital, Amman 11183, Jordan; [10]Reproductive and Developmental Biology, Maternal and Child Health Science Laboratories, Centre for Oncology and Molecular Medicine, Ninewells Hospital, University of Dundee, Dundee, DD1 9SY, Scotland
[11]These authors contributed equally to this work
*Correspondence: viville@igbmc.fr
DOI 10.1016/j.ajhg.2011.01.018.

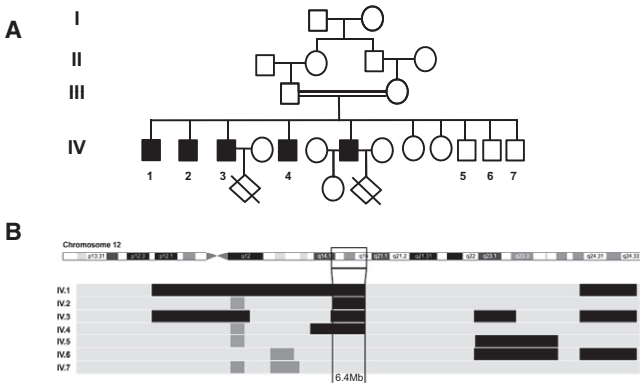

Figure 1. Pedigree and Linkage Analysis of the Jordanian Family
(A) Jordanian pedigree showing second-degree consanguinity.
(B) SNP array results of the four infertile and three fertile brothers for the region of chromosome12. Shared regions of homozygosity are visualized by the HomoSNP software, which displays one patient per line. The areas of homozygosity with 25 or more SNPs are black, whereas homozygosity regions defined by 15–25 consecutive SNPs are gray. Regions of heterozygosity are light gray. The four affected siblings, Globo1 to Globo4 (IV-1 to IV-4), share a region of homozygosity of 6.4 Mb on chromosome 12, which is not shared by the three fertile brothers, B1 to B3 (IV-5 to IV-7).

(LCRs) surrounding the DPY19L2 locus (Figure 2A). Because of the high level of conservation between *DPY19L2* and *DPY19L2P1*,[5] the design of *DPY19L2*-specific oligonucleotides was complex. PCR conditions were optimized for exons 2, 7, 9, 10, 13, 17, and 21, resulting in specific amplifications in controls but never in the patients, suggesting a large deletion of the whole gene (Figure 2B). In order to better characterize the two described LCRs surrounding *DPY19L2*, we performed a Blastn analysis[6] using a 100 kb region up- and downstream of the gene. This approach revealed a highly similar duplicated region (96.5% identity), corresponding to the two identified LCRs (Figure 2B). The telomeric LCR (named here LCR1) is 26,998 bp long and is located upstream of exon 1 (64,119,249–64,146,247), and the centromeric LCR (named here LCR2) is 28,200 bp long and starts ~1000 bp downstream of exon 22 (63,923,419–63,951,619). In order to define as precisely as possible the breakpoints within the recombined LCRs, specific oligonucleotides were designed to walk on both sides of the deletion (Figures 2B–2D and Table S4). This allowed us to define a first target segment of 107 bp, which we named DPY19L2-BP1. Further refinement is not possible because of the high percentage of identity shared by the two LCRs (Figure 2E and Figure S1). The exact size of the deletion is 195,150 bp. All three healthy brothers B1, B2 and B3 are homozygous wild-type.

The *DPY19L2* deletion was then screened in 24 globozoospermic patients from 20 independent families, originating from Algeria, France, Iran, Italy, Lybia, Morocco, and Tunisia. Twelve of them had a complete form of globozoospermia, and six had a partial form (defined as between 20% and 90% of the spermatozoa having round heads[7]). For the remaining two, we did not have any detailed information (Table 1). We found a similar deletion of *DPY19L2* in three unrelated cases presenting with a complete form of globozoospermia (Figures 2C and 2D), two cases originating from France (Globo17 and Globo18) and one familial case from Algeria comprising two brothers (Globo7 and Globo8). One of the French patients (Globo17) had two children, born after ICSI treatment (Table 1). We localized the breakpoints for all three additional patients. For Globo18, the deletion breakpoints are located, as suggested by the PCR walk (Figure 2D), in the same 107 bp LCR area (DPY19L2-BP1) as that of the Jordanian family. Both of his parents are heterozygous for the deletion, as demonstrated by the possibility of amplifying both the recombined region (i.e., absence of *DPY19L2*) and exon 10 (i.e., at least one copy of *DPY19L2*) (Figure 2C). The analysis of Globo7 and Globo8 allowed us to identify a different breakpoint, which we named DPY19L2-BP2, (Figure 2D), in an area of 296 bp, positioned 169 bp 3′ to the 107 bp area. Interestingly, Globo17 is a heterozygous composite, because the two different breakpoints (DPY19L2-BP1 and DPY19L2-BP2) could be observed (Figure 2D). The analysis of his parents showed their heterozygous status for the deletion and allowed us to detect the two different breakpoints on each allele. The maternal allele carries DPY19L2-BP2 and the paternal allele carries DPY19L2-BP1. For both parents we could amplify exon 10, demonstrating the presence of at least one copy of *DPY19L2* (Figure 2C). The deletion of *DPY19L2* was not found in the homozygous state in 105 European-descent males and 101 fertile Jordanian males.

Several studies have reported copy-number variants (CNVs) at the genome-scale level that are fully available and searchable in databases such as the Database of Genomic Variants (DGV).[8] DGV reports, in its latest version (November 2010), 66,741 CNVs from 42 studies. At the DPY19L2 locus, DGV references 29 CNVs that are described by 12 different studies encompassing different ethnic groups (Figure 3 and Table S3). Considering only the nonredundant CNVs identified in non-BAC (bacterial artificial chromosome) studies and those that are fully overlapping *DPY19L2*, only five studies can be considered.[8–13] The detailed inventory of the CNVs revealed 64 duplications and 22 heterozygous deletions out of 4886 patients studied, whereas no homozygous deletion was reported (Table 2). Interestingly, about three times more duplications than deletions can be observed at this locus. This is surprising because in most of the cases studied, deletions are more frequent than duplications.[14] We do not have a clear explanation for this, but it would be of interest to determine whether the heterozygous status is affecting male fertility, which would explain the selection against deletion. The frequency of heterozygous deletion in the surrounding region of *DPY19L2* is 1/222, implying an overall disease risk of ~1/200,000 (see Table 2). Examining the localization of the CNVs, one can observe that all deletion breakpoints

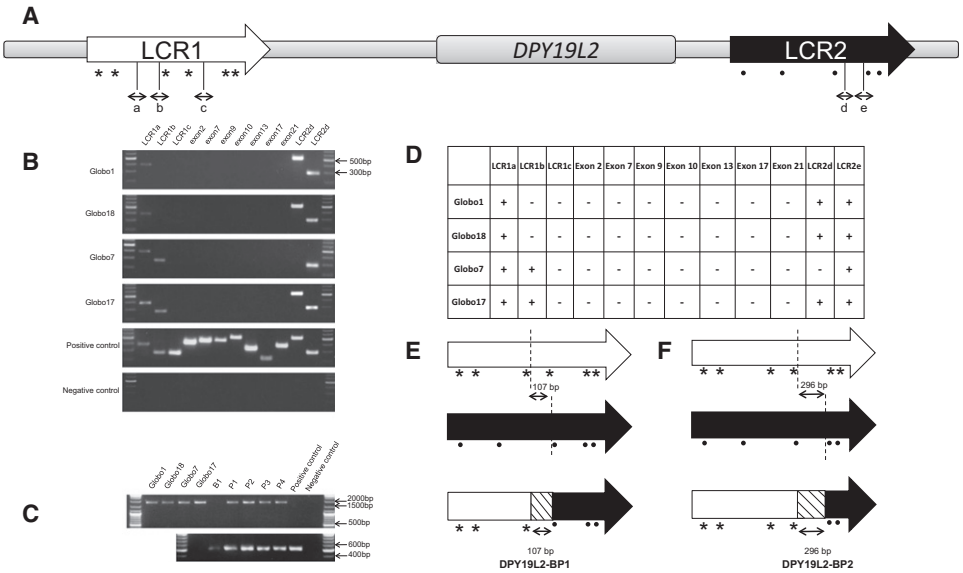

Figure 2. Analysis of *DPY19L2*

(A) Schematic representation of *DPY19L2* flanked by two LCRs, LCR1 and LCR2. LCR1 is located upstream of the gene and LCR2 downstream. Different sequences were amplified on both sides of the gene in order to determine the breakpoints: "a," "b," and "c" sequences are localized in LCR1, whereas "d" and "e" sequences are localized in LCR2. Mismatches between the two LCRs sequences are represented by (*) in the 5′ side and by (.) in the 3′ side.

(B) PCR results of *DPY19L2* exons and flanking regions. Patients Globo1, Globo7, Globo17, and Globo18 are deleted for exons 2, 7, 9, 10, 13, 17, and 21, whereas the positive control showed an amplification of all of these exons. All patients showed specific amplifications for region "a" in LCR1 and region "e" in LCR2. Globo1 and Globo18 are missing region "b" in LCR1, which is not the case for Globo7 and Globo17. Globo7 is the only patient missing region "d" in LCR2. A DNA ladder was included on both sides of each gel for accurate determination of the PCR band size.

(C) PCR of *DPY19L2* breakpoints. This PCR was performed with the use of the "LCR1a forward" and "DPY19L2-BP reverse" primers (Table S4). The four globozoospermic patients showed a fragment of 1700 bp. The fertile brother (B1, IV-5 in Figure 1A) is homozygous wild-type, because no fragment was amplified, whereas parents of Globo18 (P1 and P2) and Globo17 (P3 and P4) are heterozygous. A fertile man (positive control) was used as control of the PCR specificity. Amplification of exon 10 (lower panel) was used to control the presence or absence of *DPY19L2*.

(D) Table summarizing the PCR results of the four patients: (+) indicates presence of a PCR band at the corresponding size, and (-) indicates absence of a PCR band.

(E) Schematic representation of the breakpoint area determined for Globo1, Globo18, and the paternal allele of Globo17. The breakpoint area (DPY19L2-BP1) identified by PCR and sequencing is defined between the two discontinued lines and represented by a hatched box in the recombined LCR.

(F) Schematic representation of the homologous recombination that occurred in Globo7 and the maternal allele of Globo17. The breakpoint area (DPY19L2-BP2) identified by PCR and sequencing is defined between the two discontinued lines and represented by a hatched box in the recombined LCR.

recorded in DGV fall into the two LCRs, and this is also true for most of the duplications (Figure 3).

Considering the presence of highly conserved LCRs surrounding *DPY19L2*, the most probable mechanism to explain the multiple occurrence of this deletion would be NAHR between highly similar sequences. Indeed, such a mechanism has been reported for a large number of disorders involving deletions, duplications, inversions, or gene fusions[15–17] (reviewed in[14]). It is estimated that this is one of the most common mechanisms responsible for genomic disorders. So far, no obvious founder effect has been described for genetic disorders involving NAHR.[14] That we found the *DPY19L2* deletion in patients from three different origins (the two French cases are not known to be related), that two different breakpoints were observed, and that Globo18 is not sharing the same haplotype of the Jordanian brothers (Table S1) strongly suggest that the deletion results from recurrent events linked to the specific architectural feature of this locus rather than from a founder effect, without fully excluding a recent founder effect. It is worth noting that the two identified breakpoints are located near an Alu repeat (AluSq2), one being fully included in this repeated element (Figure S1).

Interestingly, the patients with *DPY19L2* deletion present a complete globozoospermia yet have normal or, in one case (Globo8), near-normal concentration of sperm (Table 1). This contrasts with the case in which *SPATA16* is mutated, suggesting that DPY19L2 may disrupt only spermiogenesis and not germ cell proliferation and meiosis. It should also be noted that, despite the testicular expression of the other three members of the *DPY19L* family,[5] it does not appear that there is functional redundancy. No homozygosity was found for the loci of the other functional DPY19L genes.

We selected *DPY19L2* as the primary candidate linked to globozoospermia in our Jordanian family because of its

Table 1. List of Globozoospermic Patients

Patient	Origin	SNP Screening	Sperm Concentration (Million/mL)	Progressive Motility (%)	Acrosome less Morphology (%)	Fertilization Attempts and Results
Complete Globozoospermia						
Globo1	Jordan	Yes	52	22	100	6 ICSI failed
Globo2	Jordan	Yes	223	4	100	3 ICSI failed
Globo3	Jordan	Yes	118	4	100	4 ICSI failed
Globo4	Jordan	Yes	22	30	100	3 ICSI failed (1 miscarriage)
Globo5	France	No	ND	ND	100	ND
Globo6	France	No	ND	ND	100	ND
Globo7	Algeria	No	90	10	100	ND
Globo8	Algeria	No	13.2	2	100	ND
Globo9	France	No	0.35	ND	100	1 ICSI failed
Globo10	France	Yes	38	ND	100	1 ICSI succeeded
Globo11	France	Yes	109	40	100	2 ICSI failed
Globo12	ND	No	ND	ND	100	DSI
Globo13	Iran	No	10.5	ND	100	ND
Globo14	Tunisia	No	52.6	25	100	ND
Globo15	Tunisia	No	ND	ND	100	1 ICSI failed
Globo16	Libya	No	4	8	100	3 ICSI failed
Globo17	France	Yes	78	45	100	IMSI succeeded twice
Globo18	France	Yes	71	30	100	ND
Partial Globozoospermia						
Globo19	Italy	No	50	ND	ND	1 ICSI failed
Globo20	France	Yes	0.6	ND	84	1 IUI + 3 ICSI failed
Globo21	France	Yes	2	ND	ND	2 ICSI failed
Globo22	Morocco	No	ND	ND	ND	ND
Globo23	France	Yes	ND	ND	95	1 ICSI succeeded
Globo24	France	Yes	ND	ND	98	1 ICSI succeeded
Globo25	Tunisia	No	59	50	88	ND
Globo26	Tunisia	No	100	49	99	ND
Globo27	ND	No	ND	ND	ND	ND
Globo28	ND	No	ND	ND	ND	ND

Abbreviations are as follows: ND, not determined; ICSI, intracytoplasmic sperm injection; IUI, intrauterine insemination; IMSI, intracytoplasmic morphologically selected sperm injection; DSI, donor sperm insemination.

testicular expression and function of the DPY-19 ortholog in *C. elegans*, involved in the polarization of neuroblasts.[4] This would suggest that DPY19L2 is not directly involved in the biogenesis of the acrosome, as for SPATA16[1] or PICK1 (MIM 605926),[18,19] but rather in preceding steps associated with, for example, the polarity of the spermatid, therefore allowing the correct positioning of the different organelles such as the acrosome or the flagella. However, this hypothesis is not supported by the examination of the immature germ cells in the semen of three globozoospermic patients (Globo7, Globo17, and Globo18). For Globo7, all stages of spermatogenesis could be observed (Figure 4A) and there was no obvious abnormality in the earlier stages of germ cells. In contrast, no proacrosomal vesicle was observed in the round spermatids, yet they appeared correctly polarized. Indeed, the polarization of the spermatid was indicated by the correct positioning of the chromatoid body (in the area of the attachment of the flagellum to the spermatid nucleus) observed in elongating spermatids. In Globo17 and Globo18 semen samples, we observed elongating spermatids as well as spermatozoa that are comparable to that of Globo7

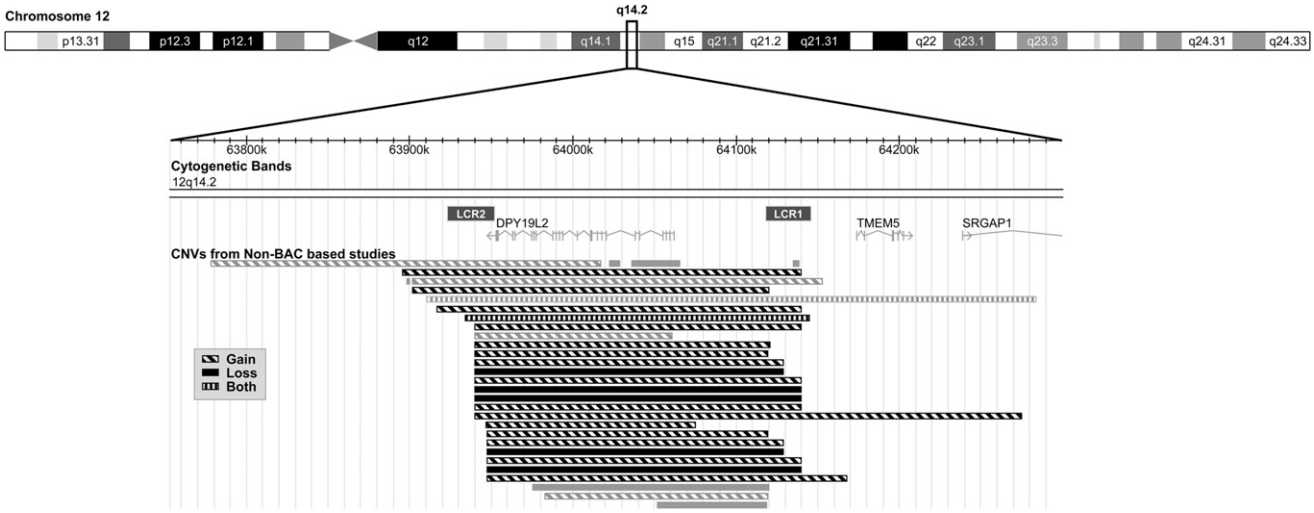

Figure 3. CNVs at the DPY19L2 Locus

CNVs are displayed as rectangles according to the following scheme: gains are drawn with hatched lines, losses with plain rectangles, and vertical lines are used when both gains and losses are observed. The CNVs excluded from the count are colored in gray. LCR1 and LCR2 are displayed as dark gray rectangles on top of the gene representation.

(Figures 4C–4E). In contrast to what is observed in *C. elegans*, the function of DPY19L2 does not seems to be involved in the polarization of the spermatozoa but, rather, in the formation of the acrosome.

DPY19L2 is the third gene identified as being associated with globozoospermia[1,19] and, according to our data, the most frequently mutated in this phenotype. Indeed, in our series of 21 unrelated cases, we identified four individual cases, meaning that the *DPY19L2* deletion is found in 19% of the globozoospermic cases. Looking at the frequency of heterozygous deletion in control populations, we would expect an incidence of 1/200,000 for homozygous individuals, whereas the incidence of globozoospermia is estimated at less than 0.1% in infertile males.[7] Although we cannot fully exclude a positioning effect, we do not believe that this could explain the observed phenotype, because the four genes surrounding the DPY19L2 locus (two centromeric: *AVPR1A* and *PPM1H*; two telomeric: *TMEM5* and *SRGAP1*) are not expressed in the testis and their function has no relation with cell polarity or acrosome formation.

Considering that round-headed spermatozoa do not have a higher incidence of chromosomal abnormalities[20–22] and that pregnancy can be obtained through ICSI, albeit at a low frequency,[2,23–26] it is reasonable to propose ICSI treatment for all globozoospermic patients. It will be interesting to enlarge the cohort of patients and establish which conditions are able to give rise to live births. Indeed, so far, no pregnancies have been obtained with patients carrying a *SPATA16* mutation[1] (data not shown), in contrast to patients with *DPY19L2* deletion. Such an analysis may suggest that a genetic test would assist in determining the best option for treatment.

Our strategy has proven its interest and should be extended to other countries, given that family clustering of male infertility cases has been found in other studies and appears to be related to patterns of consanguineous marriage over generations.[27,28]

Supplemental Data

Supplemental Data include Supplemental Material and Methods, four tables, and one figure and can be found with this article online at http://www.cell.com/AJHG.

Acknowledgments

We are very grateful to our colleagues Manuel Mark, for his help in interpreting the gametogenesis of our patients, and Michel Koenig and Regen Drouin, for their critical reading of this manuscript. We would also like to thanks all of the clinicians—C. Jimenez (Dijon), F. Carré-Pigeon (Reims), J.M. Grillot (Marseille), F. Brugnon (Clermont Ferrand), D. De Briel (Colmar), and S. Declève-Paulhac

Table 2. CNV Frequency of Individuals at the *DPY19L2* Locus and Expected Prevalence of Homozygotes

	Pinto et al.[11]	de Smith et al.[10]	Conrad et al.[9]	Itsara et al.[13]	Shaikh et al.[12]	Total	Frequency	Expected Homozygous Deletion Frequency
Gain	2	1	4	22	35	64	1/76	
Loss (Htz)	0	3	0	5	14	22	1/222	1/197,136
Population	506	50	450	1854	2026	4886		

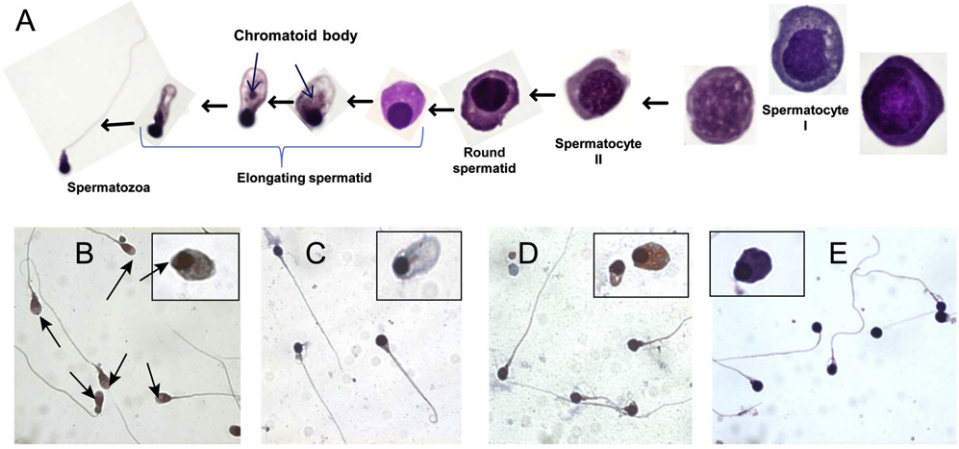

Figure 4. Seminal Germ Cells in Patients with *DPY19L2* Deletion
(A) Detailed seminal germ cell of Globo7. There is no evidence of defective spermatogenesis in immature germ cells; the spermatid polarization and elongation is consistent with controls: e.g., the correct positioning of the chromatid body (blue arrows), except for the absence of the acrosomal vesicle.
(B) Spermatozoa and spermatids of fertile controls. The acrosome is clearly observed in spermatozoa and elongating spermatid (arrows).
(C–E) Spermatozoa and spermatid of Globo17, Globo18, and Globo7, respectively. The spermatozoa present the typical round head and an absence of acrosome. No evident polarization abnormality in seminal spermatids was observed in the globozoospermic patients.

(Limoges)—who have sent us samples from globozoospermic patients and for whom we still expect to find a mutation. We would like to recognize the intellectual input of Jean-Louis Mandel, Christelle Thibault-Carpentier, and Mirna Assoum and the technical help of Anne-Sophie Jaeger and Nathalie Drouot. We would like to thank Frederic Plewniak (plewniak@igbmc.fr) from the Bioinformatics Platform of Strasbourg for the development of the HomoSNP software. We are grateful to the Institute of Genetics and Molecular and Cellular Biology (IGBMC) services and the Department of Human Genetics for their invaluable assistance. This work was supported by the French Centre National de la Recherche Scientifique (CNRS), Institut National de la Santé et de la Recherche Médicale (INSERM), the Ministère de l'Education Nationale, de l'Enseignement Supérieur et de la Recherche, the University of Strasbourg, the University Hospital of Strasbourg, and the Agence de la BioMédecine.

Received: November 22, 2010
Revised: January 6, 2011
Accepted: January 27, 2011
Published online: March 10, 2011

Web Resources

The URLs for data presented herein are as follows:

Database of Genomic Variants, http://projects.tcag.ca/variation/
UCSC Genome Browser Gateway, http://genome.ucsc.edu/cgi-bin/hgGateway
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim

References

1. Dam, A.H., Koscinski, I., Kremer, J.A., Moutou, C., Jaeger, A.S., Oudakker, A.R., Tournaye, H., Charlet, N., Lagier-Tourenne, C., van Bokhoven, H., and Viville, S. (2007). Homozygous mutation in SPATA16 is associated with male infertility in human globozoospermia. Am. J. Hum. Genet. *81*, 813–820.
2. Kilani, Z., Ismail, R., Ghunaim, S., Mohamed, H., Hughes, D., Brewis, I., and Barratt, C.L. (2004). Evaluation and treatment of familial globozoospermia in five brothers. Fertil. Steril. *82*, 1436–1439.
3. Rhead, B., Karolchik, D., Kuhn, R.M., Hinrichs, A.S., Zweig, A.S., Fujita, P.A., Diekhans, M., Smith, K.E., Rosenbloom, K.R., Raney, B.J., et al. (2010). The UCSC Genome Browser database: update 2010. Nucleic Acids Res. *38* (Database issue), D613–D619.
4. Honigberg, L., and Kenyon, C. (2000). Establishment of left/right asymmetry in neuroblast migration by UNC-40/DCC, UNC-73/Trio and DPY-19 proteins in C. elegans. Development *127*, 4655–4668.
5. Carson, A.R., Cheung, J., and Scherer, S.W. (2006). Duplication and relocation of the functional DPY19L2 gene within low copy repeats. BMC Genomics *7*, 45.
6. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.
7. Dam, A.H., Feenstra, I., Westphal, J.R., Ramos, L., van Golde, R.J., and Kremer, J.A. (2007). Globozoospermia revisited. Hum. Reprod. Update *13*, 63–75.
8. 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.
9. Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P., et al; Wellcome Trust Case Control Consortium. (2010). Origins and functional impact of copy number variation in the human genome. Nature *464*, 704–712.
10. de Smith, A.J., Tsalenko, A., Sampas, N., Scheffer, A., Yamada, N.A., Tsang, P., Ben-Dor, A., Yakhini, Z., Ellis, R.J., Bruhn, L., et al. (2007). Array CGH analysis of copy number variation identifies 1284 new genes variant in healthy white males: implications for association studies of complex diseases. Hum. Mol. Genet. *16*, 2783–2794.

11. Itsara, A., Cooper, G.M., Baker, C., Girirajan, S., Li, J., Absher, D., Krauss, R.M., Myers, R.M., Ridker, P.M., Chasman, D.I., et al. (2009). Population analysis of large copy number variants and hotspots of human genetic disease. Am. J. Hum. Genet. *84*, 148–161.
12. Pinto, D., Marshall, C., Feuk, L., and Scherer, S.W. (2007). Copy-number variation in control population cohorts. Hum. Mol. Genet. *16* (Spec. No. 2), R168–R173.
13. Shaikh, T.H., Gai, X., Perin, J.C., Glessner, J.T., Xie, H., Murphy, K., O'Hara, R., Casalunovo, T., Conlin, L.K., D'Arcy, M., et al. (2009). High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. Genome Res. *19*, 1682–1690.
14. Gu, W., Zhang, F., and Lupski, J.R. (2008). Mechanisms for human genomic rearrangements. Pathogenetics *1*, 4.
15. Tézenas Du Montcel, S., Mendizabai, H., Aymé, S., Lévy, A., and Philip, N. (1996). Prevalence of 22q11 microdeletion. J. Med. Genet. *33*, 719.
16. Hoogendijk, J.E., Hensels, G.W., Gabreëls-Festen, A.A., Gabreëls, F.J., Janssen, E.A., de Jonghe, P., Martin, J.J., van Broeckhoven, C., Valentijn, L.J., Baas, F., et al. (1992). De-novo mutation in hereditary motor and sensory neuropathy type I. Lancet *339*, 1081–1082.
17. Lakich, D., Kazazian, H.H., Jr., Antonarakis, S.E., and Gitschier, J. (1993). Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A. Nat. Genet. *5*, 236–241.
18. Xiao, N., Kam, C., Shen, C., Jin, W., Wang, J., Lee, K.M., Jiang, L., and Xia, J. (2009). PICK1 deficiency causes male infertility in mice by disrupting acrosome formation. J. Clin. Invest. *119*, 802–812.
19. Liu, G., Shi, Q.W., and Lu, G.X. (2010). A newly discovered mutation in PICK1 in a human with globozoospermia. Asian J. Androl. *12*, 556–560.
20. Carrell, D.T., Emery, B.R., and Liu, L. (1999). Characterization of aneuploidy rates, protamine levels, ultrastructure, and functional ability of round-headed sperm from two siblings and implications for intracytoplasmic sperm injection. Fertil. Steril. *71*, 511–516.
21. Machev, N., Gosset, P., and Viville, S. (2005). Chromosome abnormalities in sperm from infertile men with normal somatic karyotypes: teratozoospermia. Cytogenet. Genome Res. *111*, 352–357.
22. Viville, S., Mollard, R., Bach, M.L., Falquet, C., Gerlinger, P., and Warter, S. (2000). Do morphological anomalies reflect chromosomal aneuploidies?: case report. Hum. Reprod. *15*, 2563–2566.
23. Banker, M.R., Patel, P.M., Joshi, B.V., Shah, P.B., and Goyal, R. (2009). Successful pregnancies and a live birth after intracytoplasmic sperm injection in globozoospermia. J. Hum. Reprod. Sci. *2*, 81–82.
24. Bechoua, S., Chiron, A., Delcleve-Paulhac, S., Sagot, P., and Jimenez, C. (2009). Fertilisation and pregnancy outcome after ICSI in globozoospermic patients without assisted oocyte activation. Andrologia *41*, 55–58.
25. Coetzee, K., Windt, M.L., Menkveld, R., Kruger, T.F., and Kitshoff, M. (2001). An intracytoplasmic sperm injection pregnancy with a globozoospermic male. J. Assist. Reprod. Genet. *18*, 311–313.
26. Dirican, E.K., Isik, A., Vicdan, K., Sozen, E., and Suludere, Z. (2008). Clinical pregnancies and livebirths achieved by intracytoplasmic injection of round headed acrosomeless spermatozoa with and without oocyte activation in familial globozoospermia: case report. Asian J. Androl. *10*, 332–336.
27. Inhorn, M.C., Kobeissi, L., Nassar, Z., Lakkis, D., and Fakih, M.H. (2009). Consanguinity and family clustering of male factor infertility in Lebanon. Fertil. Steril. *91*, 1104–1109.
28. Inhorn, M.C., Kobeissi, L., Nassar, Z., Lakkis, D., and Fakih, M.H. (2009). Consanguinity and Male Infertility in Lebanon: The Power of Ethnographic Epidemiology. In Anthropology and Public Health: Bridging Differences in Culture and Society, R.A. Hahn and M.C. Inhorn, eds. (New York: Oxford University Press), pp. 165–195.

Chapter VII

Globozoospermia is mainly due to *DPY19L2* deletion via non-allelic homologous recombination involving two recombination hotspots

Human Molecular Genetics, 2012, Vol. 21, No. 16 **3695–3702**
doi:10.1093/hmg/dds200
Advance Access published on May 31, 2012

Globozoospermia is mainly due to *DPY19L2* deletion via non-allelic homologous recombination involving two recombination hotspots

Elias Ellnati[1,†], Paul Kuentz[1,2,†], Claire Redin[1], Sara Jaber[1,3], Frauke Vanden Meerschaut[4], Joelle Makarian[1], Isabelle Koscinski[1,5], Mohammad H. Nasr-Esfahani[6], Aygul Demirol[7], Timur Gurgan[7], Noureddine Louanjli[8], Naeem Iqbal[9], Mazen Bisharah[9], Frédérique Carré Pigeon[10], H. Gourabi[11], Dominique De Briel[12], Florence Brugnon[13], Susan A. Gitlin[14], Jean-Marc Grillo[15], Kamran Ghaedi[6], Mohammad R. Deemeh[6], Somayeh Tanhaei[6], Parastoo Modarres[6], Björn Heindryckx[4], Moncef Benkhalifa[16], Dimitra Nikiforaki[4], Sergio C. Oehninger[14], Petra De Sutter[4], Jean Muller[1,17] and Stéphane Viville[1,18,*]

[1]Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Institut National de Santé et de Recherche Médicale (INSERM) U964, Centre National de Recherche Scientifique (CNRS) UMR1704, Université de Strasbourg, Illkirch 67404, France, [2]Centre Hospitalier Universitaire de Besançon, Besançon CEDEX 25030, France, [3]Institut Curie, 26 rue d'UIM, Paris Cedex 75248, France, [4]Department of Reproductive Medicine, University Hospital, De Pintelaan 185, Gent B-9000, Belgium, [5]Service de Biologie de la Reproduction, Centre Hospitalier Universitaire, Strasbourg 67000, France, [6]Department of Reproduction and Development, Reproductive Biomedicine Center, Royan Institute for Animal Biotechnology, ACECR, Isfahan, Iran, [7]Clinic Women Health, Infertility and IVF Center, Ankara, Turkey, [8]Laboratoire LABOMAC, Casablanca 20000, Morocco, [9]King Faisal Specialist Hospital and Research Center, Jeddah 21499, Kingdom of Saudi Arabia, [10]Service de Gynécologie-Obstétrique, CHU Reims, Institut mère-enfant Alix de Champagne, 45 rue Cognacq-Jay, Reims F-51092, France, [11]Royan Institute Reproductive Biomedicine and Stem Cell Research Center, Tehran, Iran, [12]Service de Microbiologie, Centre Hospitalier de Colmar, 39, avenue de la Liberté, Colmar 68024, France, [13]Laboratoire de Biologie de la Reproduction, Université Clermont 1, UFR Médecine, EA 975, Clermont Ferrand Cedex 1 F-63001, France, [14]Department of Obstetrics and Gynecology, The Jones Institute for Reproductive Medicine, Eastern Virginia Medical School, Norfolk, VA, USA, [15]Département d'Urologie, AP-HM, Salvatore Marseille, France, [16]ATL R&D Laboratory, 78320 la verrière et Unilabs, Paris 75116, France, [17]Laboratoire de Diagnostic Génétique, CHU Strasbourg, Nouvel Hôpital Civil, Strasbourg 67000, France and [18]Centre Hospitalier Universitaire, Strasbourg F-67000, France

Received March 28, 2012; Revised and Accepted May 22, 2012

To date, mutations in two genes, *SPATA16* and *DPY19L2*, have been identified as responsible for a severe teratozoospermia, namely globozoospermia. The two initial descriptions of the *DPY19L2* deletion lead to a very different rate of occurrence of this mutation among globospermic patients. In order to better estimate the contribution of *DPY19L2* in globozoospermia, we screened a larger cohort including 64 globozoospermic patients. Twenty of the new patients were homozygous for the *DPY19L2* deletion, and 7 were compound heterozygous for both this deletion and a point mutation. We also identified four additional mutated patients. The final mutation load in our cohort is 66.7% (36 out of 54). Out of 36 mutated patients, 69.4% are homozygous deleted, 19.4% heterozygous composite and 11.1% showed a homozygous point mutation. The mechanism underlying the deletion is a non-allelic homologous recombination (NAHR) between the flanking

*To whom correspondence should be addressed. Tel: +33 388653322; Fax: +33 388653201; Email: viville@igbmc.fr
†These authors contributed equally to this work.

low-copy repeats. Here, we characterized a total of nine breakpoints for the *DPY19L2* NAHR-driven deletion that clustered in two recombination hotspots, both containing direct repeat elements (*AluSq2* in hotspot 1, THE1B in hotspot 2). Globozoospermia can be considered as a new genomic disorder. This study confirms that *DPY19L2* is the major gene responsible for globozoospermia and enlarges the spectrum of possible mutations in the gene. This is a major finding and should contribute to the development of an efficient molecular diagnosis strategy for globozoospermia.

INTRODUCTION

Globozoospermia is a rare and severe teratozoospermia characterized by round-headed spermatozoa lacking an acrosome. The acrosome plays a crucial role during fertilization, allowing the spermatozoa to penetrate the zona pellucida and to reach the oocyte cytoplasmic membrane (1). Therefore, patients suffering from a complete form of globozoospermia are infertile. Round-headed spermatozoa do not present chromosomal abnormalities (2–4) and pregnancy can be obtained through ICSI, although at a low frequency (5–10). Previous analysis of globozoospermia families allowed us to identify mutations in two genes, *SPATA16* and *DPY19L2* (11,12). The deletion of exon 4 in *SPATA16* was found in an Ashkenazi Jewish family with three affected brothers. No other mutations were identified in a screen of 21 patients. A large deletion of ~200 kb encompassing the entire *DPY19L2* locus was detected in a consanguineous Jordanian family and in three additional unrelated patients (12). The gene, located on 12q14.2, has 22 exons encoding for a 9 transmembrane domain protein and is flanked by two low-copy repeats (LCRs) sharing 96.5% identity. The mechanism underlying the deletion is most probably a non-allelic homologous recombination (NAHR) between the flanking LCRs (13). Indeed, sequences with high nucleotide similarity (usually >95%) can serve as substrates for NAHR or ectopic recombination (14–16). Recently, a study of recurrent rearrangements associated with Smith–Magenis syndrome (SMS, MIM 182290) showed that NAHR crossover frequencies are correlated with the flanking LCR length and are inversely influenced by the distance between the LCRs (17). Homologous recombination on the Y chromosome is also known to impair fertility (18,19). However, so far, the deletion of *DPY19L2* is the first example where CNVs on autosomes are reported to be a causative infertility factor (20).

In our initial study, 19% of our patients (4 out of 21) were observed with such a deletion (12). In contrast, Harbuz *et al.* (21) detected a higher rate of *DPY19L2*-deleted patients (75%) in a cohort mainly composed of Tunisian patients. The difference is probably due to the broader geographic distribution of our patients (i.e. Algeria, France, Iran, Italy, Libya, Morocco and Tunisia).

In order to better estimate the contribution of *DPY19L2* in globozoospermia, we screened the largest cohort of globozoospermic patients to date, including 64 patients (from 13 different countries) and corresponding to 54 genetically independent individuals for all types of mutations. Twenty of the new patients were homozygous for the *DPY19L2* deletion, and seven were compound heterozygous for this deletion and a point mutation. We also identified four additional mutated patients. The final mutation load in our cohort is 66.7% (36 out of 54). Out of 36 mutated patients, 69.4% are homozygous deleted, 19.4% heterozygous composite and 11.1% showed a homozygous point mutation.

Recombination hotspots have been associated with genomic disorders and, when analyzing the sequence in the breakpoint (BP) cluster region, a common degenerate 13mer motif (CCNCCNTNNCCNC) was found (22), which has recently been elucidated as the *PRDM9* recognition binding motif (23–25). Here, we characterized nine new BPs for the *DPY19L2* NAHR-driven deletion defining two recombination hotspots, one containing the *PRDM9* recognition motif.

In this study, we confirmed that *DPY19L2* is the major gene responsible for globozoospermia, we enlarged the spectrum of possible mutations in the gene (deletion of the whole locus, nonsense, missense, splicing mutations and partial deletion) and we present globozoospermia as a new genomic disorder. This is a major finding and should contribute to the development of an efficient molecular diagnosis strategy for globozoospermia.

RESULTS

Deletion screening

In addition to our initial cohort of 21 patients, we recruited 33 new globozoospermic patients. As a first step, we screened for the deletion of *DPY19L2* in all 33 patients by performing a PCR of exon 10 and a PCR encompassing the previously described deletion BPs. Of these 33 new patients, 13 (Globo 29–31, 36, 39, 45, 48, 51, 54, 56, 57, 59, 61) did not show any amplification of exon 10, whereas the PCR of *DPY19L2* BPs revealed a fragment of 1700 bp, suggesting a homozygous deletion and a BP located within the 1700 bp fragment (Fig. 1A). Eight patients (Globo 28, 32, 33, 47, 50, 53, 58, 60) did not show any amplification at all, suggesting a homozygous deletion, but with a BP situated outside the tested region. We confirmed the deletion of most of the gene by testing for the presence of exons 4 and 16, which could not be amplified. Using specific oligonucleotides to walk on both sides of the deletion, we were able to amplify a common region of 1600 bp for two patients (Globo 28, 32). Sequencing of the 1600 bp fragment identified two BPs within an area of 117 bp (BP8 and BP9) (Fig. 2A). This second region of recombination is localized 9450 bp 3′ of the first one (Fig. 2B). For a third patient (Globo 33), we were able to amplify a fragment of 1500 bp. Sequencing of this fragment revealed another BP within a region of 265 bp (BP7) (Fig. 2A). This third region of recombination is situated 693 bp 3′ of the first one (Fig. 2B). For four patients (Globo

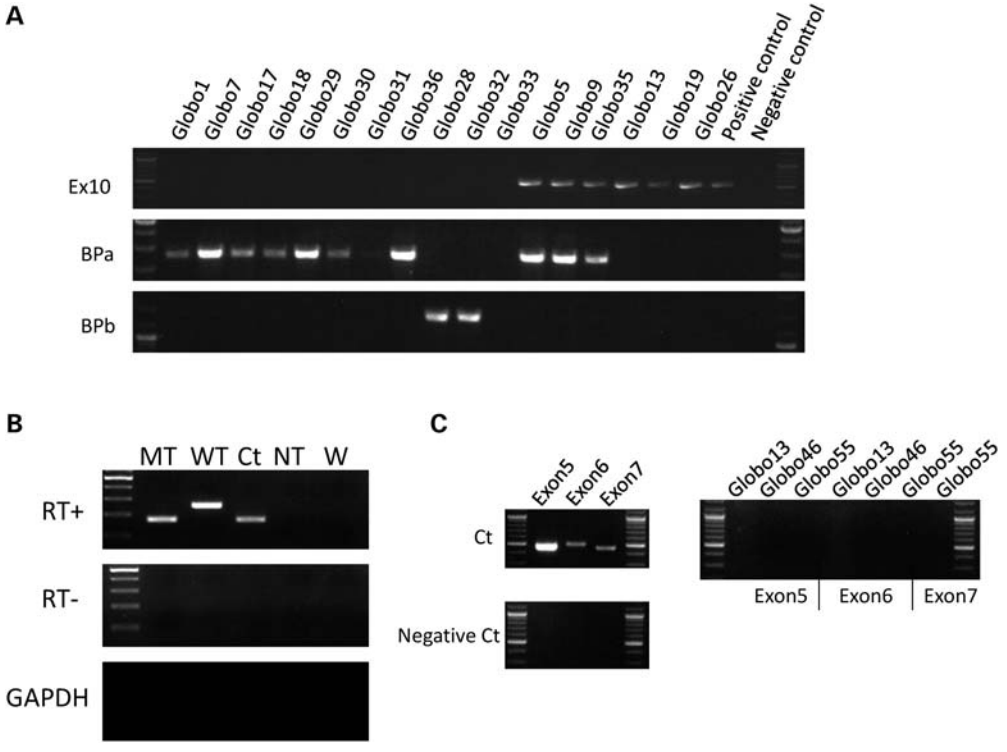

Figure 1. (**A**) PCR results of *DPY19L2* exon 10 and the BPs. Patients Globo1, Globo7, Globo17, Globo18, Globo29, Globo30, Globo31, Globo36, Globo28, Globo32, Globo33 are deleted for exon 10, whereas Gobo5, Gobo9, Gobo35, Gobo13, Gobo19, Gobo26 and the positive control showed an amplification of all of this exon. Globo1, Globo7, Globo17, Globo18, Globo29, Globo30, Globo31, Globo36, Globo28, Gobo5, Gobo9, Gobo35 showed an amplification for BP a, whereas Globo29 and Globo33 showed an amplification for BP b. For Globo39, 40, 42, 43, 45–48, 50, 51, 53, 54, 56–61, data are not shown. (**B**) Minigene constructs used to test the splicing of exon 11. Mt, mutant form of exon 11; WT, wild-type form of exon 11; Ct, construct without any cloned exon; NT, non-transfected cells; W, water. (**C**) PCR results for *DPY19L2* exons 5, 6 and 7. Globo13 and Globo46 are deleted for exons 5 and 6, whereas Globo55 is deleted for exons 5, 6 and 7. The positive control showed amplification of all these exons.

50, 53, 58, 60), we were not able to identify the BP. Five patients (Globo 35, 40, 42, 43, 46) showed an amplification of both exon 10 and the 1700 bp BP fragment, suggesting that these patients are heterozygous for the *DPY19L2* deletion (Fig. 1A). For the seven (Globo 34, 37, 38, 44, 49, 62, 63) remaining patients, exon 10 was detected, but no deletion could be found. Since we found one patient heterozygous for the deletion, we checked all remaining patients from our previous cohort. We identified two patients (Globo 5, 9) who were heterozygous for the *DPY19L2* deletion (Fig. 3A).

Point mutation screening

In a second step, in conjunction with screening for the *DPY19L2* deletion, we sequenced all coding exons and intron boundaries in the 7 compound heterozygous patients for the deletion and the 23 non-deleted patients (the exon localization of the mutations are shown in Fig. 2A). Concerning the heterozygous patients, sequence analysis of Globo5 revealed a variation in exon 8 (c.869G>A), leading to an amino acid change of a highly conserved residue (p.R290H) that was predicted to be deleterious by two programs: PolyPhen and SIFT (Supplementary Material, Table S1). Globo9 exhibited a variation in exon 9 (c.1033C>T), introducing a premature stop codon (p.Q345X). Globo35 presented a variation in exon 15 (c.1478C>G), leading to a non-synonymous mutation (p.T493R). This mutation is predicted to be deleterious by PolyPhen and tolerated by SIFT. Globo40 showed a variation in exon 21 (c.2038A>T), introducing a premature stop codon (p.K680X). Globo42 and 43, two unrelated patients, showed the same nucleotide deletion in exon 11 (c.1183delT), introducing a premature stop codon (p.S395LfsX7). Globo46 presented a deletion of exons 5 and 6. An aberrant splicing between exons 4 and 7 would give rise to a frame shift, introducing a premature stop codon (Supplementary Material, Table S1).

Among non-deleted patients, four, descending from consanguineous families, were found homozygous for a mutation. Globo26 showed a homozygous variation in exon 8 (c.892C>T), leading to a non-synonymous mutation (p.R298C), located at a strictly conserved amino acid position. This mutation is predicted to be deleterious by both PolyPhen and SIFT. Globo19 presented a homozygous donor splice-site mutation in intron 11 (c.1218+1G>A). Splice-site models (see Materials and Methods) predicted that the mutation disrupts the 5′ splice site of intron 11.

Unfortunately, the DPY19L2 protein presents a testis-restricted expression, and we were not allowed to use fresh sperm cells or to perform a testicular biopsy in these patients, in order to verify the predicted aberrant splicing *in vivo*. In order to test the prediction, minigene constructs were made, including either the wild-type (WT) or mutated form of exon 11

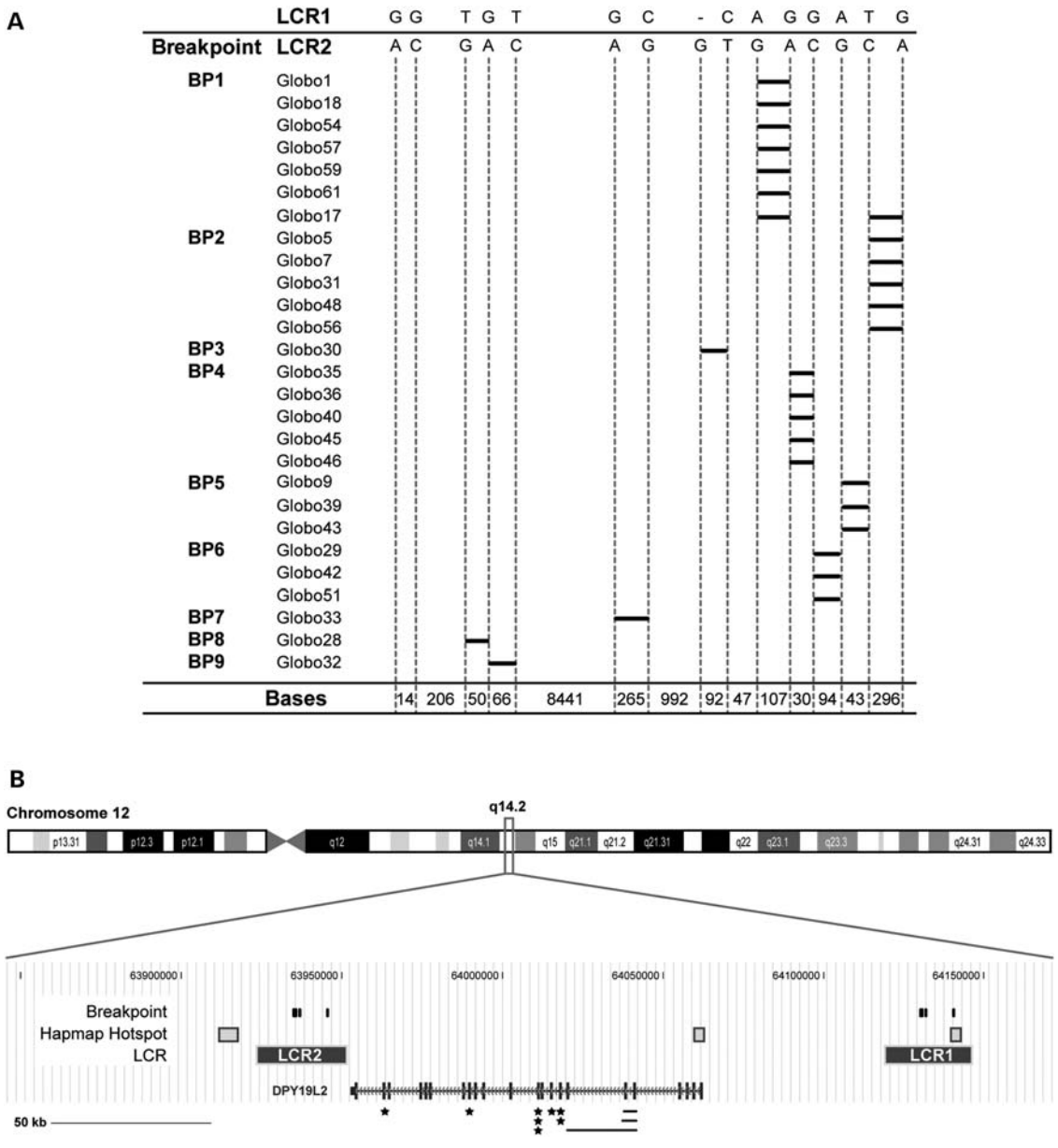

Figure 2. (**A**) For each LCR, the specific nucleotide differences within the hotspot regions are shown at the top. The length and distance between the different nucleotides are displayed at the bottom. The recombination region is represented for each patient by a black bar. The patients are ordered according to the nine BPs. (**B**) The *DPY19L2* locus is represented together with the two LCR, the HapMap recombination hotspots, the nine different BPs and the identified mutations (missenses are marked with a star and larger deletions are represented using a black horizontal line).

and flanking intronic sequences. These minigene constructs were transfected into COS1 and HeLa cells, and transcripts were analyzed by reverse transcription-PCR (RT-PCR) 48 h after the transfection. As shown in Figure 1B, WT exon 11 is invariably included in the final mRNA, as confirmed by the sequencing of the PCR product. In contrast, the mutated exon gives rise to one aberrant splicing form, leading to exon 11 skipping. An aberrant splicing of exon 11 could eventually give rise to a new splicing between exons 10 and 12 and would produce a protein with a deletion of 28 amino acids from position 378 to 406, corresponding to the seventh transmembrane helix (position 372–394).

We identified a homozygous deletion of exons 5 and 6 in Globo13. In order to pinpoint the two BP zones, subsequent amplifications on both sides of the exons were performed. Amplification across the deletion gave rise to a PCR fragment of 3500 bp for Globo13, whereas no amplification was possible from a control subject known to be fertile (Fig. 1C). Analysis of the sequence allowed the determination of the BPs and showed an insertion of 73 bp that corresponds to a part of a LINE sequence (Supplementary Material, Fig. S1). The deletion encompasses a 15.7 kb region with one BP located 8.35 kb from the 5′ side of exon 5 and the second one 4.36 kb from the 3′ side of exon 6. An aberrant splicing

A

Breakpoint zone	Point mutation	Patient	Origin
BP1	-	Globo1	Jordan
BP1	-	Globo18	France
BP1	-	Globo54	Iran
BP1	-	Globo57	Iran
BP1	-	Globo59	Iran
BP1	-	Globo61	Iran
BP2	p.R290H	Globo5	France
BP2	-	Globo7	Algeria
BP2	-	Globo31	Turkey
BP2	-	Globo48	Iran
BP2	-	Globo56	Iran
BP1/BP2	-	Globo17	France
BP3	-	Globo30	Saoudi Arabia
BP4	p.T493R	Globo35	Turkey
BP4	-	Globo36	Morocco
BP4	p.K680X	Globo40	Belgium
BP4	-	Globo45	Turkey
BP4	Ex5_6del	Globo46	USA
BP5	p.Q345X	Globo9	France
BP5	-	Globo39	Belgium
BP5	p.S395LfsX7	Globo43	Belgium
BP6	-	Globo29	Morocco
BP6	p.S395LfsX7	Globo42	Belgium
BP6	-	Globo51	Iran
BP7	-	Globo33	Turkey
BP8	-	Globo28	Morocco
BP9	-	Globo32	Turkey
-	Ex5_6del	Globo13	Iran
-	c.1218+1G>A	Globo19	Italy
-	p.R298C	Globo26	Morocco
-	Ex5_7del	Globo55	Iran
UBP	-	Globo47	Iran
UBP	-	Globo50	Iran
UBP	-	Globo53	Iran
UBP	-	Globo58	Iran
UBP	-	Globo60	Iran

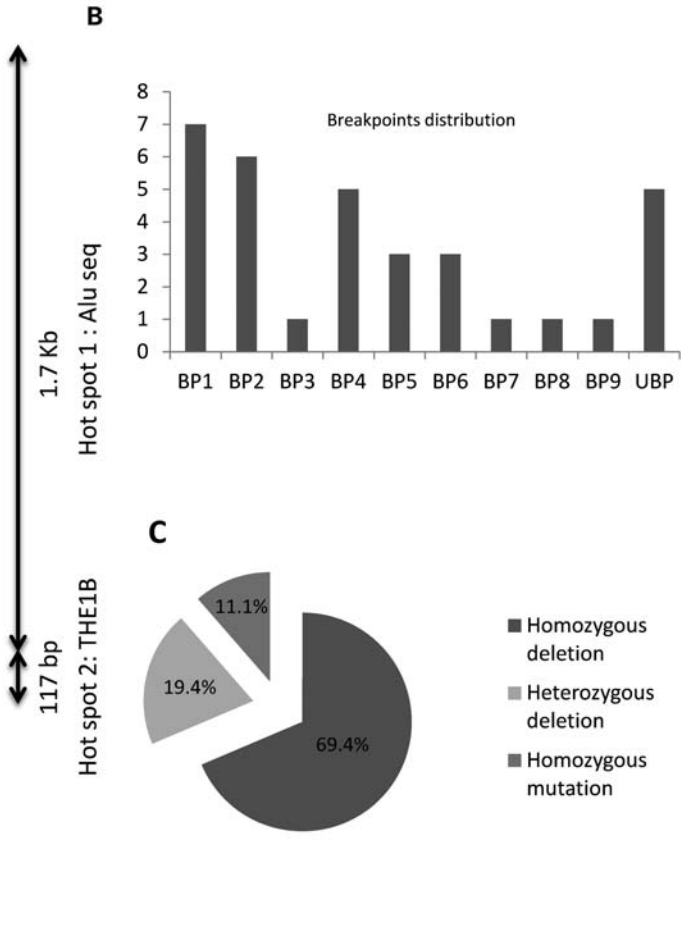

Figure 3. Distribution of BPs and mutations among patients. (**A**) The table presenting the type of BPs, mutations and patient origins. Vertical arrows delimit the two hotspots and their size. Hotspot 1 and hotspot 2 contain direct repeat elements, *Alu*Sq2 and THE1B, respectively. (**B**) The chart showing the distribution of the BPs among the patients totally deleted for *DPY19L2* (UBP, undetermined BP). (**C**) Pie chart showing the percentage of mutation types among the patients deleted and/or mutated for *DPY19L2*. Among these patients: 69.4% have a complete deletion of *DPY19L2* in the homozygous state; 19.4% have a complete deletion associated with a point mutation or a partial deletion; 11.1% have a point mutation or a partial deletion in the homozygous state.

between exons 4 and 7, if not degraded, would give rise to a frame shift, producing a truncated protein of 226 amino acids due to a premature stop codon at position 227. Since no repetitive sequence could be detected nearby, the deletion could be explained by a non-homologous end joining, which is often associated with the insertion of a DNA fragment at the BP (14). Globo46, who is compound heterozygous, also presented a deletion of exons 5 and 6, but with different BPs. Finally, Globo55 presented a deletion of exons 5, 6 and 7. An aberrant splicing between exons 4 and 8 would give rise to a protein with a deletion of 91 amino acids from position 196 to 287, corresponding to transmembrane domains 3, 4 and 5, at positions 195–215, 243–265 and 269–286, respectively (Supplementary Material, Table S1).

None of the variations described here were found, either in dbSNP v134 or when testing at least 188 control chromosomes.

BP analysis

We mapped linkage disequilibrium (LD) patterns and recombination rates in HapMap2 on chromosome 12. We observed that both LCR positions correlate with strongly suggested recombination spots, thus confirming the NAHR, which is the mechanism hypothesized to be responsible for the disease (26). Many other disorders involving NAHR between LCRs have been identified, such as DiGeorge syndrome/velocardiofacial syndrome, Williams–Beuren syndrome, Prader–Willi syndrome or Charcot–Marie–Tooth disease type 1A, which belong to the group of genomic disorders (27).

Previously, we identified two BPs (BP 1 and 2) localized close together in a region of 296 bp containing an *Alu* repetitive element (12). Sequence analysis of the amplification product across the deletions identified seven new BPs (BP3–9). Interestingly, seven of the BPs (BP1–7) are located in a small region of 1.7 kb, whereas the other two (BP8–9) are located in an area of 117 bp, 9.5 kb away from the first one (Fig. 2B). The underlying sequences of the BPs, as well as their genomic positions, are given in Supplementary Material, Figure S1. We can therefore suggest the presence of two recombination hotspots (hotspot 1 and 2, see Fig. 3A), comprising BP1–7 and BP8–9, respectively, both containing direct repeat elements (*Alu*Sq2 in hotspot 1, THE1B in hotspot 2). Interestingly, it was shown that AHR/NAHR hotspots

usually cluster within small regions of 1–2 kb of almost perfect identity, with repeat elements, transposons or minisatellites located nearby that act as substrates for double-strand breaks (22). Analysis of these hotspots in NAHR disorders identified a 13mer CCNCCNTNNCCNC motif frequently present in the flanking LCRs (22). We analyzed the sequences surrounding both identified recombination hotspots involved in the *DPY19L2* deletion and indeed found the PRDM9 13mer recognition motif (which is part of the THE1B repeat element), but only in hotspot 2 (Fig. 3A). Interestingly, hotspot 2 also coincides perfectly with a previously identified recombination hotspot (28) of chromosome 12 (Fig. 2B).

In total, we identified 32 patients with at least one deleted allele, among which BP1 occurs seven times, BP2 six times, BP4 five times, BP5 three times, BP6 three times and BP3, 7, 8, 9 were found only once (Fig. 3B). All recurrent BPs were found among patients originating from different regions, and patients from the same country showed different BPs (Fig. 3A). Out of 36 mutated patients, 69.4% were homozygous deleted, 19.4% heterozygous composite and 11.1% showed a homozygous point mutation (Fig. 3C).

DISCUSSION

The *DPY19L2* deletion was identified as being the major cause of globozoospermia in two different studies (12,21). It was shown by both groups that the most probable mechanism explaining this deletion is NAHR mediated by two LCRs surrounding *DPY19L2*. Thus, globozoospermia can be considered a new genomic disorder (27,29). Although both reports described a deletion rate of high frequency, a large difference was observed. Indeed, 19% (4 out of 21) of globozoospermic patients were found deleted in our study, whereas the other reported a frequency of 75% (15 out of 20). This difference could be due to a bias either in patient recruitment or in the limited number of patients analyzed. The higher frequency observed in Harbuz *et al*. (21) could be explained by the fact that most of the patients were Tunisians, whereas in our study, the patients were from seven different countries. Unfortunately, Harbuz *et al*. (21) were unable to localize the BPs, making it impossible to exclude a local founder effect in this cohort of globozoospermic patients. We present here the analysis of a larger cohort of globozoospermic patients, which allowed us, first, to refine the frequency rate of the deletion in our cohort, and, second, to enlarge the mutation spectrum. Indeed, in the context of this study and our previous work, we identified 24 homozygous deleted patients, 7 patients heterozygous for the deletion and presenting a point mutation in the remaining allele and 4 homozygous patients presenting a non-sense, a splicing mutation or deletions of exons 5 and 6 or exons 5 to 7 (Fig. 3A). In total, we identified nine BP zones, of which seven are clustered in a region of 1.7 kb and two in an area of 117 bp situated ∼9 kb away from the first spot of recombination. Both regions seem to define two ectopic recombination hotspots for *DPY19L2* deletion (Supplementary Material, Fig. S1), each being <2 kb (in LCRs of ∼27 kb) and containing a repeat element. Hotspot 1 accounts for almost all NAHR-driven *DPY19L2* deletions. In spite of its lower frequency, hotspot 2 contains the PRDM9 recognition binding motif, further supporting its involvement in the recombination mechanism (23,30). In addition, the fact that hotspot 2 coincides perfectly with a previously identified recombination hotspot supports the observation that NAHR and AHR share common features, including association with identical hotspot motifs.

We did not find any predominating BP. The fact that the same BPs are shared by patients from completely different regions and that patients from the same country can show different BPs tends to exclude any founder effect, even a recent one, and strongly suggests that the deletion results from recurrent events linked to the specific genomic architectural feature of this locus.

Our study allows us to calculate a more accurate prevalence of *DPY19L2* involvement in globozoospermia as a new genomic disorder. We analyzed a cohort of 54 globozoospermic patients and found that 36 of them were mutated for *DPY19L2* (66.7%). Despite the identification of subtle mutations, the most frequent alteration remains the deletion of the whole gene. Indeed, more than two-thirds (69%) of our patients are homozygously deleted, which also suggests that in one out of three of the cases, a search for point mutations is justified. We still have 18 patients with no identified mutation in *DPY19L2*, including 3 pairs of brothers, suggesting that new genes remain to be identified.

Since they do not present any other symptoms, globozoospermic patients are always recruited via *in vitro* fertilization centers. Considering the low frequencies of fertilization and birth obtained so far, one can wonder whether it is justified to propose that these couples go through the burden of such a heavy technology. A recent report, describing one globozoospermic patient, suggested that globozoospermia could be associated with the absence of phospholipase C zeta (PLCζ). This might explain the absence of fertilization, since PLCζ is the main physiological actor responsible for oocyte activation (31). An artificial activation, via a treatment with calcium ionophore, could improve fertilization rates and help to obtain embryos, thus increasing the chance of pregnancy (32). Such calcium ionophore activation treatment has been previously used successfully, considerably improving the fertilization and pregnancy rates in patients deficient for oocyte activation, including globozoospermic patients (33,34). It would be of great interest to see, in a larger cohort of globozoospermic patients, whether calcium ionophore activation would allow a better pregnancy rate and whether there is any correlation between the presence of a mutation in either *DPY19L2* or *SPATA16* and the pregnancy outcome. We are currently collecting DNA from globozoospermic patients for whom ICSI attempts have been done with or without oocyte activation.

MATERIALS AND METHODS

Patient recruitment and DNA preparation

Patients were selected by *in vitro* fertilization centers through a semen analysis implemented according to World Health Organization recommendations. Globozoospermia was diagnosed using a spermocytogram after a Harris–Shorr coloration.

Genomic DNA was extracted either from peripheral blood leukocytes using QIAamp DNA Blood Midi Kit (QIAGEN, Germany) or from saliva using Oragene DNA Self-Collection Kit (DNAgenotech, Ottawa, Canada), according to the manufacturer's instructions. This study was approved by the local Ethical Committee (Comité de protection de la personne, CPP) of Strasbourg University Hospital. For each case analyzed, informed written consent was obtained according to CPP recommendations.

PCR analysis

PCR covering all the exons of DPY19L2 and their exon–intron boundaries or the previously identified deletion (previously identified BPs BP1 and BP2) was performed using genomic DNA, and amplicons were sequenced by GATC (Konstanz, Germany). For the newly identified deletions, 5′ and 3′ walks were then carried out to identify the BPs. All primer sequences and PCR conditions are available in Supplementary Material, Table S2. Because of the high conservation level of duplicated regions, special care was taken to choose specific oligonucleotides with a unique sequence specifying a single location.

Plasmids and transfection

To study the effect of the c.1218+1G>A splice mutation, we constructed WT and mutated hybrid minigenes, using the pDUP33 vector (35). The genomic DNA region from Globo26 or from a control containing exon 11 (87 bp) and intronic flanking sequences (654 bp upstream from the 5′ exon end and 329 bp downstream from the 3′ exon end) were PCR-amplified using a forward primer (DPY19L2e11BamHI; 5′-GGGCCCtacatatggtattggtgatatcca-3′) and a reverse primer (DPY19L2e11BglII; 5′-AGATCTcactgcaatgagtacttaacc-3′). The 1063 bp PCR products were purified from agarose gel, using the Gel Band Purification Kit (Amersham Biosciences), and were sequentially digested by *Bam*HI and *Bgl*II. The insert was directionally cloned into the first intron of β-globin in the Dup33 plasmid. Recombinant plasmids were sequenced to confirm the presence of the mutated or the WT sequence. Transfection of plasmid DNA (WT, mutant and control pDUP33) in Cos and HeLa cells and RT-PCR were performed as previously described (11).

Bioinformatic analysis

Single-nucleotide variations (SNVs) were analyzed using the Alamut software (Interactive BioSoftware), which systematically gives access to several prediction algorithms. In particular, missense variations were tested with SIFT (36) and PolyPhen v2 (37). Missense and intronic variations were analyzed for splicing effects, using MaxEntScan (38), NNSPLICE (39) and HSF (40).

Allelic crossover hotspots [build 37 (36)], recombination rates from HapMap Phase II [build 36 (28)] and LD patterns from the 1000 Genomes project [build 37 (41)] were mapped onto the *DPY19L2* locus using the UCSC browser (42).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We are very grateful to James R Lupski, Claudia Carvalho and our colleagues Jean-Louis Mandel and Julie Thompson for their critical reading of this manuscript. We are also grateful for the services of the Institute of Genetics and Molecular and Cellular Biology (IGBMC).

Conflict of Interest statement. None declared.

AUTHORS' ROLES

Contributors E.E., P.K. and S.V. designed the study; E.E., P.K., S.J., J.M. and I.K. performed genetic analysis and mutation screening; F.V.M., M.H.N.-E., A.D., T.G., N.L., N.I., M.B., F.C.P., H.G., D.D., F.B., S.A.G., J.-M.G., S.C.O., P.D. recruited patients and collected clinical data; C.R. and J.M. contributed to the bioinformatics analysis; E.E., P.K., C.R., J.M. and S.V. contributed to analysis and interpretation of data; E.E., C.R., J.M. and S.V. wrote the manuscript.

FUNDING

This work was supported by the French Centre National de la Recherche Scientifique (CNRS), Institut National de la Santé et de la Recherche Médicale (INSERM), the Ministère de l'Education Nationale, the Enseignement Supérieur et de la Recherche, the University of Strasbourg, the University Hospital of Strasbourg and the Agence de la BioMédecine.

REFERENCES

1. Ikawa, M., Inoue, N., Benham, A.M. and Okabe, M. (2010) Fertilization: a sperm's journey to and interaction with the oocyte. *J. Clin. Invest.*, **120**, 984–994.
2. Carrell, D.T., Emery, B.R. and Liu, L. (1999) Characterization of aneuploidy rates, protamine levels, ultrastructure, and functional ability of round-headed sperm from two siblings and implications for intracytoplasmic sperm injection. *Fertil. Steril.*, **71**, 511–516.
3. Machev, N., Gosset, P. and Viville, S. (2005) Chromosome abnormalities in sperm from infertile men with normal somatic karyotypes: teratozoospermia. *Cytogenet. Genome Res.*, **111**, 352–357.
4. Viville, S., Mollard, R., Bach, M.L., Falquet, C., Gerlinger, P. and Warter, S. (2000) Do morphological anomalies reflect chromosomal aneuploidies? Case report. *Hum. Reprod.*, **15**, 2563–2566.
5. Banker, M.R., Patel, P.M., Joshi, B.V., Shah, P.B. and Goyal, R. (2009) Successful pregnancies and a live birth after intracytoplasmic sperm injection in globozoospermia. *J. Hum. Reprod. Sci.*, **2**, 81–82.
6. Bechoua, S., Chiron, A., Delcleve-Paulhac, S., Sagot, P. and Jimenez, C. (2009) Fertilisation and pregnancy outcome after ICSI in globozoospermic patients without assisted oocyte activation. *Andrologia*, **41**, 55–58.
7. Coetzee, K., Windt, M.L., Menkveld, R., Kruger, T.F. and Kitshoff, M. (2001) An intracytoplasmic sperm injection pregnancy with a globozoospermic male. *J. Assist. Reprod. Genet.*, **18**, 311–313.
8. Dirican, E.K., Isik, A., Vicdan, K., Sozen, E. and Suludere, Z. (2008) Clinical pregnancies and livebirths achieved by intracytoplasmic injection of round headed acrosomeless spermatozoa with and without oocyte activation in familial globozoospermia: case report. *Asian J. Androl.*, **10**, 332–336.

9. Kilani, Z., Ismail, R., Ghunaim, S., Mohamed, H., Hughes, D., Brewis, I. and Barratt, C.L. (2004) Evaluation and treatment of familial globozoospermia in five brothers. *Fertil. Steril.*, **82**, 1436–1439.
10. Sermondade, N., Hafhouf, E., Dupont, C., Bechoua, S., Palacios, C., Eustache, F., Poncelet, C., Benzacken, B., Levy, R. and Sifer, C. (2011) Successful childbirth after intracytoplasmic morphologically selected sperm injection without assisted oocyte activation in a patient with globozoospermia. *Hum. Reprod.*, **26**, 2944–2949.
11. Dam, A.H., Koscinski, I., Kremer, J.A., 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.
12. Koscinski, I., Elinati, E., Fossard, C., Redin, C., Muller, J., Velez de la Calle, J., Schmitt, F., Ben Khelifa, M., Ray, P.F., Kilani, Z. *et al.* (2011) DPY19L2 deletion as a major cause of globozoospermia. *Am. J. Hum. Genet.*, **88**, 344–350.
13. Stankiewicz, P. and Lupski, J.R. (2002) Genome architecture, rearrangements and genomic disorders. *Trends Genet.*, **18**, 74–82.
14. Gu, W., Zhang, F. and Lupski, J.R. (2008) Mechanisms for human genomic rearrangements. *Pathogenetics*, **1**, 4.
15. Inoue, K. and Lupski, J.R. (2002) Molecular mechanisms for genomic disorders. *Annu. Rev. Genomics Hum. Genet.*, **3**, 199–242.
16. Lupski, J.R. (2004) Hotspots of homologous recombination in the human genome: not all homologous sequences are equal. *Genome Biol.*, **5**, 242.
17. Liu, P., Lacaria, M., Zhang, F., Withers, M., Hastings, P.J. and Lupski, J.R. (2011) Frequency of nonallelic homologous recombination is correlated with length of homology: evidence that ectopic synapsis precedes ectopic crossing-over. *Am. J. Hum. Genet.*, **89**, 580–588.
18. Blanco, P., Shlumukova, M., Sargent, C.A., Jobling, M.A., Affara, N. and Hurles, M.E. (2000) Divergent outcomes of intrachromosomal recombination on the human Y chromosome: male infertility and recurrent polymorphism. *J. Med. Genet.*, **37**, 752–758.
19. McLachlan, R.I. and O'Bryan, M.K. (2010) Clinical review: state of the art for genetic testing of infertile men. *J. Clin. Endocrinol. Metab.*, **95**, 1013–1024.
20. Carvalho, C.M., Zhang, F. and Lupski, J.R. (2011) Structural variation of the human genome: mechanisms, assays, and role in male infertility. *Syst. Biol. Reprod. Med.*, **57**, 3–16.
21. 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.
22. Myers, S., Freeman, C., Auton, A., Donnelly, P. and McVean, G. (2008) A common sequence motif associated with recombination hot spots and genome instability in humans. *Nat. Genet.*, **40**, 1124–1129.
23. Baudat, F., Buard, J., Grey, C., Fledel-Alon, A., Ober, C., Przeworski, M., Coop, G. and de Massy, B. (2011) PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science*, **327**, 836–840.
24. McVean, G. and Myers, S. (2010) PRDM9 marks the spot. *Nat. Genet.*, **42**, 821–822.
25. Myers, S., Bowden, R., Tumian, A., Bontrop, R.E., Freeman, C., MacFie, T.S., McVean, G. and Donnelly, P. (2010) Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science*, **327**, 876–879.
26. Lindsay, S.J., Khajavi, M., Lupski, J.R. and Hurles, M.E. (2006) A chromosomal rearrangement hotspot can be identified from population genetic variation and is coincident with a hotspot for allelic recombination. *Am. J. Hum. Genet.*, **79**, 890–902.
27. Lupski, J.R. (2009) Genomic disorders ten years on. *Genome Med.*, **1**, 42.
28. International HapMap Consortium (2003) The International HapMap Project. *Nature*, **426**, 789–796.
29. Lupski, J.R. (1998) Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet.*, **14**, 417–422.
30. Berg, I.L., Neumann, R., Lam, K.W., Sarbajna, S., Odenthal-Hesse, L., May, C.A. and Jeffreys, A.J. (2011) PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. *Nat. Genet.*, **42**, 859–863.
31. Kashir, J., Heindryckx, B., Jones, C., De Sutter, P., Parrington, J. and Coward, K. (2010) Oocyte activation, phospholipase C zeta and human infertility. *Hum. Reprod. Update*, **16**, 690–703.
32. Taylor, S.L., Yoon, S.Y., Morshedi, M.S., Lacey, D.R., Jellerette, T., Fissore, R.A. and Oehninger, S. (2010) Complete globozoospermia associated with PLCzeta deficiency treated with calcium ionophore and ICSI results in pregnancy. *Reprod. Biomed. Online*, **20**, 559–564.
33. Heindryckx, B., De Gheselle, S., Gerris, J., Dhont, M. and De Sutter, P. (2008) Efficiency of assisted oocyte activation as a solution for failed intracytoplasmic sperm injection. *Reprod. Biomed. Online*, **17**, 662–668.
34. Heindryckx, B., Van der Elst, J., De Sutter, P. and Dhont, M. (2005) Treatment option for sperm- or oocyte-related fertilization failure: assisted oocyte activation following diagnostic heterologous ICSI. *Hum. Reprod.*, **20**, 2237–2241.
35. Dominski, Z. and Kole, R. (1992) Cooperation of pre-mRNA sequence elements in splice site selection. *Mol. Cell. Biol.*, **12**, 2108–2114.
36. Kumar, P., Henikoff, S. and Ng, P.C. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.*, **4**, 1073–1081.
37. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. and Sunyaev, S.R. (2010) A method and server for predicting damaging missense mutations. *Nat. Methods*, **7**, 248–249.
38. Yeo, G. and Burge, C.B. (2004) Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.*, **11**, 377–394.
39. Reese, M.G., Eeckman, F.H., Kulp, D. and Haussler, D. (1997) Improved splice site detection in Genie. *J. Comput. Biol.*, **4**, 311–323.
40. Desmet, F.O., Hamroun, D., Lalande, M., Collod-Beroud, G., Claustres, M. and Beroud, C. (2009) Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.*, **37**, e67.
41. 1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. *Nature*, **467**, 1061–1073.
42. Karolchik, D., Hinrichs, A.S. and Kent, W.J. (2011) The UCSC Genome Browser. *Curr. Protoc. Hum. Genet.*, **18**, 18.6.1–18.6.33.

Chapter VIII

Confirmation of the pathogenicity of *SPATA16* mutations and identification of its partners

1- Introduction

In 2007, we identified *SPATA16* (Spermatogenesis associated 16) as the first autosomal gene involved in human globozoospermia in three affected brothers of an Ashkenazi Jewish family. We demonstrated that a homozygous mutation (c.848G→A) in *SPATA16* led to the production of roundheaded and acrosomless spermatozoa, by disrupting the splicing site of exon 4 coding for a highly conserved tetratricopeptide repeat domain (TPR domain). The TPR domain is a protein-protein interaction module that acts as an organizing centre for complexes regulating a multitude of biological processes (Krachler, Sharma et al. 2010). The encoded protein, specifically expressed in human testis, is localized in the Golgi apparatus and then shifted with Golgi vesicules to the acrosome. Here, we report another mutation of *SPATA16* and we present a study to select genes potentially involved in human globozoospermia using two approaches: i) the determination of Spata16 partners by GST pulldown and ii) the candidate gene approach.

Our analysis of 54 globozoospermic patients originating from different countries showed that 36 patients are mutated for *DPY19L2*, which represents 67% of our cohort. However, for the 18 remaining patients, we did not find any causative mutation, indicating that acrosome biogenesis is a complex process that implicates many genes. In order to identify the gene responsible for globozoospermia in these patients, we initiated our study by screening for *SPATA16* and *PICK1* mutations, since both have been shown to be implicated in human acrosome biogenesis. Using the candidate gene approach, we selected two additional genes known as *AGFG1 (*ArfGAP with FG repeats 1) and *GOPC* (Golgi-associated PDZ- and coiled-coil motif-containing protein) because phenotypes in the null mutant mice are very similar to human globozoospermia: round, spherical and acrosomless sperm (Christensen, Ivanov et al. 2006; Xiao, Kam et al. 2009). For instance, in *Agfg1*-/- mice, proacrosomic vesicules form but fail to fuse (Kang-Decker, Mantchev et al. 2001). A lack of acrosome formation was observed in *Gopc* null mutant mice due to the fragmentation of the acrosome in the early stage of spermiogenesis (Yao, Ito et al. 2002). The same phenotype was also observed in male mice deficient in Pick1, a protein colocalized with GOPC in the golgi regions (Xiao, Kam et al. 2009). Taking into consideration the mice phenotype in case of failure of *Agfg1*, *Gopc*, *Pick1*, as well as the human cases reported implicating *SPATA16* and *PICK1*(Dam, Koscinski et al. 2007; Liu, Shi et al. 2010), we screened the entire coding region

and intron/exon boundaries of the 4 genes cited above in 18 globozoospermic patients for whom no mutation in *DPY19L2* was found.

2- Materials and Methods

Patients and controls

A total of 18 patients originating from France, Italy, Tunisia, Turkey, Libya and Morocco were recruited for this study. All the patients were healthy and represent nonsyndromic infertility. None of them had chromosomal abnormalities or were positive for Y microdeletions. Any exposure to toxic chemicals, radiation or treatment by drugs known to impair fertility was reported. Patients were identified at the University Hospital of Strasbourg or provided by collaborators. This study was approved by the local Ethical Committee (Comité de protection de la personne, CPP) of Strasbourg University Hospital. For each case analysed, informed written consent was obtained according to the CPP recommendations.

DNA preparation

Genomic DNA from each patient was extracted either from peripheral blood leucocytes using QIAmp DNA Blood Midi Kit (Qiagen, Germany) or from saliva using Oragene DNA Self-Collection Kit (DNAgenotech, Canada).

Linkage analysis

The whole genome scan was performed on 4 patients using the GeneChip Human mapping 10 K 2.0 Xba from Affymetrix. Single nucleotide polymorphisms (SNP) obtained were analysed with the HomoSNP program (plewniak@igbmc.u-strasbourg.fr) that was set to identify homozygous regions of 25 or more consecutive SNP.

PCR reactions and DNA sequencing

Mutation screening was performed by polymerase chain reaction (PCR) and direct sequencing of the exons and intronic flanking sequences of the 4 genes. PCR primer sequences and amplification conditions are listed in table 2. 30 cycles of PCR reactions were carried out in the T-Gradient PCR system (Biometra) in a volume of 25 µL using Taq DNA polymerase (Roche). DNA amplicons were purified and double strand sequencing of each DNA fragment were performed with primers used for the PCR amplification. Sequence analyses were carried out using an ABI PRISM 3730xl Genetic Analyser (Applied Biosystems, GATC, Germany).

Cloning SPATA16 into a pGex expression vector

The fusion protein GST-SPATA16 was constructed by inserting the cDNA encoding the full open reading frame of mouse SPATA16 in the multiple cloning site of the pGex-4T-1.
The murine cDNA was ordered from Biovalley (MMM1013-99826784) and amplified using a forward primer that contained the EcoRI (5'- GAATTCgattctggcaagagtaggag -3') site and a reverse primer that contained a flag encoding sequence in green, a stop codon in red and XhoI site (5'- caaactgtacagcaaaac**gattacaaggacgacgatgacaag**tgaccgCTCGAGcgg -3'). The FLAG tag sequence was added to the C terminus of SPATA16 in order to check the complete translated protein. The 1755bp PCR products were purified from the agarose gel by use the Gel Band Purification Kit (Amersham Biosciences) and were sequentially digested with EcoRI and XhoI. The insert was directionally cloned into the EcoRI and XhoI sites of the pGex-4T-1 vector. The recombinant plasmids were amplified in XL1blue competent cells and purified by use of the Minipreps DNA Purification System (Promega). The correct sequence was confirmed by sequencing with primers in the plasmid (pGEX3-RP:TCAAGAATTATACACTCCG and pGEX5-FP: AACGTATTGAAGCTATCCC).

Purification of SPATA16

The recombinant protein GST–SPATA16 and protein GST were individually expressed in *E. coli* BL21 cell. The induction by 0.1mM of IPTG was performed at a cell density of A_{600} = 0.5 for 3 hours at 22°C tested by 10% SDS-PAGE. Cleared lysates were prepared and the soluble fusion proteins were purified by batch procedure incubating the lysate with

Glutathione Sepharose (Amersham Biosciences) for 1 hour at 4°C with gentle shaking. Then the Glutathione Sepharose was collected by centrifugation and washed five times with lysis buffer (Tris HCl 50mM pH9, NaCl 500mM, dTT 2mM). After boiling, the bound proteins were analyzed by 10% SDSPAGE followed by Western blotting using anti-GST monoclonal antibody (Kangwei, China).

Protein extraction

Forty testes were collected from wild type adult mice. Protein are then extracted by adding a lysis buffer (50 mM Tris HCl pH8, 150 mM NaCl, 1 mM EGTA, 1 mM $MgCl_2$, 0.1% NP40, 1 mM PMSF, 1 mM DTT and protease inhibitor cocktail) and by mixing for 2 min on ice. The soluble fraction is then collected after centrifugation at 8000 rpm for 10 min at 4°C.

GST pull-down assay and mass spectrometry

The pull down was performed on Glutathione Sepharose coupled to GST and GST-SPATA16 separately. The soluble fraction of the testes proteins are incubated with the Glutathione Sepharose at 4°C with gentle shaking for overnight. Then the beads are washed 5 times with the lysis buffer before analyzing by SDS-Page. Mass spectrometry was performed at Taplin mass spectrometry in Harvard after running the samples about 2-3cm into the resolving gel (10% acrylamide) and staining with coomassie blue. The gel is then cut and the bands were sent in 1.5mL tubes with mQ water to Taplin.

Production of Rabbit Polyclonal antibody

A rabbit polyclonal antibody was generated at the IGBMC and directed against the mouse recombinant protein GST-SPATA16. Two rabbits were immunizated by injecting 300 micrograms of protein per rabbit, at concentration of about 1 mg/ml: 200 µg was used for a first injection, 100 µg for the boost. 1 ml of antigen solution (containing 200 µg of protein) was mixed with 1 ml of “Complete Freund Adjuvant” to produce 2 ml emulsion. Emulsion is injected intradermically on 40 to 60 different locations. One month after immunization, blood sample (30 ml) is collected each week for a one month period. Blood is centrifuged at 3000 rpm, 20 min, at room temperature on Jouan CR411, and antiserum (about 10 ml) is harvested

and stored at –20°C. The boost is performed by sub-cutaneous injection of 2 ml emulsion formed by mixing 1 ml antigen solution with 1 ml of "Incomplete Freund Adjuvant". Twelve days later, the rabbit is anesthetized and then sacrificed. Blood (130-150 ml) is collected by intracardiac puncture, and then centrifuged to generate about 70-80 ml of antisera.

Western Blot analysis

The protein extracts were boiled in loading buffer (1M Tris pH 6.8, 8% SDS, 1 ml glycerol; Betamercaptoéthanol; 0.004% bromophenol blue) and resolved on 10% SDS-Page. In order to make the proteins accessible to the detection by the antibody, they are blotted onto a nitrocellulose membrane (Schleicher & Schuell). After transfer, the membrane nitrocellulose is blocked in a solution PBS1X (0.05% Triton) with 5% milk for 1h at room temperature before being incubated overnight with primary antibody directed against SPATA16 diluted in a solution containing 1% chicken albumin + PBS1X 0.05% Tween-20. Antibodies against SPATA16 are diluted to 1:10000. The membrane is then washed three times for 5 min in PBS1X and then incubated with goat secondary antibody directed against rabbit IgG (GAR: Goat anti-rabbit IgG) diluted to 1/20000 in 5% milk in PBS 1X 0.1% Tween-20.

The membrane was washed 3 times for 5 min each and incubated after with anti-IgG secondary antibody rabbit coupled to horseradish peroxidase (horseradish peroxidase) for 2 hours at room temperature. The HRP catalyzes the conversion of a substrate highly sensitive chemiluminescent (SuperSignal West Pico Chemiluminescent substrate, Pierce) chemiluminescent molecules, thereby producing light. The detection is carried out in using autoradiographic film.

Immunofluresence

After blocking of nonspecific binding sites in 4% fetal calf serum (Sigma-Aldrich; A9647) in PBS1X 0.1% tritron for 1 h at room temperature, deparaffinized testis sections (5 µm thick) from adult wild type mice were incubated with the primary antibody, that is, anti-SPATA16 (dilution 1:500) followed by anti-rabbit Alexa Fluor conjugated (Invitrogen; #A11011, #A11008) secondary antibodies. A nucleus counterstaining to determine nuclear shape and chromatin condensation was carried out with DAPI (dilution 1:20000). In samples used as control, primary antibodies were omitted. Testis sections were examined on a Leica DMRB

microscope equipped with standard filter sets for green (Alexa 488), red (Alexa 568) and blue (DAPI) fluorescence. Fluorescence images were captured with a Leica DCF480 charge-coupled device camera (Leica Microsystems) by using Imaging software (Microsoft Corporation) and were elaborated with Adobe Photoshop (Mountain View).

3- Results

Mutation screening of *SPATA16, AGFG1, GOPC* and *PICK1*

In our previous work, 18 patients did not show any mutation for *DPY19L2*. Among these patients, three consanguineous families (A, B and C) that each contained two globozoospermic brothers (Globo10, Globo20 and Globo23), have been analysed by 10 K SNP microarrays. Arrays with a higher resolution were not available at this time. A wide genome scan was also applied on a Tunisian globozoospermic patient (Globo65), as requested by a collaborator. A homozygous region was observed at the locus of *SPATA16* in Globo65, whereas Globo10, Globo20 and Globo23 were heterozygous. Therefore, we performed PCR of all the exons and exon-intron junctions of *SPATA16* on the genomic DNA of Globo65, which showed a homozygous deletion of exon 2, since specific amplification was detected in the control but never in the patient **(Figure 25)**. In order to pinpoint the breakpoints within intron 1 and intron 2, subsequent amplifications on both sides of exon 2 were performed.

Amplification across the deletion using the primers, 5'SP16n7 as forward primer and 3'SP16n4 as reverse primer, was performed as a control of the deletion **(Figure 25)**. Globo65 showed an amplification of 2000 bp, but this was not the case for the fertile control. Sequencing the 2000 bp PCR product allowed us to precisely define the breakpoint, which is located at 20.5 Kb at 5' side and at 1.5 kb at the 3' side of the exon, and hence to determine the exact size of the deletion which is 22.6 kb.

We pursued our study by screening *SPATA16* mutations in the 14 remained patients who were not analysed by SNP microarray. No variation was detected in any of these patients. Furthermore, no mutation was identified in *AGFG1*, *GOPC* and *PICK1* in the 17 remaining patients.

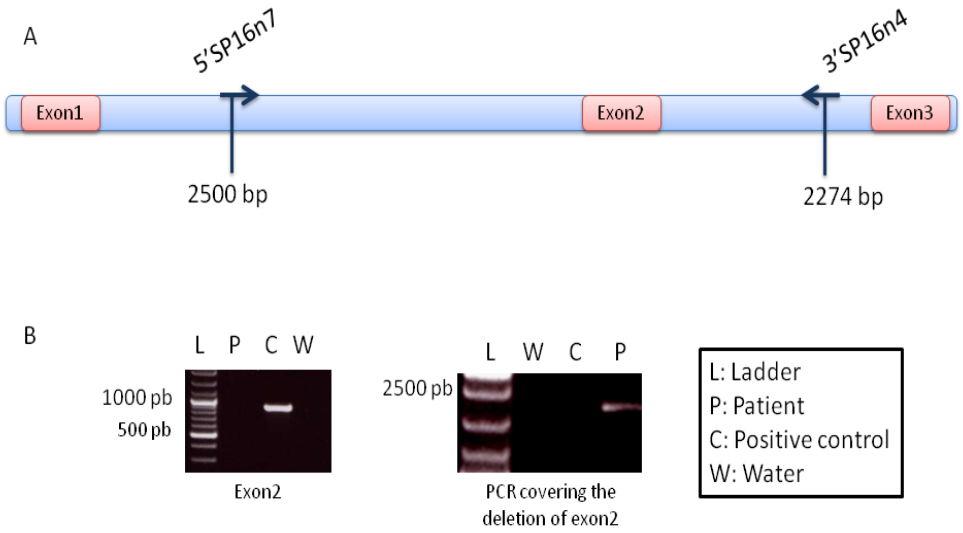

Figure 25 : Analysis of *SPATA16* gene. (A) Schematic representation of exons 1, 2 and 3 of *SPATA16.* Sequence amplifications on both sides of exon 2 were performed using the primers 5'SPA16n7 at 2500 bp from exon1 and 3'SP16n4 at 2274bp from exon2 in order to determine the breakpoints of the deletion. (B) PCR results of *SPATA16* exon2 and breakpoints. The patient is deleted for exon2 while the positive control showed amplification. The PCR covering the deletion of exon2 was performed using the primers 5'SPA16n7 and 3'SP16n4. The deleted patient showed a fragment of 2000bp while the positive control was homozygous wild type since no fragment was amplified.

Identification of Spata16 partners and selection of new candidate genes

Since Spata16 contained a highly conserved TPR domain, known to be a center of protein interactions especially co-chaperone proteins, we wanted to investigate the partners with which Spata16 interacts. Consequently, we applied the strategy of GST pulldown. At this time, we could not use co-immunoprecipitation, since we did not have a specific antibody against the protein. Thus, producing the fusion protein GST-Spata16 allowed us, on the first hand to perform the pulldown, and on the second hand to synthesis polyclonal antibodies by immunizing rabbits with the purified protein. The fusion protein was expressed in *E. coli* BL21 and then purified using a batch to Glutathione beads. The purified protein was then incubated with an extract of wild type mice testis proteins. In order to eliminate the false positive proteins, we analyzed the proteins that also interact with GST alone. The two samples were later analyzed by 10% SDS-PAGE and showed specific bands in the GST-Spata16 sample which were not detected in the negative control. We were able to identify the potential partners of Spata16 by mass spectrometry. The proteins shared between the sample and the negative control were omitted, leading to 35 potential partners **(Table 1)**.

Table1: Spata16 partners identified by GST pulldown. The partners presented below are selected after eliminating the proteins identified in the negative control. Mouse proteins highlighted in green showed a predominant expression in testes.

Spata16 partners		
Col6a3	Nid1	Irgc1
Hspa8	Tcp1	Hspa4l
Tubb3	Clpx	Dnaja1
Tubb5	Actbl2	Ldhc
Myh11	Hnrnph1	Dnaja4
Col6a1	Spata16	Kdm5b
Myh9	Cltc	
Myh10	Gm5409	
Lamb2	Slc2a3	
Rcn2	Eef1a1	
Col6a2	Axin2	
Atp5a1	Pcbp1	
Eef1a2	Grpel1	
Sucla2	Myl6b	
Dnaja2		

The identified partners could be considered as potential candidate genes involved in human globozoospermia, especially since five had shown an exclusive expression in human testes **(Table 1)**. Subsequently, we decided to sequence the candidates in the 17 remaining globozoospermic patients who were not mutated for *DPY19L2* or *SPATA16*, hoping to identify new genes implicated in human globozoospermia.

We started the study by combining the linkage analysis of the three consanguineous globozoospermic families and the GST pulldown outcome. Analysis of the three families as one entity, assuming that the same gene is responsible for the phenotype observed in each family, revealed a unique homozygous region located on chromosome 8. The region of 4.7 Mb contains 5 genes, but none was identified as a partner of Spata16. Examining the families separately, three genes were chosen among the 35 partners: *IRGC*, *DNAJA4* and *TCP1*. In fact, brothers of family A showed homozygosity for *IRGC* and *DNAJA4* loci, while brothers of family B were homozygous for the *TCP1* locus. Amplification and sequencing of the coding sequence of each gene did not reveal any nucleotide variation in the corresponding patients.

Spata16 expressed in the testis and localized in murine acrosome

Western blot analysis performed on the protein extract of adult mice testis, using a rabbit polyclonal antibody against murine Spata16, confirms the expression of Spata16 in the testis. Immunofluoresence performed on mice testis sections showed that Spata16 is located in the developing acrosome of the germ cells, as depicted in **Figure 26**.

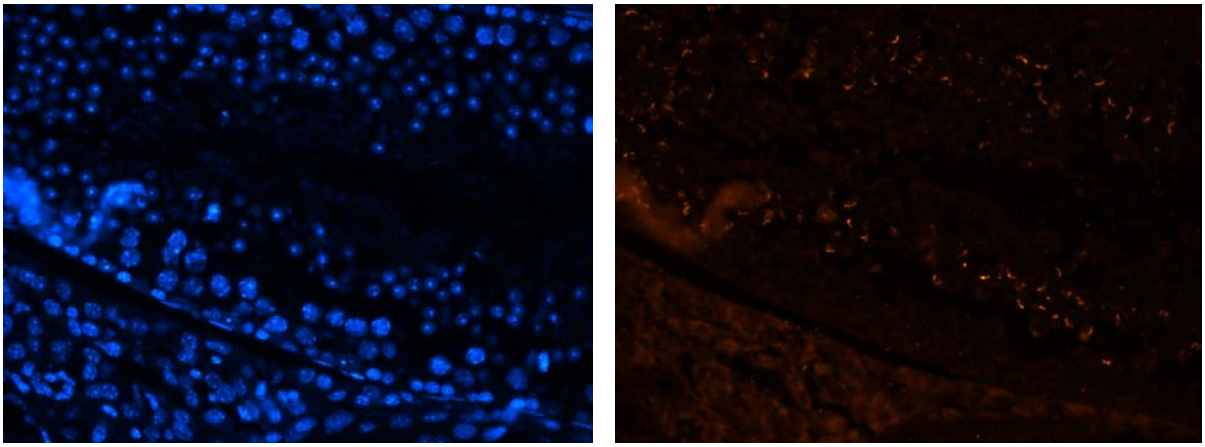

Figure 26 : Spata16 a component of the acrosome. Wild Type adult mouse testes sections stained with DAPI (Blue, on the left) and SPATA16 (Red, on the right). Spata16 appears to be localized in acrosomes.

4- Discussion

Even if some genetic causes of male infertility have been identified, the genetics of infertility remain largely an unexplored field. Here, we presented the second globozoospermic case due to a variation of the encoding sequence of *SPATA16.* Hence, we confirmed the pathogenicity of *SPATA16* mutations by identifying a deletion of 22.6 Kb encompassing exon 2 and we presented two approaches to select genes potentially involved in human globozoospermia: i) candidate gene approach and ii) identification of Spata16 partners. The deletion of exon 2 was identified in one of the 18 patients not mutated for *DPY19L2.* The molecular mechanism that leads to the synthesis of a non-functional protein has not been studied, but we hypothesise that the absence of exon 2, which is the first coding exon (exon 1 is non-coding) and contains the start codon ATG, lead to the absence of the protein or to a truncated protein. Alternatively, an eventual splice between exon 1 and 3 would lead to the production of an mRNA containing

three ATG codons in the exon 3. None of these codons are likely to be used since they do not present a Kozac sequence. The first and third are in frame with each other but are not in frame with the normal protein and would introduce three premature termination codons (PTC) 88; 31; 10 nucleotides 5' to the next 3-4 exon-exon junction for the first ATG and a PTC 75 nucleotides 5' to the next 4-5 exon-exon junction. In such a case, it is likely that the exon 2 deleted Spata16 mRNA is eliminated by a nonsense-mediated mRNA decay (NMD) mechanism. The second ATG is in frame with the *SPATA16* sequence and if used, it would produce a truncated protein with 36 amino acids missing out of 116 (31%) of the highly conserved TPR domain, and thus probably producing a non-functional protein.

Using the candidate genes approach, three genes described from site-directed mutagenesis experiments in animals were of interest: *Agfg1*, *Gopc* and *Pick1*, since mice deficient for these proteins exhibit a phenotype very similar to human globozoospermia. The human *AGFG1* gene is located on chromosome 2 at position q36.3 and comprises 13 coding exons. In wild-type mice, *Agfg1* associates with the cytosolic surface of proacrosomic transport vesicles that fuse to create a single large acrosomic vesicle in step 3 of mouse spermiogenesis. Whereas in mutant mice, proacrosomic vesicles form, but fail to fuse, blocking acrosome development (Kang-Decker, Mantchev et al. 2001). In addition to many sperm in knockout mice having a classic globozoospermic appearance, a report also shows that *Agfg1*-/- mice have multiple centriolar anomalies, leading to deformed or multiple flagella (Juneja and van Deursen 2005). The *GOPC* gene, also commonly identified as a PDZ domain protein interacting specifically with TC10, is located at 6q21 and contains 9 exons. The protein is localized in the trans-Golgi region in round spermatids of mice and appears to play a role in vesicle transport from the Golgi apparatus to the acrosome. The primary defect in mutant mice is fragmentation of the acrosome in early round spermatids and abnormal acrosomal vesicles that fail to fuse. Later-stage spermatozoa also demonstrate nuclear malformation and abnormal mitochondria arrangement (Yao, Ito et al. 2002). Finally, *PICK 1* contains 13 exons encoding a 415 amino-acid protein. This cytosolic protein interacts with many membrane proteins via its postsynaptic density-95/Discs large/zona occludens-1 (PDZ) domain, resulting in the regulation of their subcellular targeting and surface expression. The primary defect in the testes of *Pick1* knockout mice was fragmentation of acrosomes in the early stages of spermiogenesis. This fragmentation was followed by defects in nuclear elongation and mitochondrial sheath formation, which led to round headed sperm, reduced sperm count and severely impaired sperm motility (Liu, Shi et al. 2010). Despite the fact that we did not find a

definitive mutation with a clear link to human globozoospermia, our results do not imply that *AFGF1*, *GOPC*, and *PICK1* do not play a role in human globozoospermia. The mouse knockout models for these genes, previously described, show how major disruptions of any of the 3 genes can clearly lead to infertility. Until sufficient numbers of patients have been screened, these genes should be considered as candidate genes.

The pull down was mainly successful due to the identification of chaperone proteins among the identified partners. This was expected because of the TPR domain of Spata16. Three partners, showing at least two peptide matches after mass spectrometry, were selected to be screened in globozoospermic patients: *TCP1*, *DNAJA4* and *IRGC*. Homozygosity was observed at the loci of these genes. *In silico* analysis showed that *TCP1* is ubiquitously expressed, while *IRGC* and *DNAJA4* are exclusively expressed in the testis.

TCP1 is a molecular chaperone implicated in tubulin biogenesis. DNAJA4 is a sterol regulatory element binding protein involved in the cholesterol biosynthesis pathway (Robichon, Varret et al. 2006). Abdul *et al*. showed that the same protein protects the heart muscle cells from various stresses. They showed that Dnaja4 constituted about 1% of the total protein in heart and is localized in the cytoplasm under normal conditions, whereas after heat shock, it is localized in the nucleus (Abdul, Terada et al. 2002). Hence, we supposed that DNAJA4 maintains the acrosome proteins during temperature fluctuations in the testis. Finally, there is no published data describing the role of *IRGC*. The fact that no mutation was identified does not exclude their involvement in human globozoospermia.

Concerning the function of Spata16, Min Xu *et al.* showed, by transfecting the fusion protein GFP-Spata16 in different cell types, that it is located at the Golgi apparatus and then shifts to Golgi vesicles (Xu, Xiao et al. 2003). Thus, they speculated that Spata16 might be involved in the formation of the acrosome during spermiogenesis. The immunohistochemistry on adult mice testis sections strengthened this hypothesis, because the protein was located at the acrosome of the round and elongated spermatids. Indeed, Spata16 might be required for the acrosome biogenesis from the early stages, since Min Xu et al detected its presence in the Golgi apparatus but the precise step at which Spata16 directly interferes has not yet been elucidated.

5- Conclusion

Overall, the results of the present study confirm the pathogenicity of the *SPATA16* mutations and their implication in acrosome biogenesis. The deletion of exon 2 is the cause of a new case of human globozoospermia. The involvement of Spata16 in globozoospermia was confirmed by immunofluorescence that revealed the localization of Spata16 in the acrosome. This would suggest that a mutation in *SPATA16* leads to a non-functional protein and thus disturbs the biogenesis of the acrosome. The pull down allowed us to determine new candidate genes potentially involved in globozoospermia, since they cooperate with Spata16. Among the 35 identified proteins, *IRGC, DNAJA4 and TCP1* were in a homozygous state in two consanguineous families and were screened for mutation, but unfortunately no mutation was detected. In a future study, these genes will be screened in the 14 remaining patients who were not the subject of a SNP genome wide screening. Furthermore, the pull down should be complemented by co-immunoprecipitation, in order to confirm the partners that cooperate with Spata16. Once the partners are confirmed, they will be subsequently screened for mutation. Colocalization studies and immunogold electron microscopy are also being considered to better characterize at which step Spata16 is involved in the acrosome biogenesis. Moreover, no mutation was identified in *AFGF1*, *GOPC*, and *PICK1* that were chosen based on the mice knock-out phenotypes.

The fact that no variation was detected in the coding sequence of the seven selected genes does not mean that they are not involved in human globozoospermia. Mutations could concern the non-coding sequence such UTRs and promoters that were not sequenced.

Since globozoospermia is a rare teratozoospermia, the main issue remains the lack of sufficient number of patients. A larger globozoospermic cohort must be studied in order to accelerate the process of identifying new genes involved in such phenotypes.

Azoospermia

Chapter IX

Linkage analysis and exome sequencing to find mutation causing non-obstructive azoospermia with maturation arrest

1- Introduction

Azoospermia, defined as a complete absence of spermatozoa in the semen, affects approximately 1% of all men and about 15% of infertile men (Oates 2012). It is related to many factors acting at the pre-testicular, post-testicular or directly at the testicular level and is classified into two categories i) obstructive azoospermia and ii) non-obstructive azoospermia (Lee, Dada et al. 2011).

Azoospermia at the pre-testicular stage includes two pathological conditions: a dysfunction of the hypothalamus and coital disorders (Krausz 2010). In fact, a dysfunction of the hypothalamus and pituitary glands leads to a deficiency of LH and FSH, resulting in disruption of the gonadal functions of the testis. Thus, testosterone synthesis in Leydig cells and spermatogenesis within the seminiferous tubules are impaired. This hormone deficiency may be due to congenital and acquired factors (Krausz 2010). The congenital form is associated with delayed puberty, gynecomastia and very low testicular volume and mostly caused by gene mutations. The acquired form in adulthood is due to tumors, infections or radiation treatments and manifests with several symptoms, such as a reduction of the volume of the ejaculate, of beard growth and asthenia. However, coital disorders such ejaculatory disorders are very a rare cause of infertility. The common example is anejaculation, also known as retrograde ejaculation, which occurs especially in diabetic patients. In this case, spermatozoa are absent in the semen, but they are detected in the urine.

Testicular azoospermia is characterized by spermatogenic defects despite the proper functioning of the hypothalamus and the pituitary glands. It may include a total absence of germ cells, known as Sertoli cell only syndrome (SCO), as well as maturation arrest of spermatogenesis (Oates 2012). In Sertoli cell only syndrome, a higher level of FSH is detected in the plasma due to the absence of retronegative control provided by germ cells (Sadeghi-Nejad and Farrokhi 2006). In contrast, azoospermia due to spermatogenic arrest is characterized by a normal level of FSH. Although the FSH value reflects the degree of spermatogenic failure, there is no absolute value that predicts the presence or the absence of spermatozoa in the testis. Causes for testicular failure include chromosomal abnormalities and Y microdeletion, as well as a large number of pathologies. Cryptorchidism defined as the absence of one or both testes from the scrotum, is the most frequent pathology that causes testicular failure (Virtanen and Toppari 2008). Acquired issues such as infections,

gonadotoxic medications, tumors, chemotherapy and radiotherapy may also disrupt spermatogenesis.

Since in pre-testicular and testicular azoospermia the reproductive ductal structures are present and do not show any abnormality, both of the above etiologies constitute subclasses of non-obstructive azoospermia. However, post testicular azoospermia concerns the blockage of the sperm at any level of the transport system, from the rete testis to the ejaculatory ducts and thus is classified as obstructive azoospermia (Lee, Dada et al. 2011). In this case, spermatogenesis is not altered and meiosis pursued normally. The most common cause of post testicular azoospermia is the congenital absence of the vas deferens due to *CFTR* mutations discussed previously in Chapter V section 1-E.

Although some causes of azoopermia have been established, a large amount remains idiopathic and can be associated with autosomal monogenic mutations. The aim of the following study is to identify autosomal genes responsible for non-obstructive azoospermia with maturation arrest in order to better understand spermatogenesis. For this reason, linkage analysis was performed on a Turkish consanguineous family including two azoospermic brothers and a sister and an aunt who have repetitive hydatidiform moles. Hydatidiform moles are abnormal conceptions of excessive trophoblast development, resulting in abnormal human pregnancies with no embryo and cystic degeneration of the chorionic villi and they are divided into complete and partial types (Williams, Hodgetts et al. 2010). The complete hydatidiform moles, known as androgentic moles, represent 75% of clinically ascertained moles and are solely paternally derived. In 90% of cases, they are the result of an empty oocyte fertilization followed by duplication of the haploid sperm, usually resulting in 46, XX karyotype. In the remaining cases, complete hydatiform moles result from fertilization by two haploid sperm. Their chromosome constitution is divided equally between 46, XX and 46, XY. In some rare cases, hydatidiform moles can have a familial biparental origin, with two sets of both maternal and paternal chromosomes (Bestor and Bourc'his 2006). This case is linked to recessive autosomal mutations that behave as maternal effect mutations, such as preventing the establishment of genomic imprints during oogenesis. On the other hand, partial moles are triploid resulting from dispermic fertilization and therefore contain one set of maternal haploid chromosomes and two sets of paternal haploid chromosomes. Contrary to the complete hydatidiform moles, partial moles show fetal tissues (Williams, Hodgetts et al. 2010). We hypothesized that a mutation in the same gene is responsible for the two phenotypes observed among the affected men and women. The hypothesis was as follows: i)

since gametogenesis occurs differently in men and women, the mutation will block spermatogenesis, while oogenesis will occur normally leading to the production of an anucleus oocyte which, upon fertilization, will lose its nucleus leading to a duplication of the paternal genome. ii) the fertilizable oocytes may have an imprinting alteration responsible for the hydatidiform moles, which in this case have bi-parental origin.

2- Materials and methods

Patients

This study includes a Turkish consanguineous family consisting of two brothers with non-obstructive azoospermia with maturation arrest at primary spermatocyte stage and a sister and an aunt who both have repetitive hydatidiform moles. It also includes 38 Turkish men with NOA with maturation arrest and 15 women who had repetitive hydatidiform moles.

DNA preparation

Genomic DNA from each patient was extracted either from peripheral blood leucocytes using QIAmp DNA Blood Midi Kit (Qiagen, Germany) or from saliva using Oragene DNA Self-Collection Kit (DNAgenotech, Canada).

Linkage analysis

This study started in 2006. A genome wide scan was performed on the DNA of the four patients of the Turkish family, as well as on the DNA of the fertile sister and of the parents using the GeneChip Human mapping 10 K 2.0 Xba from affymetrix. Single nucleotide polymorphisms (SNP) obtained were analysed with the HomoSNP program (plewniak@igbmc.u-strasbourg.fr) that was set to identify homozygous regions of 25 or more consecutive SNP.

PCR reactions and DNA sequencing

Mutation screening was performed by polymerase chain reaction (PCR) and direct sequencing of the exons and intronic flanking sequences of the selected genes. PCR primer sequences and amplification conditions are listed in table x. 30 cycles of PCR reactions were carried out in the T-Gradient PCR system (Biometra) in a volume of 25 µL using Taq DNA polymerase (Roche). DNA amplicons were purified and double strand sequencing of each DNA fragment were performed with primers used for the PCR amplification. Sequence analysis was carried out using an ABI PRISM 3730xl Genetic Analyser (Applied Biosystems, GATC, Germany).

Exome Sequencing and variants verification

Sequencing was conducted on an Illumina® HiSeq-2000 sequencer (IGBMC, Strasbourg, France). 10 µg of genomic DNA was extracted from the blood of the two Turkish azoospermic brothers and was fragmented into 200 to 350 nt fragments, to which nucleotide adapters were ligated. Agilent SureSelect target enrichment system was used to collect the protein coding regions of human genome DNA. Each sample was pooled, and sufficient sequencing was performed so that each sample had an average coverage of 95.5% with 72-bp paired-end reads using an Illumina HiSeq 2000 sequencer. The two samples were analyzed using Illumina Pipeline RTA (Real-Time Analysis) version 1.7 and BAM files were exported from this process (using the hg18 human genome draft). These BAM files were then run through a variant caller pipeline utilizing the Genome Analysis Toolkit (GATK). This pipeline implements best-practice approaches for SNP and indel calling, and variants for each individual sample were compiled into a single variant call format (VCF) file. Once all variants were obtained, annotated, and assessed from the exome sequencing process, variant filtration was performed. The sequencing reads were aligned to the human reference genome (NCBI Build 36.3). Data were provided as lists of sequence variants (SNPs and short indels) relative to the reference genome. Sanger sequencing was used to determine whether any of the remaining variants co-segregated with the disease phenotype in this family. Primers flanking the candidate loci were designed based on genomic sequences of Human Genome (hg18/build36.3). All shared variants of the two affected individuals after filtering were then confirmed by direct polymerase chain reaction (PCR) and analyzed on an ABI PRISM 3730xl Genetic Analyser (Applied

Biosystems, GATC, Germany). Sequencing data were compared pair-wisely with the Human Genome database.

3- Results

Our study started with the analysis of a consanguineous Turkish family of 5 individuals, including a fertile sister, two infertile brothers with NOA with maturation arrest at the primary spermatocyte stage and a sister and an aunt who has both presented repetitive complete hydatidiform moles.

Because of the consanguinity, we postulated that the genetic abnormality was transmitted as an autosomal-recessive disorder. We performed a genome-wide scan analysis of the five family members and the parents, using 10K SNP arrays (Affimetrix Genechip). Regions of homozygosity were defined by the presence of at least 25 consecutive homozygous SNPs. We supposed that the same mutation is responsible for the two phenotypes observed in the family. We identified a unique region of 80 homozygous SNPs shared by the brothers, the infertile sister and the aunt; the fertile sister and the parents were heterozygous for this region (**Figure 27**). This region of about 27 Mb on chromosome 11 (positions 60635238 to 87433821, on the GRCh36/hg18 version of the human genome, corresponding to SNPs rs522073 and rs953353) contains 477 genes according to the UCSC Genome Browser.

We selected seven candidate genes according to their predominant expression in the testis and the ovary: *MTL5*, *CCDC89*, *FGF4*, *INCENP*, *PIWIL4*, *POLA2* and *TSGA10IP*. Sequence analysis of the seven genes did not reveal any mutations. Therefore, our next step was to perform high throughput exome sequencing on the genomes of the two azoospermic brothers.

We did not carry out a targeted sequencing, covering the unique homozygous region shared between the four patients, for many reasons. First, there is a possibility that our first hypothesis might not be valid. Two different genes may be responsible for the two phenotypes observed in the family. Consequently, when analyzing the two azoospermic brothers alone, five regions of 115 Mb in total **(**see **Table 1)** were found to be homozygous between them and not between the parents. Moreover, the 10K Affimetrix array does not cover the entire genome and it is possible that the causative gene may be located in an uncovered region.

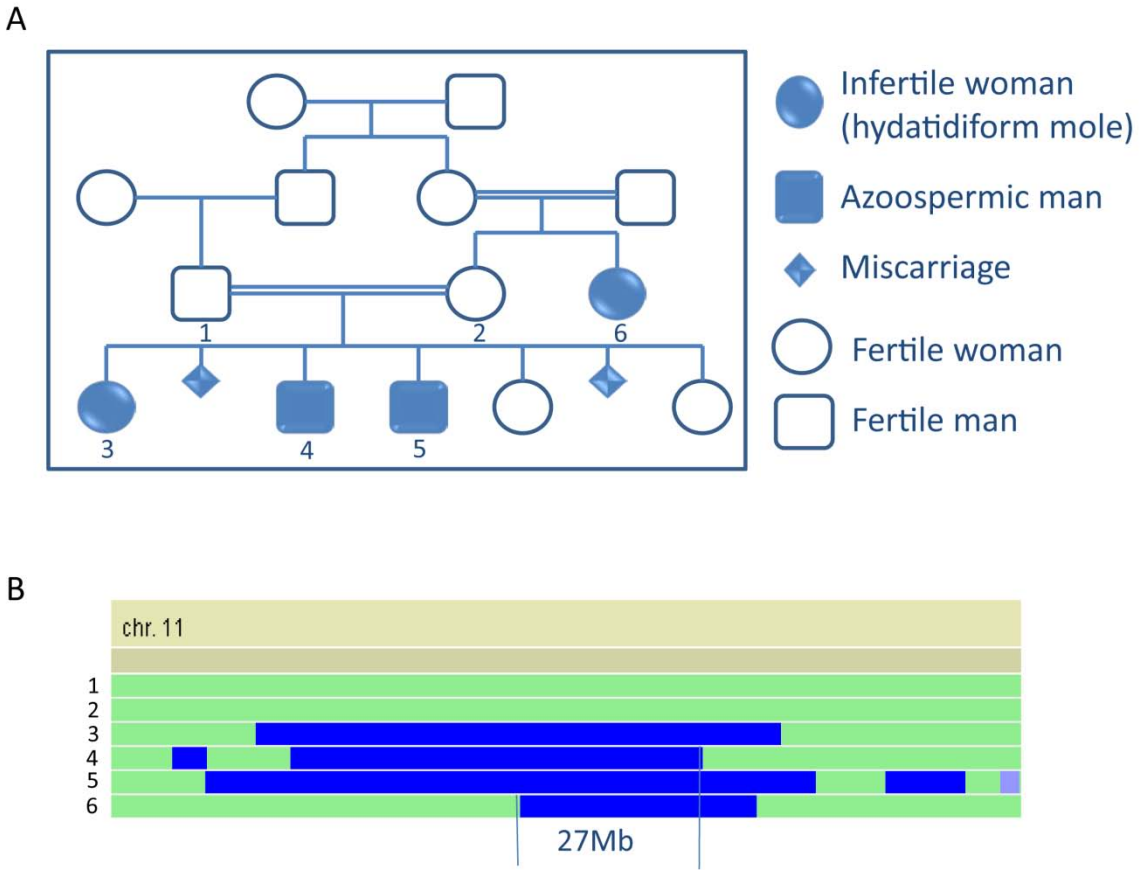

Figure 27 : Pedigree and Linkage Analysis of the Turkish Family. (A) Turkish pedigree showing second-degree consanguinity. (B) SNP array results of the four infertile patients, the fertile sister and the parents for the region of chromosome 11. Shared regions of homozygosity are visualized by the HomoSNP software, which displays one patient per line. The areas of homozygosity with 25 or more SNPs are blue, whereas homozygosity regions defined by 15–25 consecutive SNPs are violet. Regions of heterozygosity are green. The four affected siblings share a region of homozygosity of 27 Mb on chromosome 11, which is not shared by the fertile sister and the parents.

Table 1: Homozygous regions shared between the two azoospermic brothers.

Chromosome	SNPs number	Start position	End position	Length (Mb)
3	116	4730979	29032190	24
6	81	106080157	125875447	20
9	38	239391	6606648	6
11	205	27359908	87433821	60
14	33	19490525	24650499	5

Exome sequencing was performed with 72-bp paired-ends and 16,174 variants were found to be shared among the two individuals. By restricting the variants to those which were homozygous between the two patients and which were present within the region of interest (chromosome 11), 587 variants were retained. Among these variants, 142 caused non synonymous variations or disrupted the coding sequence and only 4 were not reported in dbSNP. These mutations were further confirmed by Sanger sequencing performed on the four patients, the fertile sister as well as the parents. Only one variant segregated with the disease in the family. In fact, we identified a missense mutation in exon 1 of *ARHGEF17* leading to amino acid change p.Gln805Glu. Protein alignment analysis of 41 different systems showed that Glutamine 805 is conserved in mammals, with the exception of Sarcophilus arisii, Monodelphis domestica, Macropus eugenii and Cavia porcellus, which have a histidine (H) at this position. Birds and the reptile Anolis carolinensis have a serine at this position, and in fish, this residue is variable [C, N, D, A, S]. In addition, secondary structure predictions indicated that this residue is located in a disordered region in an area without any structural or functionally assigned domain. Finally, we screened this mutation and other exonic mutations of the gene in 38 patients with NOA with maturation arrest and 15 women with recurrent complete hydatidiform moles, but unfortunately no variation was detected. .

4- Discussion

The presence of a unique homozygous region of 27 Mb on chromosome 11 of the four affected siblings reinforces the hypothesis that the same gene is involved in the two phenotypes observed in the Turkish family. In fact, many examples showed that the same gene may affect gametogenesis differently in both sexes. For instance, male *Dnmt3l*-KO mice showed a complete azoospermia with maturation arrest at the pachytene stage, while female *Dnmt3l*-KO mice produce morphologically normal and fertilizable oocytes (Liao, Tai et al. 2012). In male mice, deletion of the *Dnmt3l* gene prevented the de novo methylation of both long terminal repeat (LTR) and non-LTR retrotransposons, which were transcribed at high levels in spermatogonia and spermatocytes (Bourc'his and Bestor 2004). In addition to hypomethylation in early stages of spermatogenesis, *Dnmt3l* -/- spermatocytes showed abnormalities during later stages of development. Homologous chromosomes failed to align and to form synaptonemal complexes, leading to spermatogenetic arrest and loss of spermatocytes through apoptosis (Webster, O'Bryan et al. 2005). Mice embryos derived from

female *Dnmt3l*-KO oocyte die around 9.5 dpc due to a default of parental imprinting methylation during oogenesis (Kobayashi, Sakurai et al. 2012). Moreover, human diseases were associated with this faulty imprinting, such as molar pregnancies and ovarian teratomas (Liao, Tai et al. 2012). In our study, we focused on genes which were located within the unique homozygous region at chromosome 11 and with a high expression in the gonads, supposing that they might have a similar role to *Dnmt3l.* Since the region of interest contains around 472 genes and our first screen for mutations in *MTL5*, *CCDC89*, *FGF4*, *INCENP*, *PIWIL4*, *POLA2* and *TSGA10IP* was not within our expectations, we preferred to combine the linkage analysis strategy with the whole exome sequencing. The exome sequencing performed on the two azoospermic brothers revealed a mutation in *ARHGEF17* after several filtering steps that eliminated the known variants in the database and those that did not segregate with the disease. The substitution c.2413C>G in exon 1 of *ARHGEF17* leads to amino acid change p.Gln805Glu, which is located before the dbl homology domain, also known as RhoGEF. In fact, the function of ARHGEF17 is unknown, but the protein contains a RhoGEF domain and belongs to the Guanine nucleotide exchange factor family. Guanine Nucleotide Exchange Factors (GEFs) are proteins involved in the activation of small GTPases, which act as molecular switches in intracellular signaling pathways. One family member was shown to be implicated in microtubule biogenesis, which allowed us to think that a mutation of *ARHGEF17* may disrupt meiosis. So far, we could not confirm the pathogenicity of the mutation and thus we suppose that it is not deleterious and may not be responsible for infertility in this family.

5- Conclusion

To date, no autosomal recessive mutation has been identified as responsible for human azoospermia. We supposed that the same mutation is responsible for the non-obstructive azoospermia and for complete hydatidiform moles, which we considered had biparental origin. The conjunction of linkage analysis and exome high throughput sequencing is a powerful strategy, but still has some limitations. The fact that no mutation is identified can be due to the fact that the variation concerned a non-covered exon in the region. All the genes within the region of interest will be checked in order to select the non-covered exons in a first step and then to verify the presence of a variation by Sanger sequencing.

We did not exclude the possibility that two distinct genes can be involved in the observed

phenotypes. For this purpose, the two diseases will be studied separately if no common mutation is highlighted. Furthermore, we cannot omit the possibility of a wrong diagnosis since we did not have access to the testicular biopsy results. Finally, we think that recruiting more families presenting the same case of infertility will accelerate the process and will increase the chances to identify the causative gene(s).

Conclusions and perspectives

In the presented study, we identified recessive autosomal mutations that impair human male fertility. In fact, by studying a consanguineous Jordanian family, we identified a whole deletion of *DPY19L2* as a major cause of globozoospermia (67% of patients are mutated for *DPY19L2*). The deletion of 200 Kb is due to a non-allelic homologous recombination (NAHR) between the two low-copy repeats (LCRs) regions sharing 96.5 % of identity and flanking *DPY19L2*. We have analyzed a total of 54 patients: 31 were homozygous deleted for *DPY19L2*, 7 were heterozygous composite for the deletion and a point mutation and 4 were homozygous for a point mutation. This study, which is the largest study so far on human globozoospermia, allowed us to identify 9 breakpoints that clustered in two hot spots of recombination, both containing direct repeat elements (*Alu*Sq2 in hotspot #1, THE1B in hotspot #2). Thus, we defined globozoospermia as the first autosomal genomic disorder that caused infertility in men. The 18 remaining patients, for whom we could not find any mutation in *DPY19L2*, were screened for mutations in *SPATA16*, and we identified a new mutation in one patient, which deletes exon 2 and confirms the pathogenicity of *SPATA16* mutations. I also showed that Spata16 is expressed in the mouse testis and is located at the acrosome. For the 18 patients who did not show any mutation either in *DPY19L2* or in *SPATA16*, other genes are probably responsible for globozoospermia. In fact, more than one gene needs to be identified, since we could not determine a unique homozygous causative locus shared by three consanguineous families of our cohort. So far, the strategy of GST-Spata16 pulldown did not identify any partner involved in human globozoospermia, probably due to the limited number of patients. However, concerning the azoospermia, no autosomal mutation was identified in our study of a consanguineous Turkish family. The failure to identify causative genes in some cases can be explained by the fact that human gametogenesis is a complex process and infertility may be the result of multigenic and multifactorial causes. To overcome this problem, a large cohort of well diagnosed informative patients must be studied.

These studies are of tremendous importance at different levels. They represent a unique way to study human gametogenesis and to decipher the underlying mechanisms, which in turn will have clinical and fundamental repercussions: (i) patients will benefit by having a molecular diagnosis followed by appropriate counsel and care; (ii) results will contribute to a better understanding of the physiopathological process of human reproduction; (iii) in the longer term a better understanding of the genes involved in spermatogenesis will help to identify the extrinsic factors that compromise spermatogenesis and prevent the deterioration of fecundity;

(iv) together this should also bring improvement in the practice of artificial reproductive technologies, not only to mutated patients but to all couples.

Couples experiencing male infertility problems should undergo genetic counseling before pursuing assisted reproductive treatment in order to suggest an appropriate solution. ICSI cannot be proposed as a treatment for all cases of infertility. For example, patients with enlarged headed spermatozoa, due to a mutation in AURKC, have only tetraploid gametes (Dieterich, Soto Rifo et al. 2007); it is therefore useless to offer them an ICIS treatment. Contrariwise, it is reasonable to propose ICSI treatment for globozoospermic patients since round headed spermatozoa do not present a higher rate of chromosomal abnormalities (Machev, Gosset et al. 2005). However, it has been shown that a low fertilization rate is obtained when ICSI is performed with globozoospermic sperm because of its inability to initiate oocyte activation. This is due to the absence of *PLCζ*, the major factor responsible for oocyte activation which is located at the acrosome of the sperm (Kashir, Heindryckx et al. 2010). Following sperm/oocyte fusion, PLCζ is released into the cytoplasm of the oocyte, where it facilitates the hydrolysis of membrane-bound PIP2 to DAG and IP3, triggering Ca2+ release from intracellular Ca2+ stores, leading to Ca2+ induced Ca2+ release and oocyte activation (Ramadan, Kashir et al. 2012). To overcome oocyte activation failure and failed fertilization following ICSI, artificial oocyte activation (AOA) has been applied. Several activating agents have been shown to be able to overcome fertilization failure, such as electrical pulses, strontium chloride and calcium ionophores (Yanagida, Katayose et al. 1999; Eldar-Geva, Brooks et al. 2003; Chi, Koo et al. 2004; Heindryckx, Van der Elst et al. 2005; Yanagida, Morozumi et al. 2006). These artificial activation agents cause a single prolonged calcium rise, but fail to elicit physiological patterns of calcium release. The most common type of AOA implies the use of a calcium ionophore, which has been shown to improve fertilization and pregnancy rates. Fertilization rates were significantly increased after AOA application to a normal level (76.3%, 33 cycles) and successful pregnancies were established in all groups of patients (48.5% average pregnancy rate) (Heindryckx, Van der Elst et al. 2005; Heindryckx, De Gheselle et al. 2008).

The aim of our next study is to compare results of ICSI with or without AOA in several globozoospermic patients mutated for *DPY19L2*, or not, in order to be able to give a prognosis based on the molecular etiology of this phenotype. We have been able to collect data concerning ICSI attempts for 34 couples and 83 cycles. Male and female baseline characteristics were collected as well as cycle parameters, such as the number of retrieved

metaphase II (MII) oocytes, the overall fertilization rate, the number of good quality embryos, the number of transferred embryos and the outcomes of these transfers in terms of hCG result, ongoing pregnancy, and delivery. Couples were divided into four groups depending on whether they are mutated or not for *DPY19L2* and whether their injected oocytes underwent or not AOA. Our preliminary results showed that AOA- patients showed a much lower fertilization rate, which led to less transferable embryos and lower ongoing pregnancy rates compared to the AOA+ group of patients, which showed results comparable to the general ICSI population independently of *DPY19L2* mutations.

On the other hand, functional studies of Spata16 are still on going. A co-immunoprecipitation should be performed to validate the partners of the protein. This approach will allow us to determine candidate genes potentially involved in human globozoospermia, which will be screened in a larger cohort of patients.

Immunohistochemistry analysis will allow us to follow the expression profile of Spata16 in the seminiferous tubules at different stages. Immunoblot analysis will be performed on ejaculated mature sperm to see whether Spata16 remains at the acrosome at the final steps of spermiogenesis or is eliminated with the cytoplasmic droplet. Immunogold electron microscopy can be applied on sections of wild type testis mice to specifically localize the protein.

Azoospermia remains a more complex phenotype and we believe that recruiting other families will greatly facilitate the analysis. Considering the two phenotypes observed in the Turkish family are caused by two different genes will lead us to investigate more homozygous loci and thus to screen more candidate genes.

Finally, the results of this study can have broader implications. Indeed, a general decline in fertility is observed, which is obviously not due to genetic causes but rather caused, at least in part, by increased concentrations of environmental endocrine disruptor, such as environmental oestrogen like molecules (Sharpe and Skakkebaek 1993; Skakkebaek, Rajpert-De Meyts et al. 2001). We believe that a better understanding of spermatogenesis, starting with in depth knowledge of all the genes involved, will in the long term be instrumental in identifying and understanding the effects of these toxic chemicals and finding solutions to stop their adverse effects.

Genetics of infertility is a new field that will, undoubtedly, take on greater importance in the near future, even if trying to do genetics of infertility can sound paradoxical. Indeed, genetics consist of the study of the transmission from one generation to the next of genetic traits and in

contrast, infertility studies try to understand why genetic traits cannot be transmitted to the next generation. These studies will allow the identification not only of genes involved in spermatogenesis but also risk factors. More particularly, it will enable the establishment of the most exhaustive list of combinations of genetic variations and environmental factors that could lead to spermatogenesis defaults.

Appendix

French abstract

Résumé de thèse

1- Etat de la question

L'infertilité est définie comme l'incapacité de concevoir après un an de rapports sexuels non protégés. L'OMS estime que 70 à 80 millions de personnes souffrent d'infertilité dans le monde ce qui en fait un véritable problème de santé publique. En effet, environ 9% des couples sont confrontés à ce problème et dans 50% des cas l'infertilité est due à un facteur masculin caractérisé par une absence totale de spermatozoïdes (azoospermie) ou par l'incapacité de l'homme de produire des spermatozoïdes en nombre suffisant (oligozoospermie), de mobilité adéquate (asthénozoospermie) ou de morphologie normale (tératozoospermie). Les causes de l'infertilité sont multiples et comprennent les facteurs environnementaux, les dérégulations hormonales, les troubles sexuels et les facteurs génétiques. Dans d'autres cas, l'infertilité est associée à des syndromes complexes comme la mucoviscidose, le syndrome de klinefelter, le sex reversal et autres. D'autre part, un certain nombre d'infertilités reste inexpliqué (30% des cas) et peut être attribué à une origine génétique sans exclure la possibilité d'une origine multifactorielle. Nous sommes intéressés au laboratoire par le versant génétique de l'infertilité et notamment la globozoospermie et l'azoospermie. Notre but est d'identifier des gènes autosomales impliqués dans l'infertilité humaine non syndromique et d'étudier leurs fonctions afin de mieux comprendre la physiologie de la gamétogenèse humaine.

2- Stratégie de travail

Notre travail repose sur le recrutement de familles consanguines ou de personnes originaires d'une région isolée (dans le cas d'une mutation à effet fondateur) et qui dans les deux cas ne présentent aucune pathologie expliquant leur infertilité. L'approche utilisée est celle du clonage positionnel et plus particulièrement, la cartographie par homozygotie. Les études de liaison génétique (cartographie par homozygotie) permettent de lier le gène morbide à une région génomique donnée en montrant la ségrégation de la maladie avec certains allèles des marqueurs de cette région. La cartographie par homozygotie est basée sur l'hypothèse qu'un enfant, issu d'un mariage consanguin et atteint d'une maladie génétique récessive, est homozygote par descendance et reçoit les deux allèles portant la mutation d'un ancêtre commun. Les individus atteints présenteront donc une région d'homozygotie au locus morbide alors que les individus sains présenteront une région hétérozygote pour le même

locus. Une étude *in silico* des gènes présents dans cette (ces) région(s) permettra de sélectionner des gènes candidats : en priorité des gènes impliqués dans la gamétogenèse (particulièrement la méiose) et exprimés préférentiellement dans les gonades.

3-Approches expérimentales et Résultats

La globozoospermie est une tératozoospermie monomorphe particulière rare (<0.1%) caractérisée par la présence de spermatozoïdes à tête ronde sans acrosome **(Figure1)**. Les patients présentant un tel défaut sont infertiles car l'acrosome permet aux spermatozoïdes de traverser la zone pellucide de l'ovocyte lors de la fécondation.

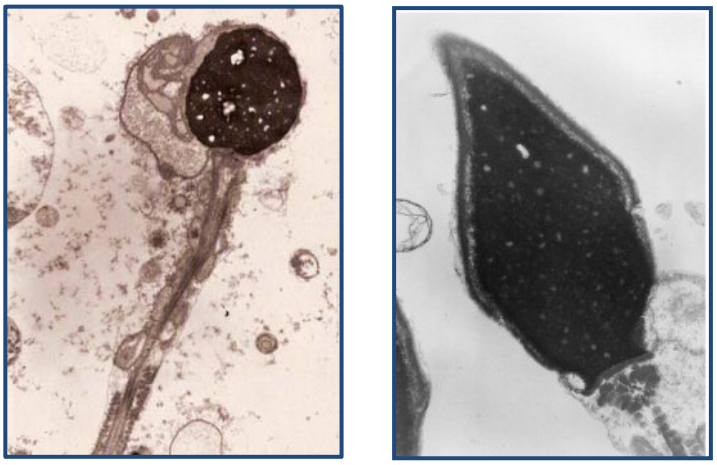

Figure 1: Microscopie électronique d'un spermatozoide globozoospermique (gauche) et spermatozoide normal (droite).

Le génotypage sur puce Affymetrix 10K d'une famille jordanienne consanguine constituée de 4 frères globozoospermiques et de 3 frères fertiles, a permis de localiser un nouveau gène responsable de la globozoospermie situé dans un intervalle de 6.4Mb en 12q14.2 **(Figure2)**. Au regard de son expression prédominante dans le testicule et de l'implication de son orthologue, chez *C. elegans*, dans la polarisation cellulaire, le gène *DPY19L2* est un gène candidat parfait. Le gène, codant pour une protéine transmembranaire, est flanqué par deux séquences répétées (LCRs) qui partagent 96,5% d'identité. Une délétion de 200kb englobant l'ensemble du gène a été mise en évidence chez les 4 frères infertiles de cette famille jordanienne ainsi que chez 3 autres patients non apparentés (Koscinski *et al*. 2011). Le mécanisme sous-jacent de la délétion est très probablement une recombinaison homologue non allélique (NAHR) entre les LCRs.

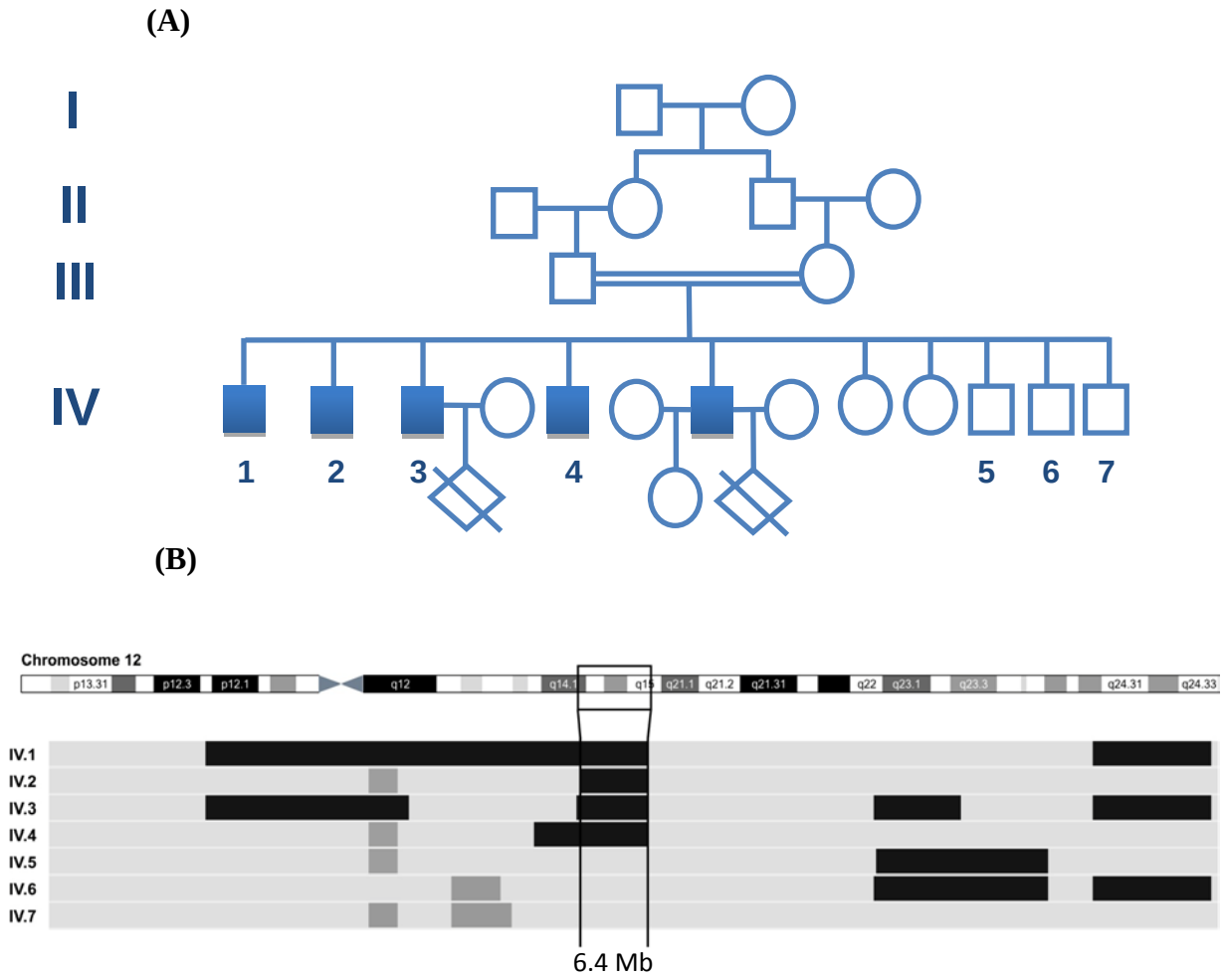

Figure 2 : Pedigree et étude de liaison de la famille jordanienne. (A) Pedigree de la famille jordanienne montrant une consanguinité du second degré. (B) Résultats des puces SNP des quatres frères infertiles et des trois frères fertiles. Une région homozygote de 6.4 Mb sur le chromosome 12 est partagée entre les frères infertiles.

Dans notre étude initiale, nous avions identifié uniquement 19% de nos patients présentant une telle délétion. En revanche, un taux beaucoup plus élevé de 75% a été décrit dans une cohorte de 20 patients tunisiens (Harbuz *et al*. 2011). Cet écart est probablement lié à la différence de recrutement des deux cohortes (nos patients ont différentes origines éthniques). Afin d'obtenir une meilleure idée de la fréquence de l'implication de *DPY19L2* dans la globozoospermie, nous avons recruté une plus grande cohorte de patients. Parmi ces patients, 20 sont homozygotes pour la délétion de *DPY19L2*, et 7 sont hétérozygotes composites associant la délétion hétérozygote et une mutation ponctuelle. En outre, nous avons identifié 4

patients avec des mutations ponctuelles homozygotes dont un présente une délétion des exons 5 et 6 et un autre une délétion des exons 5, 6 et 7 certainement due à un mécanisme de non-homologous end joining (NHEJ) **(Figure 3)**.

Par conséquent, la fréquence d'implication de *DPY19L2* s'élève à 66,7%. En séquençant les jonctions, 9 points de cassures au total sontidentifiés. Ils sont regroupés en deux régions au sein des LCRs définissant ainsi deux hot-spots stimulant la recombinaison: la séquence Alu et l'élément répété THE1B. Ce dernier coïncide avec un hotspot méiotique déjà identifié par le projet HapMap.

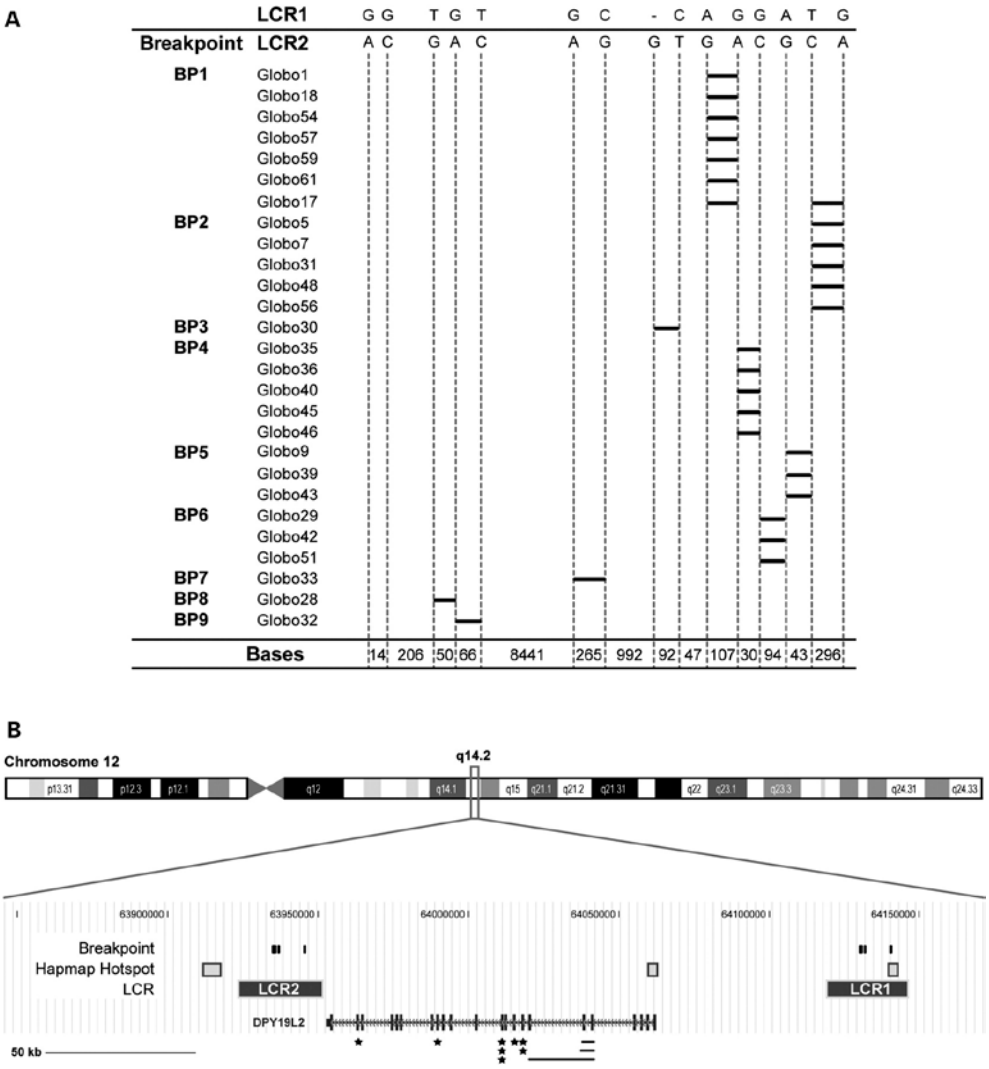

Figure 3. (A) Les nucléotides spécifiques et différents entre les deux LCRs au sein des régions hotspots sont présentés en haut du diagramme. La longueur et la distance entre les nucléotides différentes sont affichées au fond. La région de recombinaison pour chaque patient est représentée par une barre noire. Les patients sont classés en fonction des neufs points de cassures (BPs). (B) Le locus *DPY19L2* est représenté avec les deux LCRs, les points chauds de recombinaison du projet HapMap, les neuf points de cassures différents et les mutations identifiées (les mutations ponctuelles sont marqués d'une étoile et les grandes délétions sont représentés en utilisant une ligne horizontale noire)

D'autre part, en 2007 notre équipe a décrit *SPATA16* comme premier gène autosomale responsable d'une globozoospermie humaine. Une mutation perturbant l'epissage de l'exon 4 de ce gène (c.848G>A) a été mise en évidence en analysant par puce SNP une famille consanguine.

Cette mutation a été recherchée chez des patients globozoospermiques ne présentant pas de mutation pour *DPY19L2*, mais aucun n'était porteur. En revanche, une autre mutation, une délétion de l'exon 2 (premier exon codant) de *SPATA16*, a été détectée chez un patient d'origine turque. La délétion de 22.6 Kb entraine la délétion du premier ATG introduisant un codon stop prématuré ainsi que la délétion de 36 acides aminés du domaine de répétition tetratricopeptide (TPR) connu comme domaine d'interaction protéiques. Afin de mieux comprendre la fonction de Spata16, j'ai produit des anticorps de lapin dirigés contre la protéine murine recombinante GST-Spata16 produite dans une souche bactérienne *E.coli* BL21. Par immunohistochimie sur des coupes de testicule de souris adulte, nous avons détecté la présence de Spata16 au niveau de l'acrosome des cellules germinales **(Figure 4)**.

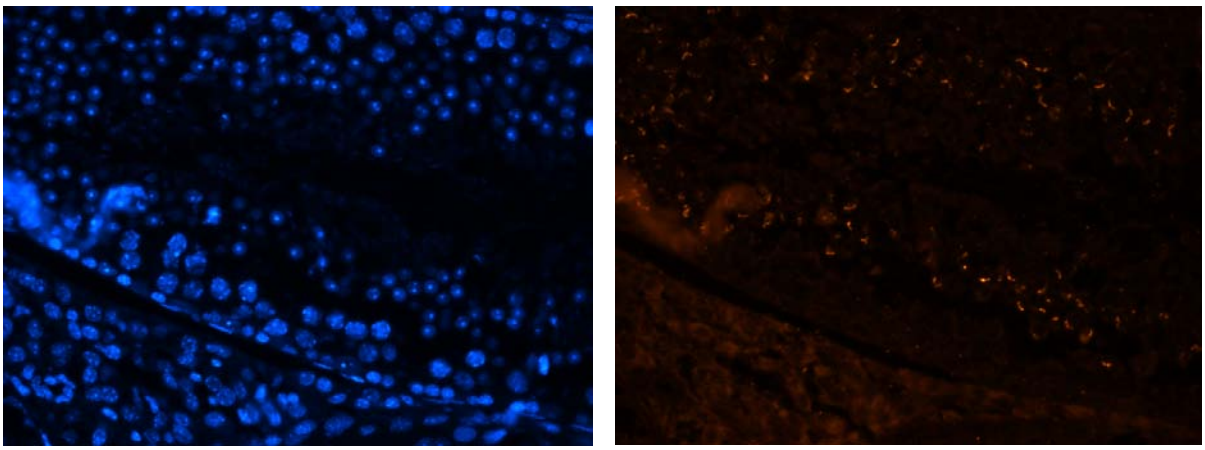

Figure 4 : Localization de Spata16 au niveau de l'acrosome des cellules germinales. Des sections de testis de souris adultes sauvages sont colorées au DAPI (gauche) et marquées par l'anticorps anti-SPATA16 (droite)

En plus, un GST pulldown en présence d'un extrait protéique testiculaire de souris a été réalisé afin d'identifier les partenaires de Spata16 **(Tableau1)**. Cette approche nous a permis de sélectionner ainsi de nouveaux gènes candidats pour la globozoospermie. Une étude *in silico* de ces gènes combinée à l'étude de liaison a été faite pour sélectionner ceux qui seraient impliqués dans la biogenèse de l'acrosome et dans le trafic vésiculaire et localisés dans des

régions homozygotes chez des patients issues de familles consanguines. Trois gènes ont été sélectionnés : IRGC, DNAJA4 et TCP1 et ont été ensuite analysés par amplification et séquençage des exons et des jonctions exons/introns, mais aucune mutation n'a été détectée.

Tableau 1 : Partenaires de Spata16 identifiés par GST-pulldown. Les protéines sélectionnées en vert sont exprimées exclusivement dans les testis.

Spata16 partners		
Col6a3	Nid1	Irgc1
Hspa8	Tcp1	Hspa4l
Tubb3	Clpx	Dnaja1
Tubb5	Actbl2	Ldhc
Myh11	Hnrnph1	Dnaja4
Col6a1	Spata16	Kdm5b
Myh9	Cltc	
Myh10	Gm5409	
Lamb2	Slc2a3	
Rcn2	Eef1a1	
Col6a2	Axin2	
Atp5a1	Pcbp1	
Eef1a2	Grpel1	
Sucla2	Myl6b	
Dnaja2		

Concernant l'azoospermie, nous disposons d'une famille consanguine turque comportant deux frères présentant une azoospermie non obstructive avec un arrêt de méiose au stade spermatocyte I, une sœur et une tante ayant déjà fait des moles hydatiformes. Les moles hydatiformes sont des moles androgénotes résultant d'une fécondation d'un ovocyte dépourvu de noyau ou présentant des défauts d'épigenèse par un spermatozoïde suivi d'une duplication du matériel génétique paternel. L'hypothèse posée est que les deux frères et les 2 femmes ont la même mutation qui s'exprime différemment chez les deux sexes. Un arrêt de méiose est donc observé chez les hommes alors que chez les femmes l'ovogenèse n'est pas perturbée et se poursuit normalement. Pour une raison encore inconnue l'ovogenèse s'achève avec soit une perte du matériel génétique soit des défauts d'épigenèse. Plusieurs exemples montrent que la gamétogenèse humaine se déroule différemment chez les deux sexes. En effet, une translocation chez les hommes perturbe la spermatogenèse et cause un phénotype

allant de l'oligozoospermie à l'azoospermie. Alors que la même translocation chez les femmes ne cause pas un défaut de l'ovogenèse. Le génome des patients ainsi que celui des parents a été analysé sur des puces SNP 10k. Une seule région homozygote localisée sur le chromosome 11 est partagée entre les patients. La région fait 25 Mb contenant au moins 430 gènes. Sept gènes ont été sélectionnés selon leur profil d'expression qui est prédominant dans les gonades et leur rôle potentiel dans la méiose. Le séquençage des exons et des jonctions introns/exons de ces gènes n'a mis en évidence aucune mutation responsable du phénotype observé. Nous avons poursuivi l'analyse en utilisant la technique de séquençage haut débit afin de séquencer tous les exons du génome (exome high throughput sequencing) des frères azoospermiques pour nous assurer de ne pas passer à côté d'une mutation dans une région non mise en évidence par les puces SNP.

L'analyse du séquençage consiste à identifier la mutation en cause de la maladie parmi les 25000 variants générés. A cette fin, un tri est appliqué consistant à éliminer i) les variations déjà répertoriées dans les bases de données comme dbSNP et 1000 génomes en tant que polymorphismes non délétères ; ii) les variations hétérozygotes puisque nous étudions des familles consanguines, nous sommes à la recherche de mutations homozygotes ; iii) et les mutations faux sens n'altérant pas la fonction de la protéine. Seuls les variations homozygotes et qui sont identiques chez les deux frères sont retenues. Dans une première étude seules les variations de la région du chromosome 11 ont été analysées. Un seul gène correspondait à la ségrégation de la maladie dans la famille : TEM4. En effet, il s'agit d'une mutation faux-sens substituant la glutamine en glutamate en position 805 de la protéine. La mutation n'affecte pas un domaine conservé de la protéine et les prédictions bioinformatiques (Alamut) ne montrent aucun effet nuisant à la fonction de la protéine. Afin de confirmer la pathogénicité de cette mutation et l'implication du gène dans un tel phénotype, un screen de cette mutation et d'autres concernant le même gène a été effectué chez 38 hommes azoospermiques et 15 femmes ayant fait des moles d'hydatiformes mais aucun changement n'a été mis en évidence.

4- Conclusions et Perspectives

L'étude menée sur la globozoospermie, qui représente la plus large étude réalisée, nous a permis d'identifier *DPY19L2* comme le gène principal de la globozoospermie et d'élargir le spectre des mutations possible dans ce gène (deletion entière, mutations faux sens, non-sens et délétion partielle). Ceci est un résultat majeur puisqu'il permet de définir une stratégie de

diagnostic efficace. Il est important de ne pas limiter la recherche de mutation à la seule délétion complète du gène. Nous confirmons aussi le mécanisme sous-jacent de la délétion du gène qui est due au NAHR - liée à l'architecture particulière du locus - plutôt qu'un effet fondateur. Ces travaux nous ont permis également de confirmer la pathogénicité de *SPATA16* et de montrer par immunohistochimie que la protéine est impliquée dans la biogénèse de l'acrosome. Une étude concernant les résultats d'ICSI (Intra Cytoplasmic Sperm Injection) sera menée sur des patients présentant une mutation de *DPY19L2* ou de *SPATA16* afin de voir s'il y a une corrélation entre le taux de réussite de grossesses et la mutation observée chez ces patients. Concernant Spata16, une co-immunoprécipitation est envisagée pour valider les résultats du GST-pulldown. Tous les partenaires seront ensuite séquencés chez les patients qui ne présentent aucune mutation ni pour *DPY19L2* ni pour *SPATA16*.

Jusqu'à présent, aucune mutation autosomale récessive n'a été décrite comme responsable d'une azoospemie humaine. Ceci peut être expliqué par la complexité et l'hétérogénéité de la spermatogénèse sans exclure un effet multifactoriel. Une approche différente sera appliquée pour analyser les phénotypes observés chez la famille turque dont nous disposons. En effet les deux phénotypes seront étudiés séparément suggérant que deux gènes différents seraient responsables de l'infertilité chez les deux sexes. Finalement, le recrutement d'autres familles accélérera l'identification des gènes probablement responsables d'infertilité humaine.

L'originalité de ce travail consiste à caractériser des gènes impliqués dans une infertilité, non héréditaire par définition. C'est le moyen le plus puissant permettant la compréhension de la gamétogenèse humaine. Ce travail permet également de délivrer une meilleure information aux patients ayant recours à l'assistance médicale à la procréation.

Supplementary data

Chapter VI

DPY19L2 deletion as a major cause of globozoospermia

The American Journal of Human Genetics, Volume *88*

Supplemental Data

DPY19L2 Deletion as a Major Cause of Globozoospermia

Isabelle Koscinski, Elias ElInati, Camille Fossard, Claire Redin, Jean Muller, Juan Velez de la Calle, Françoise Schmitt, Mariem Ben Khelifa, Pierre Ray, Zaid Kilani, Christopher L.R. Barratt, and Stéphane Viville

Supplemental Material and Methods

Patients and controls

The pedigree was identified at the Farah Hospital, Amman, Jordan. Other patients were identified at the University Hospital of Strasbourg or provided by collaborators Pr. C. Jimenez (Dijon), Dr. F. Carré-Pigeon (Reims), Pr. J.M. Grillot (Marseille), Dr. F. Brugnon (Clermont Ferrand), Dr. D. De Briel (Colmar), Dr. S. Declève-Paulhac (Limoges). This study was approved by the local Ethical Committee (Comité de protection de la personne, CPP) of Strasbourg University hospital. For each case analyzed informed written consent was obtained accordingly to the CPP recommendations.

DNA preparation

Genomic DNA was extracted either from peripheral blood leucocytes using QIAamp DNA Blood Midi Kit (QIAGEN, Germany) or from saliva using Oragene DNA Self-Collection Kit (DNAgenotech, Ottawa, Canada).

SNP mapping

A genomewide scan was performed on the 4 affected and the 3 healthy brothers of the family using the Affymetrix GeneChip®Mapping 10K array (Affymetrix, Santa Clara, CA). Sample processing and labeling were performed according to the manufacturer's instructions (Affymetrix Mapping 10K 2.0 Assay Manual, Version 1.0, 2004). The hybridization was performed using a GeneChip Hybridization oven 640, washed with the GeneChip Fluidics Station 450 and scanned with a GeneChip Scanner 3000. SNP allele calls were generated by the GeneChip DNA Analysis Software version 3.0.2 (GDAS). Regions of homozygosity were defined by the presence of more than 25 consecutive homozygous SNPs. The unique region of homozygosity by descent

segregating with the disease in the family was entered and screened in the UCSC Genome Browser (http://genome.ucsc.edu , version March 2006) for selection of candidate genes.

Deletion analysis

PCR products covering exon 2, 7, 9, 10, 13, 17 and 21 of *DPY19L2* and their exon-intron boundaries were amplified using genomic DNA from a control and one of the affected brothers and subsequently of the 24 patients with globozoospermia. A walk 5' and 3' was then carried out to identify the deletion, 3 amplicons were tested on the 5' and 2 on the 3' region of the gene. All primer sequences and PCR conditions are available in the supplementary data (Table 1). Because of the high conservation of the duplicated regions special care was taken to choose specific oligonucleotides. Even so, all amplicons were sequenced in order to control for the specificity of the PCR.

CLUSTALW 2.0.12 multiple sequence alignment

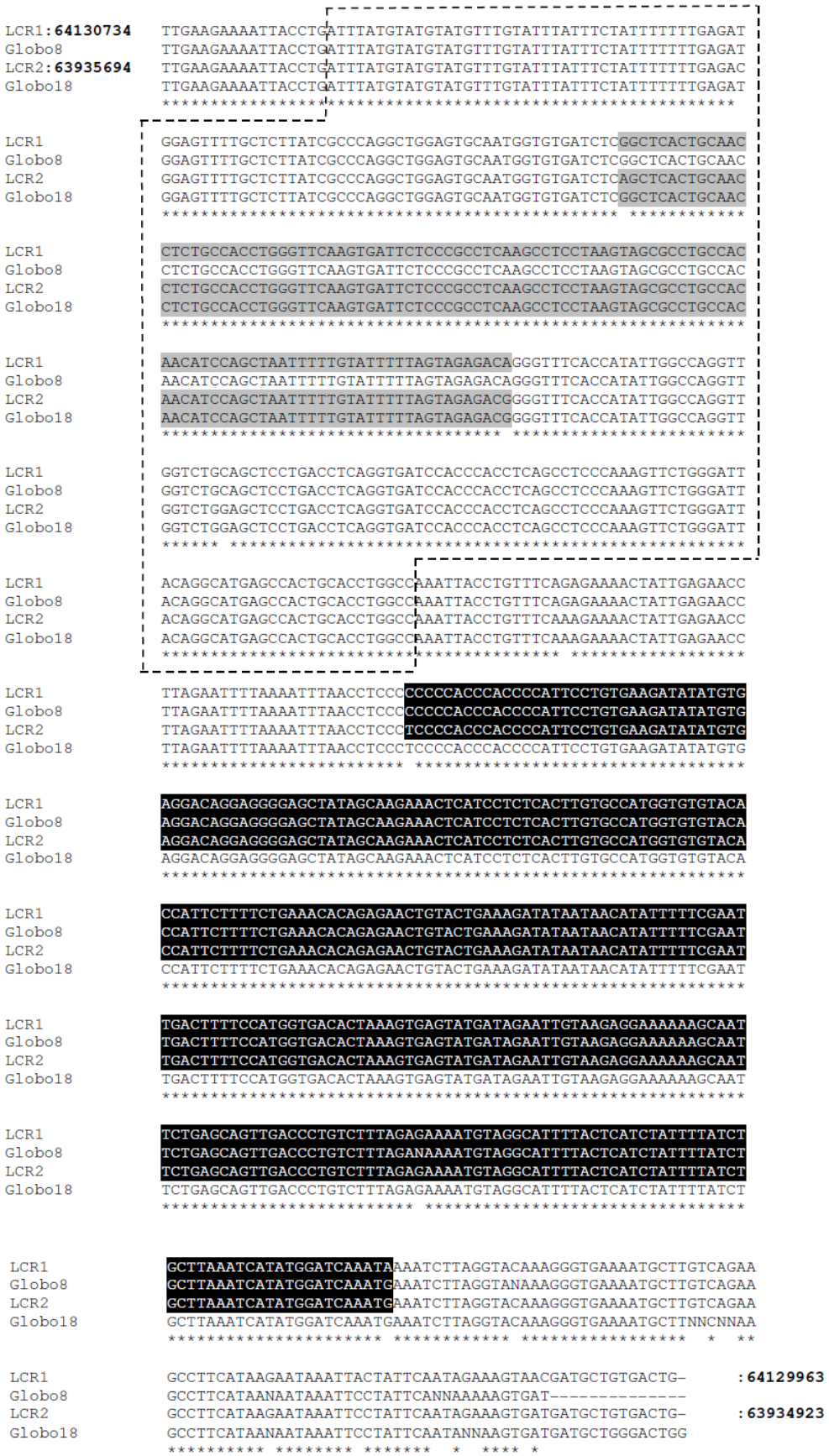

Figure S1. **Multiple Sequence Alignment of the Genomic Region Surrounding the Identified Breakpoints**

The alignment was performed using the clustalw 2.0 program[1] using as query sequences: the reference human sequence for both LCR1 (64 119 249-64 146 247 bp) and LCR2 (63 923 419-63 951 619 bp) and the sequences from 2 representative patients (Globo8 and Globo18) for each Breakpoint (BP) are presented. The exact genomic positions of both LCRs fragments shown are : LCR1-fragment: 64 129 963 - 64 130 734 bp; LCR2-fragment: 63 934 923 - 63 935 694 bp. Identical residues are marked with a star. Breakpoint 1 (BP1) is highlighted in gray and BP2 is highlighted in black. The block surrounded with dashed-lines indicates the AluSq2 repeat element.

Table S1. Results of the SNP Array for the Four Affected Brothers

Chromosome 12	Genotype Call in Sibling				
SNP	Globo1	Globo2	Globo3	Globo4	Globo18
rs345945	BB	BB	BB	BB	AA
rs505071	BB	BB	BB	BB	BB
rs2839798	BB	BB	BB	BB	AB
rs699603	AA	AA	AA	AA	AA
rs722918	AA	AA	AA	AA	BB
rs1146122	AA	AA	AA	AA	BB
rs722526	BB	BB	BB	BB	AA
rs1445442	BB	BB	BB	BB	BB
rs1373877	AA	AA	AA	AA	AA
rs4129000	BB	BB	BB	BB	BB
rs974349	AA	AA	AA	AA	AB
rs952562	BB	BB	BB	BB	AB
rs2029692	BB	BB	BB	BB	BB
rs1343807	BB	BB	BB	BB	BB
rs1405467	AA	AA	AA	AA	AA
rs1480065	BB	BB	BB	BB	AB
rs2870793	BB	BB	BB	BB	BB
rs3850590	AA	AA	AA	AA	AA
rs1905444	BB	BB	BB	BB	AB
rs722748	AA	AA	AA	AA	AB
rs722749	BB	BB	BB	BB	AB
rs2193047	AA	AA	AA	AA	BB
rs2870951	BB	BB	BB	BB	BB
rs2870950	AA	AA	AA	AA	AA
rs973328	BB	BB	BB	BB	AB
rs952218	BB	BB	BB	BB	AB
rs1908660	BB	BB	BB	BB	AB
rs1908682	AA	AA	AA	AA	AB
rs2172989	BB	BB	BB	BB	AB

The thirty homozygous SNP shared by the four infertile brothers which cover an 6.4 Mb region on the chromosome 12 and the analysis of Globo18 patient showing a region of homozygosity covering the *DPY19L2* locus, showed in grey.

Table S2. List of Genes in the Homozygous Region Shared between the Infertile Patients

Ensembl Gene ID	Description
ENSG00000251857	5S ribosomal RNA [Source:RFAM;Acc:RF00001]
ENSG00000240075	
ENSG00000252883	Small nucleolar RNA SNORD112 [Source:RFAM;Acc:RF01169]
ENSG00000238475	Small nucleolar RNA U13 [Source:RFAM;Acc:RF01210]
ENSG00000238440	Small nucleolar RNA U13 [Source:RFAM;Acc:RF01210]
ENSG00000223294	Small nucleolar RNA SNORD83 [Source:RFAM;Acc:RF00137]
ENSG00000238592	Small nucleolar RNA U13 [Source:RFAM;Acc:RF01210]
ENSG00000238528	Small nucleolar RNA U13 [Source:RFAM;Acc:RF01210]
ENSG00000206650	Small nucleolar RNA SNORA70 [Source:RFAM;Acc:RF00156]
ENSG00000200814	U6 spliceosomal RNA [Source:RFAM;Acc:RF00026]
ENSG00000202034	U6 spliceosomal RNA [Source:RFAM;Acc:RF00026]
ENSG00000200296	U1 spliceosomal RNA [Source:RFAM;Acc:RF00003]
ENSG00000212298	U6 spliceosomal RNA [Source:RFAM;Acc:RF00026]
ENSG00000251788	U5 spliceosomal RNA [Source:RFAM;Acc:RF00020]
ENSG00000221564	U6atac minor spliceosomal RNA [Source:RFAM;Acc:RF00619]
ENSG00000207099	U6 spliceosomal RNA [Source:RFAM;Acc:RF00026]
ENSG00000252770	U7 small nuclear RNA [Source:RFAM;Acc:RF00066]
ENSG00000241412	
ENSG00000245867	
ENSG00000246324	
ENSG00000250314	
ENSG00000250748	
ENSG00000248947	
ENSG00000247008	
ENSG00000245302	
ENSG00000251695	
ENSG00000247363	
ENSG00000111530	cullin-associated and neddylation-dissociated 1 [Source:HGNC Symbol;Acc:30688]
ENSG00000127334	dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 2 [Source:HGNC Symbol;Acc:3093]
ENSG00000111537	interferon, gamma [Source:HGNC Symbol;Acc:5438]
ENSG00000111536	interleukin 26 [Source:HGNC Symbol;Acc:17119]
ENSG00000127318	interleukin 22 [Source:HGNC Symbol;Acc:14900]
ENSG00000111554	Mdm1 nuclear protein homolog (mouse) [Source:HGNC Symbol;Acc:29917]
ENSG00000127314	RAP1B, member of RAS oncogene family [Source:HGNC Symbol;Acc:9857]
ENSG00000111581	nucleoporin 107kDa [Source:HGNC Symbol;Acc:29914]
ENSG00000175782	solute carrier family 35, member E3 [Source:HGNC Symbol;Acc:20864]
ENSG00000135679	Mdm2 p53 binding protein homolog (mouse) [Source:HGNC Symbol;Acc:6973]
ENSG00000226959	cDNA, FLJ92387 [Source:UniProtKB/TrEMBL;Acc:B2R594]
ENSG00000135678	carboxypeptidase M [Source:HGNC Symbol;Acc:2311]
ENSG00000111605	microRNA 1279 [Source:HGNC Symbol;Acc:35357]
ENSG00000090382	lysozyme [Source:HGNC Symbol;Acc:6740]

Ensembl Gene ID	Description
ENSG00000127337	YEATS domain containing 4 [Source:HGNC Symbol;Acc:24859]
ENSG00000166225	fibroblast growth factor receptor substrate 2 [Source:HGNC Symbol;Acc:16971]
ENSG00000166226	chaperonin containing TCP1, subunit 2 (beta) [Source:HGNC Symbol;Acc:1615]
ENSG00000198812	leucine rich repeat containing 10 [Source:HGNC Symbol;Acc:20264]
ENSG00000242922	
ENSG00000214304	keratin 8 pseudogene 19 [Source:HGNC Symbol;Acc:33371]
ENSG00000213363	
ENSG00000241941	
ENSG00000213352	
ENSG00000139239	
ENSG00000250517	
ENSG00000240027	
ENSG00000242087	
ENSG00000243024	ribosomal protein S11 [Source:HGNC Symbol;Acc:10384]
ENSG00000243020	
ENSG00000215208	
ENSG00000213344	
ENSG00000241749	ribosomal protein SA pseudogene 52 [Source:HGNC Symbol;Acc:35752]
ENSG00000213343	
ENSG00000228144	
ENSG00000225422	RNA binding motif, single stranded interacting protein 1 pseudogene 1 [Source:HGNC Symbol;Acc:9908]
ENSG00000244432	
ENSG00000236946	
ENSG00000240087	
ENSG00000241825	
ENSG00000241765	
ENSG00000198673	family with sequence similarity 19 (chemokine (C-C motif)-like), member A2 [Source:HGNC Symbol;Acc:21589]
ENSG00000135655	ubiquitin specific peptidase 15 [Source:HGNC Symbol;Acc:12613]
ENSG00000061987	MON2 homolog (S. cerevisiae) [Source:HGNC Symbol;Acc:29177]
ENSG00000221949	Putative uncharacterized protein C12orf61 [Source:UniProtKB/Swiss-Prot;Acc:Q8N7H1]
ENSG00000111110	protein phosphatase, Mg2+/Mn2+ dependent, 1H [Source:HGNC Symbol;Acc:18583]
ENSG00000166148	arginine vasopressin receptor 1A [Source:HGNC Symbol;Acc:895]
ENSG00000177990	Protein dpy-19 homolog 2 (Dpy-19-like protein 2) [Source:UniProtKB/Swiss-Prot;Acc:Q6NUT2]
ENSG00000118600	transmembrane protein 5 [Source:HGNC Symbol;Acc:13530]
ENSG00000196935	SLIT-ROBO Rho GTPase activating protein 1 [Source:HGNC Symbol;Acc:17382]
ENSG00000174206	UPF0536 protein C12orf66 [Source:UniProtKB/Swiss-Prot;Acc:Q96MD2]
ENSG00000185306	Uncharacterized protein C12orf56 [Source:UniProtKB/Swiss-Prot;Acc:Q8IXR9]
ENSG00000184575	exportin, tRNA (nuclear export receptor for tRNAs) [Source:HGNC Symbol;Acc:12826]
ENSG00000183735	TANK-binding kinase 1 [Source:HGNC Symbol;Acc:11584]
ENSG00000153179	Ras association (RalGDS/AF-6) domain family member 3 [Source:HGNC Symbol;Acc:14271]
ENSG00000135677	glucosamine (N-acetyl)-6-sulfatase [Source:HGNC Symbol;Acc:4422]
ENSG00000111490	TBC1 domain family, member 30 [Source:HGNC Symbol;Acc:29164]
ENSG00000156076	WNT inhibitory factor 1 [Source:HGNC Symbol;Acc:18081]

Ensembl Gene ID	Description
ENSG00000174106	LEM domain containing 3 [Source:HGNC Symbol;Acc:28887]
ENSG00000174099	methionine sulfoxide reductase B3 [Source:HGNC Symbol;Acc:27375]
ENSG00000149948	high mobility group AT-hook 2 [Source:HGNC Symbol;Acc:5009]
ENSG00000197301	Putative uncharacterized protein ENSP00000348547 [Source:UniProtKB/TrEMBL;Acc:B7WPI5]
ENSG00000139233	LLP homolog, long-term synaptic facilitation (Aplysia) [Source:HGNC Symbol;Acc:28229]
ENSG00000155957	transmembrane BAX inhibitor motif containing 4 [Source:HGNC Symbol;Acc:24257]
ENSG00000090376	interleukin-1 receptor-associated kinase 3 [Source:HGNC Symbol;Acc:17020]
ENSG00000127311	helicase (DNA) B [Source:HGNC Symbol;Acc:17196]
ENSG00000155974	glutamate receptor interacting protein 1 [Source:HGNC Symbol;Acc:18708]
ENSG00000199179	hsa-let-7i [Source:miRBase;Acc:MI0000434]
ENSG00000252801	
ENSG00000207546	hsa-mir-548c [Source:miRBase;Acc:MI0003630]
ENSG00000211577	
ENSG00000252547	
ENSG00000221422	
ENSG00000252660	Y RNA [Source:RFAM;Acc:RF00019]
ENSG00000222744	7SK RNA [Source:RFAM;Acc:RF00100]

The first column reports the Ensembl Gene number of the 101 genes contained in the homozygous region. From this, forty genes have been selected according to their testis expression and their putative role in the gametogenesis. These genes are presented in the second column.

Supplementary data

Chapter VII

Globozoospermia is mainly due to *DPY19L2* deletion via non-allelic homologous recombination involving two recombination hotspots

List of primers used to amplify *DPY19L2*

Exon	Primer	Annealing temperature	Sequence	Length
1	Forward	64,9 °C	gacccagctccaccatactcctt	504 pb
	Reverse	62,6 °C	ggccaacttctttctactcggac	
2	Forward	59,3 °C	gacaggattagctggccg	386 pb
	Reverse	54,8 °C	agcaaaaatattttaattcataagtg	
3	Forward	62,7 °C	gcaacgtacctggccacag	430 pb
	Reverse	59,5 °C	tatagcttatgaaacagtgcagttga	
4	Forward	53,9 °C	caaaatagcgagaagtgattag	414 pb
	Reverse	49,5 °C	ttctactcaactataaggatacac	
5	Forward	52,1 °C	agcttcatccatgtcactat	432 pb
	Reverse	53,6 °C	agccttctcagaaaactatttt	
6	Forward	53,9 °C	gggtaaataattaaacacagca	462 pb
	Reverse	53,9 °C	aaacaacagaataaaagggat	
7	Forward	52,1 °C	aatttatacgtacacttttagaatta	420 pb
	Reverse	57,3 °C	atttaaacatttcaatcaacatgc	
8	Forward	57,7 °C	tggacatggtagttaattgctg	371 pb
	Reverse	59,2 °C	tcccaaagtgctgaattgaa	
9	Forward	50,4 °C	ttaatggtaaattacaatgttaata	429 pb
	Reverse	53,7 °C	catggcatacatttacctaca	
10	Forward	50,0 °C	catccatcttttaattctg	479 pb
	Reverse	53,9 °C	atacatttccagttttccttct	
11	Forward	49,9 °C	ttggccaagagtcatt	516 pb
	Reverse	51,0 °C	aacctcctcaagtgacttag	
12	Forward	49,2 °C	ggtatttaagtgaggaaataat	422 pb
	Reverse	49,1 °C	tttaagacagtagctatttattaac	
13	Forward	53,1 °C	aatttttctatgtcatttgtagac	306 pb
	Reverse	57,1 °C	ccaataactcgtctagagaccttag	
14	Forward	50,4 °C	attactgtctataggctataatcac	475 pb
	Reverse	56,0 °C	ttccaagtggcctagattatc	
15	Forward	56,3 °C	attattttattaaggcatggaagac	315 pb
	Reverse	57,0 °C	aattctgagcaatttgcattc	
16	Forward	56,4 °C	ttctgctttcttataaacccct	453 pb
	Reverse	50,7 °C	tcatatttcctttcttaatcac	
17	Forward	51,1 °C	tctaagatcaagcaaatgaa	181 pb
	Reverse	54,2 °C	ctttgtcaattatcctcaaactac	
18	Forward	50,5 °C	gaatactgataggcttattactaag	331 pb
	Reverse	53,7 °C	ttagtcagcaaagccaca	
19	Forward	59,5 °C	tgagaatttattgttgaccctacg	544 pb
	Reverse	57,1 °C	tgtctgagagttaaaggacataactta	
20	Forward	56,6 °C	caatttctagccccaagatagt	505 pb
	Reverse	59,8 °C	tccagaggcaacaggtacg	
21	Forward	52,6 °C	ctgttttgagtcatgtatatcg	345 pb
	Reverse	50,7 °C	attcttaaagaatcaggactacta	

22	Forward	52,6 °C	attgtctctagacagcaatacat	313 pb
	Reverse	52,1 °C	gtgtctgttattaaagcttgtg	

Primers sequences			Annealing temperature	Length
DPY19L2-BP				
BPa	Forward	ATGCCATGTTGCCTGCT	64.1°C	1700bp
	Reverse	TCTTCTGGGAAAGGTATTATCGTAG		
BPb	Forward	CTCTATATCCATGAGAGACTGGCTCT	61.8°C	1500bp
	Reverse	CTTCCCCAAAAGAACATTATGGT		
BPc	Forward	GGGAGGAGCTCTTAACACCTGA	65°C	1500bp
	Reverse	CAAACAGCACACCTGGTCCT		

							Splice prediction									Missense prediction			
							MaxEnt			NNS			HSF			SIFT		PolyPhen	
Gene	Build	Transcript	Protein	Genomic position	Nt change	AA change	WildType	SNV Score	Variation	WildType	SNV Score	Variation	WildType	SNV Score	Variation	Prediction	Score	Prediction	Score
DPY19L2	hg19	NM_173812.4	NP_776173.3	g.64017904C>T	c.869G>A	p.Arg290His	6.46595	6.46595	0	0.918438	0.855584	6.84	89.07	89.07	0	Deleterious	0.01	Probably damaging	1.0
DPY19L2	hg19	NM_173812.4	NP_776173.3	g.63989801G>C	c.1478C>G	p.Thr493Arg	6.36196	6.36196	0	ND	ND	ND	81.489	81.489	0	Tolerated	0.38	Possibly damaging	0.708
DPY19L2	hg19	NM_173812.4	NP_776173.3	g.64017881G>A	c.892C>T	p.Arg298Cys	6.46595	6.46595	0	0.918438	0.918438	0	89.07	89.07	0	Deleterious	0.00	Probably damaging	1.0
DPY19L2	hg19	NM_173812.4	NP_776173.3	g.64011083C>T	c.1218+1G>A	-	10.2385	0	100	0.99324	0	100	93.162	0	100				

Supplementary Data Figure 1

Hotspot # 1

GGTAGGAGCGGTGCCCACTCAGACCCCAAAGTGGAGAGGTCACAGAAGCTGCTCCACTTG

GGGCAGTTCGCTGACTGCTTTCATCCCCGCTGTGGCTGCAGCTGGACTCCATTCCAGCCC

*

CAAGGGAATGTTTTGCTTCAGAGACTTCTCATCTGCATTCCCTTGCCTGAAATACTCTTT

CCTCAGATTCCTCAGCCTCCAAGTCTTTGCTTAAATATCACAATCTTGTTGATTTAGGCC

BP7

TCCTCTGAGCCCCCTATTTCAAAATGTCAGTCCCTTACCGTATTTCCTCCTTGTCTCTCT

TTCCTACTTTATTTTTCACCATAGCACTTACCGTCATCTGCCCCGGATGTCTTCCAGCCC

*

CAGCCCCATATTCACTTGCCACCGTCTCAGCCCTGCCACGTGTACCAGGAGGCTGACCGG

Hotspot # 2

*

BP9

*

GAAGGCAAAGGAGGAGCAAAAGCACGACTCACGTGGCGGCAGGCAAAAGGGCGTGTGCGG

*

BP8

AGGAACTCCTGTTTATAAAACCACCAGATCTCGCGAGACTTATTCACTACCATGAGAACA

*

GTATAGGGGAAACCACCCCCGTGATTCAATTATCTCCACCTGGCCCCAACCTTGACACGT

Junction sequence of exons 5-6 deletion (Globo 13)

ATGTTGGGTCAGCTTTCCCTTGGATGTAGAAAGCCAAAAAGAACATTGCTTGTGGCCTTC

CAGCAATGTCAGCAGCAACAACAACAAAAGCCAGACAGTCTTCAAATTTACAACTTTTCC

TGAAGCGGTCAGGAAGCTGAGGTCACAGGGCAATCAAATAACCCAAAATAAATTAAAAAG

ATAAAGCATGTGCAAGGAGAACAGAAACAAGCATTTAATTACATGGGTCATGCCACTAGA

CACAATGGTAAAGAATTAAACTCAAATTGTTCACCAGTCAGTCAGTAAATGCCTAGTATG

GGTTAGCAAGGGTATTTGCAATTGCTTACAGACATTTTAACACAGACCTCATAAGGTGCT

GACAAAAGATTAGCAAGAGTGCTCAGAAAGCATCCTCCATGGTAAAGCCTGGGAGAACAG

GAGCCACTTAGAGAAAGGCGTGAAACCCACCCCCCGACCCTTCCTTATCTCCCCTATAA**G**

GCTTACACATGTTCTTGTGATAGTTTAGTGAGAATGATGATTTCCAATTTCATCCATGTC

CCTACAAAGGACA**T**GAACTCATCATTTTTTATGGCTGCATAGTATTCCATGGTGTATATG

TGCCACATTTTCTGAATCCAGTCTATCATTGTTGGACATTTGGGTTGGTTCCAAGTCTTT

GCTATTGTGAATAGTGCCACAATANACATATGTGTGCATGTGTCTTTATAGCAGCATGAT

TTATAATCCTTTGGGTATATACCCAGTAATGGGATGGCTGGGTCANATGGTATTTCTAGT

TCTAGATCTTTGAGGAATTGCCACACTGTCTTCCACAGTGGTTGAACTAGTTTACAGTCC

CACCAACAGTGTANAAGTGTTCCTATTTCTCCACATCCTCTCCGGCACC

BPs Genomic positions on LCR 2 (hg 19)

BP1 : 63935480 63935586
BP2 : 63935013 63935308
BP3 : 63935636 63935727
BP4 : 63935449 63935478
BP5 : 63935310 63935352
BP6 : 63935354 63935447
BP7 : 63945428 63945477
BP8 : 63945361 63945374
BP9 : 63936714 63936978

BPs Genomic positions on LCR 1 (hg19)

BP1 : 64130520 64130626
BP2 : 64130053 64130348
BP3 : 64130676 64130767
BP4 : 64130489 64130518
BP5 : 64130350 64130392
BP6 : 64130394 64130487
BP7 : 64140492 64140541
BP8 : 64140426 64140439
BP9 : 64131762 64132026

Figure 1: Sequences of the Genomic Region Surrounding the Identified Breakpoints.
Mismatches between the two LCRs are marked with a star.
In the junction sequence, the sequence before of the highlighted nucleotide G (g.64049979) corresponds to the last nucleotide of exon 4 while the highlighted nucleotide T (g.64033425) corresponds to the first nucleotide of exon 6. The sequence between these two nucleotides corresponds to 73bp of a LINE element.

Supplementary data

Chapter VIII

Confirmation of the pathogenicity of *SPATA16* mutations and identification of its partners

Table 1: List of primers

	PICK1		
1	Forward GGTTAGACGCTGTCAGCCTGT Reverse CAAGTGCCTAAATGCCAACG	55,9	508
2	Forward GGCGTTGGCATTTAGGCACT Reverse ACCGCTGAGAAGCAAGGGTC	64,1	258
3	Forward CAGTGGAGCCCCTCAGGA Reverse CAGGTGGTCAGAAAGCCCCT	64,1	341
4	Forward GAGCAGAGGGTAGAGTGGAAGACAG Reverse ACAAGGAAGGGGGCGGTG	64,1	358
5	Forward AGGAGTCTCAGTCCAGAACAGTCTT Reverse TTGGTCAGAGGTCAGAGCC	64,1	301
6	Forward TTGGTCAGAGGTCAGAGCC Reverse TGGTGACTTCTCAGTTCCACG	58,2	268
7	Forward CACCTTTGAACCCAGCTGA Reverse GTAGGCTGGCATTCCTAGTG	55,9	323
8	Forward CCTGGGCCAGTGTGGTC Reverse CTCAGCATGGGAGTGAGAGG	60,6	193
9	Forward CTCTTGGAGGTGTCAATCCA Reverse AGAGGTCAATGGCAAGGC	63	515
10	Forward ATTCATTGTCACCCTGGAC Reverse TAACCCATTGTTTAAAGGCAACAA	55,9	418
11	Forward AACACCTGCGCCAGCT Reverse TCAAACACTTAAAAGCGCCT	58,2	417
12	Forward GAACAGCCGTGGCTTTGA Reverse AGGCCAGGCTCTGGGCA	61,8	363
13	Forward TCAGGGAAGTGCCGAGTG Reverse TTCGCTGTGCGAGGACAGA	58,2	542
	AGFG1		
1	Forward CAGTACAGCCAAGCCGCTG Reverse AAGGCTGCCCAGAACGTAA	58,2	571
2	Forward CACTGTTTCAAGTTTGTGTGTTGG Reverse GACCGTGAAATTCAACGTTT	58,2	301
3	Forward GCTTCCTGAAAGAATCATAGTATC Reverse ATCCTTCATTGTGTCAGACTAACCT	58,2	289
4	Forward TATTCTGTGTGAGAGGGTTTTCAG Reverse ATGCTACTGGCATACAAGATTATG	58,2	405
5	Forward CTCAGCATACTGTAACTCTTCGTATGT Reverse GACATTACATCCTTTCCCCATT	58,2	513
6	Forward TACCTGTAATTGGGTTACATGCTG Reverse CGCAGGAAAAGGAGGAGAAT	58,2	533

7	Forward CGTCACTTGTAGTCAAAAGCCA Reverse AAATTTCTATACAACAACTGAGAAAAG	58,2	480
8	Forward CTAAATCCATTTTAAGTTTACCTGTGTTT Reverse GATAAAACACAAAGCAGGAGGG	59,4	324
9	Forward CTGATGCTGTGATACACGATATGTT Reverse TATCAAGACAGAAAACTCAAGGATTC	59,4	309
10	Forward CCCAAATATATATTGTGTAGCATTTTC Reverse GTTCCATTAAGACTGCATGAGTGA	60,6	573
11	Forward GGAATTTTGGAATTAACCCTGC Reverse TCTGAGAATGACGGGATGCT	60,6	546
12	Forward GCTGCTCTGGTAGGAATGTCTC Reverse TATCAATATAACACCAAATACCATCTAG	60,6	303
13	Forward CTAATTAAAAAAATTAAATTCAGTTTAATG Reverse GGTTAGAAGGTGTTACTGGAGCC	58,2	324
DNAJA4			
1	Forward ACAGTTGTCGGAGGGC Reverse AAGCTTCCGCAAGGAGAC		256
2	Forward GGGCGGGAGCTACAAGC Reverse GCTTAATAACCCTTGGCCCC		341
3	Forward CAGGTTTGCTTAACAAAACAACTG Reverse ATCCTGATTCCAACAACACG		315
4	Forward AACTAAGGGTCATATGGGTGG Reverse GCATTGTTTCTGAATCTAAGGG		245
5	Forward CCAGCAGGGACACTGGTAAG Reverse ACGATTAGCACCCAGACAGC		351
6	Forward GCATGAGGTTGGGAGACTTG Reverse CAAAGTATCACATGCACCCC		480
7-8	Forward GAGGGAGAACGACCTCTGC Reverse CGATGTCATTTTCATTGTGCC		556
IRGC1			
1	Forward AGAAAGATCTCTCTGGAGTCTTCA Reverse AAGACATTGTTGCCTCCAGAT		489
2a	Forward CTTCACATCCGTGGTATCTGTC Reverse CAGGTCCTCGTCCACCTTG		668
2b	Forward AAGCAGGTAGACTTCAGCCG Reverse GGTCACCTGCCTGTTTGC		598
2c	Forward TTGCTCATCCACTCACTG Reverse GTTGGTAATTTGGATTTTGT		581
TCP1			
1	Forward GCGATACTGGTTGCCCG Reverse TTTCTGCGGAGGTAAAGGAG		225

2-3	Forward GCTGTTGGATGTCATTTACCTTTA Reverse TGACCCACTTATGGCTTCAAC		574
4	Forward CACGCCCAGGTTTCTAATAAC Reverse AATCAAAACGTTAAGACAACACAG		256
5	Forward TGATGAATGCAAAGAGTTTGAGTAG Reverse ACATACCACCATGCCCATC		249
6	Forward TTGTAACATTTTGGTGACAAGTAATG Reverse CAGTGATTCTCCTTCCTCAGC		588
7	Forward AATGCAATGAATTAGTCTAAATTT Reverse GAAGTAAAATAGTTTGACTTGGCT		324
8	Forward ATTTGGGTTGGGGAGAGCT Reverse AAATGTGTTGGTAGCACAAAGC		513
9	Forward CAAATGCTACTGAGTCCAAGG Reverse ACAAGCAGCAAATAAATGGG		284
10-11	Forward TGAGAAATTGATGGTCTTGGC Reverse GACAAGTCTGTTACTTATGTTTGCTGAC		638
12	Forward ACCTACAAGATCTCTAGTCTGA Reverse CAGAGGACCAAAGTACAGATGG		486

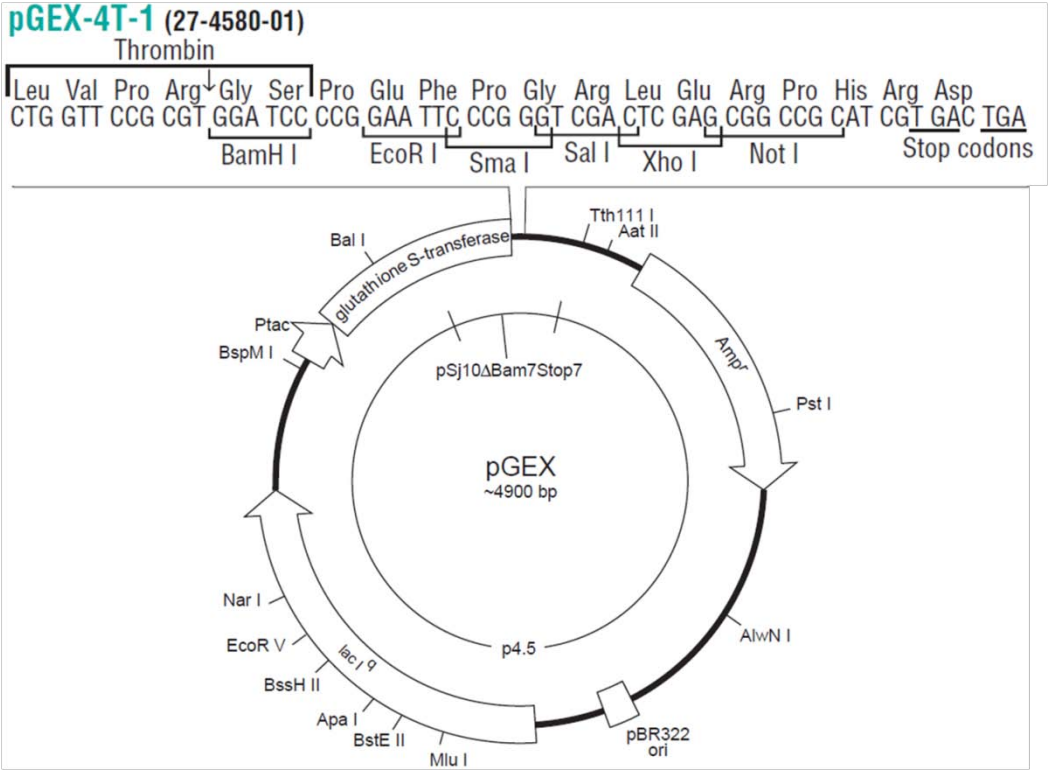

Figure 1: Map of the gluthatione S-transferase fusion vector.

References

Abdul, K. M., K. Terada, et al. (2002). "Characterization and functional analysis of a heart-enriched DnaJ/ Hsp40 homolog dj4/DjA4." Cell Stress Chaperones **7**(2): 156-166.

Abou-Haila, A. and D. R. Tulsiani (2000). "Mammalian sperm acrosome: formation, contents, and function." Arch Biochem Biophys **379**(2): 173-182.

Adzhubei, I. A., S. Schmidt, et al. (2010). "A method and server for predicting damaging missense mutations." Nature Methods **7**(4): 248-249.

Amann, R. P. (2008). "The cycle of the seminiferous epithelium in humans: a need to revisit?" J Androl **29**(5): 469-487.

Anakwe, O. O. and G. L. Gerton (1990). "Acrosome biogenesis begins during meiosis: evidence from the synthesis and distribution of an acrosomal glycoprotein, acrogranin, during guinea pig spermatogenesis." Biol Reprod **42**(2): 317-328.

Aoki, V. W., G. L. Christensen, et al. (2006). "Identification of novel polymorphisms in the nuclear protein genes and their relationship with human sperm protamine deficiency and severe male infertility." Fertil Steril **86**(5): 1416-1422.

Aponte, P. M., M. P. van Bragt, et al. (2005). "Spermatogonial stem cells: characteristics and experimental possibilities." APMIS **113**(11-12): 727-742.

Aston, K. I. and D. T. Carrell (2009). "Genome-wide study of single-nucleotide polymorphisms associated with azoospermia and severe oligozoospermia." J Androl **30**(6): 711-725.

Aston, K. I., C. Krausz, et al. (2010). "Evaluation of 172 candidate polymorphisms for association with oligozoospermia or azoospermia in a large cohort of men of European descent." Hum Reprod **25**(6): 1383-1397.

Avenarius, M. R., M. S. Hildebrand, et al. (2009). "Human male infertility caused by mutations in the CATSPER1 channel protein." Am J Hum Genet **84**(4): 505-510.

Bailey, J. A. and E. E. Eichler (2006). "Primate segmental duplications: crucibles of evolution, diversity and disease." Nat Rev Genet **7**(7): 552-564.

Bailey, J. A., G. Liu, et al. (2003). "An Alu transposition model for the origin and expansion of human segmental duplications." Am J Hum Genet **73**(4): 823-834.

Bailey, J. A., A. M. Yavor, et al. (2001). "Segmental duplications: organization and impact within the current human genome project assembly." Genome Res **11**(6): 1005-1017.

Balabanic, D., M. Rupnik, et al. (2011). "Negative impact of endocrine-disrupting compounds on human reproductive health." Reprod Fertil Dev **23**(3): 403-416.

Ballow, D., M. L. Meistrich, et al. (2006). "Sohlh1 is essential for spermatogonial differentiation." Dev Biol **294**(1): 161-167.

Bamshad, M. J., S. B. Ng, et al. (2011). "Exome sequencing as a tool for Mendelian disease gene discovery." Nat Rev Genet **12**(11): 745-755.

Banker, M. R., P. M. Patel, et al. (2009). "Successful pregnancies and a live birth after

intracytoplasmic sperm injection in globozoospermia." J Hum Reprod Sci **2**(2): 81-82.

Bashamboo, A., B. Ferraz-de-Souza, et al. (2010). "Human male infertility associated with mutations in NR5A1 encoding steroidogenic factor 1." Am J Hum Genet **87**(4): 505-512.

Batzer, M. A. and P. L. Deininger (2002). "Alu repeats and human genomic diversity." Nat Rev Genet **3**(5): 370-379.

Baudat, F., J. Buard, et al. (2010). "PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice." Science **327**(5967): 836-840.

Baudat, F. and A. Nicolas (1997). "Clustering of meiotic double-strand breaks on yeast chromosome III." Proc Natl Acad Sci U S A **94**(10): 5213-5218.

Beauregard, A., M. J. Curcio, et al. (2008). "The take and give between retrotransposable elements and their hosts." Annu Rev Genet **42**: 587-617.

Bechoua, S., A. Chiron, et al. (2009). "Fertilisation and pregnancy outcome after ICSI in globozoospermic patients without assisted oocyte activation." Andrologia **41**(1): 55-58.

Ben Khelifa, M., R. Zouari, et al. (2011). "A new AURKC mutation causing macrozoospermia: implications for human spermatogenesis and clinical diagnosis." Mol Hum Reprod **17**(12): 762-768.

Benet, J., M. Oliver-Bonet, et al. (2005). "Segregation of chromosomes in sperm of reciprocal translocation carriers: a review." Cytogenet Genome Res **111**(3-4): 281-290.

Berg, I. L., R. Neumann, et al. (2010). "PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans." Nat Genet **42**(10): 859-863.

Berruti, G. and C. Paiardi (2011). "Acrosome biogenesis: Revisiting old questions to yield new insights." Spermatogenesis **1**(2): 95-98.

Bhalla, N. and A. F. Dernburg (2008). "Prelude to a division." Annu Rev Cell Dev Biol **24**: 397-424.

Bi, W., S. S. Park, et al. (2003). "Reciprocal crossovers and a positional preference for strand exchange in recombination events resulting in deletion or duplication of chromosome 17p11.2." Am J Hum Genet **73**(6): 1302-1315.

Bier, F. F., M. von Nickisch-Rosenegk, et al. (2008). "DNA microarrays." Adv Biochem Eng Biotechnol **109**: 433-453.

Blatch, G. L. and M. Lassle (1999). "The tetratricopeptide repeat: a structural motif mediating protein-protein interactions." Bioessays **21**(11): 932-939.

Bloch-Zupan, A., X. Jamet, et al. (2011). "Homozygosity mapping and candidate prioritization identify mutations, missed by whole-exome sequencing, in SMOC2, causing major dental developmental defects." Am J Hum Genet **89**(6): 773-781.

Boivin, J., L. Bunting, et al. (2007). "International estimates of infertility prevalence and treatment-seeking: potential need and demand for infertility medical care." Hum Reprod **22**(6): 1506-

1512.

Bourc'his, D. and T. H. Bestor (2004). "Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L." Nature **431**(7004): 96-99.

Brick, K., F. Smagulova, et al. (2012). "Genetic recombination is directed away from functional genomic elements in mice." Nature **485**(7400): 642-645.

Buard, J., P. Barthes, et al. (2009). "Distinct histone modifications define initiation and repair of meiotic recombination in the mouse." EMBO J **28**(17): 2616-2624.

Carlson, A. E., R. E. Westenbroek, et al. (2003). "CatSper1 required for evoked Ca2+ entry and control of flagellar function in sperm." Proc Natl Acad Sci U S A **100**(25): 14864-14868.

Carrell, D. T. (2012). "Epigenetics of the male gamete." Fertil Steril **97**(2): 267-274.

Carrell, D. T., B. R. Emery, et al. (1999). "Characterization of aneuploidy rates, protamine levels, ultrastructure, and functional ability of round-headed sperm from two siblings and implications for intracytoplasmic sperm injection." Fertil Steril **71**(3): 511-516.

Carson, A. R., J. Cheung, et al. (2006). "Duplication and relocation of the functional DPY19L2 gene within low copy repeats." BMC Genomics **7**: 45.

Chi, H. J., J. J. Koo, et al. (2004). "Successful fertilization and pregnancy after intracytoplasmic sperm injection and oocyte activation with calcium ionophore in a normozoospermic patient with extremely low fertilization rates in intracytoplasmic sperm injection cycles." Fertility and sterility **82**(2): 475-477.

Choi, Y., S. Jeon, et al. (2010). "Mutations in SOHLH1 gene associate with nonobstructive azoospermia." Hum Mutat **31**(7): 788-793.

Christensen, G. L., I. P. Ivanov, et al. (2006). "Identification of polymorphisms in the Hrb, GOPC, and Csnk2a2 genes in two men with globozoospermia." J Androl **27**(1): 11-15.

Clarke, L., X. Zheng-Bradley, et al. "The 1000 Genomes Project: data management and community access." Nat Methods **9**(5): 459-462.

Clermont, Y. (1966). "Renewal of spermatogonia in man." Am J Anat **118**(2): 509-524.

Clermont, Y. (1966). "Spermatogenesis in man. A study of the spermatogonial population." Fertil Steril **17**(6): 705-721.

Clermont, Y. (1972). "Kinetics of spermatogenesis in mammals: seminiferous epithelium cycle and spermatogonial renewal." Physiol Rev **52**(1): 198-236.

Clermont, Y. and C. P. Leblond (1955). "Spermiogenesis of man, monkey, ram and other mammals as shown by the periodic acid-Schiff technique." Am J Anat **96**(2): 229-253.

Cobb, J. and M. A. Handel (1998). "Dynamics of meiotic prophase I during spermatogenesis: from pairing to division." Semin Cell Dev Biol **9**(4): 445-450.

Coetzee, K., M. L. Windt, et al. (2001). "An intracytoplasmic sperm injection pregnancy with a

globozoospermic male." J Assist Reprod Genet **18**(5): 311-313.

Cohen, P. E., S. E. Pollack, et al. (2006). "Genetic analysis of chromosome pairing, recombination, and cell cycle control during first meiotic prophase in mammals." Endocr Rev **27**(4): 398-426.

Consortium, T. I. H. (2003). "The International HapMap Project." Nature **426**(6968): 789-796.

Coop, G., X. Wen, et al. (2008). "High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans." Science **319**(5868): 1395-1398.

Dam, A. H., I. Feenstra, et al. (2007). "Globozoospermia revisited." Hum Reprod Update **13**(1): 63-75.

Dam, A. H., I. Koscinski, et al. (2007). "Homozygous mutation in SPATA16 is associated with male infertility in human globozoospermia." Am J Hum Genet **81**(4): 813-820.

de Graaff, E., P. Rouillard, et al. (1995). "Hotspot for deletions in the CGG repeat region of FMR1 in fragile X patients." Hum Mol Genet **4**(1): 45-49.

de Kretser, D. M., K. L. Loveland, et al. (1998). "Spermatogenesis." Hum Reprod **13 Suppl 1**: 1-8.

Dej, K. J. and T. L. Orr-Weaver (2000). "Separation anxiety at the centromere." Trends Cell Biol **10**(9): 392-399.

Dettin, L., N. Ravindranath, et al. (2003). "Morphological characterization of the spermatogonial subtypes in the neonatal mouse testis." Biol Reprod **69**(5): 1565-1571.

Dieterich, K., R. Soto Rifo, et al. (2007). "Homozygous mutation of AURKC yields large-headed polyploid spermatozoa and causes male infertility." Nat Genet **39**(5): 661-665.

Dieterich, K., R. Zouari, 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**(7): 1301-1309.

Dirican, E. K., A. Isik, et al. (2008). "Clinical pregnancies and livebirths achieved by intracytoplasmic injection of round headed acrosomeless spermatozoa with and without oocyte activation in familial globozoospermia: case report." Asian J Androl **10**(2): 332-336.

Duman, J. G. and J. G. Forte (2003). "What is the role of SNARE proteins in membrane fusion?" Am J Physiol Cell Physiol **285**(2): C237-249.

Dym, M. and D. W. Fawcett (1970). "The blood-testis barrier in the rat and the physiological compartmentation of the seminiferous epithelium." Biol Reprod **3**(3): 308-326.

Ehmcke, J., C. M. Luetjens, et al. (2005). "Clonal organization of proliferating spermatogonial stem cells in adult males of two species of non-human primates, Macaca mulatta and Callithrix jacchus." Biol Reprod **72**(2): 293-300.

Ehmcke, J. and S. Schlatt (2006). "A revised model for spermatogonial expansion in man: lessons from non-human primates." Reproduction **132**(5): 673-680.

Ehmcke, J., D. R. Simorangkir, et al. (2005). "Identification of the starting point for spermatogenesis and characterization of the testicular stem cell in adult male rhesus monkeys." Hum Reprod

20(5): 1185-1193.

Ehmcke, J., J. Wistuba, et al. (2006). "Spermatogonial stem cells: questions, models and perspectives." Hum Reprod Update **12**(3): 275-282.

Eldar-Geva, T., B. Brooks, et al. (2003). "Successful pregnancy and delivery after calcium ionophore oocyte activation in a normozoospermic patient with previous repeated failed fertilization after intracytoplasmic sperm injection." Fertility and sterility **79 Suppl 3**: 1656-1658.

Elliott, D. J. and H. J. Cooke (1997). "The molecular genetics of male infertility." Bioessays **19**(9): 801-809.

Feldman, H. A., I. Goldstein, et al. (1994). "Impotence and its medical and psychosocial correlates: results of the Massachusetts Male Aging Study." J Urol **151**(1): 54-61.

Ferlin, A., F. Raicu, et al. (2007). "Male infertility: role of genetic background." Reprod Biomed Online **14**(6): 734-745.

Fritz, I. B. (1994). "Somatic cell-germ cell relationships in mammalian testes during development and spermatogenesis." Ciba Found Symp **182**: 271-274; discussion 274-281.

Gardner, S. M., K. Takamiya, et al. (2005). "Calcium-permeable AMPA receptor plasticity is mediated by subunit-specific interactions with PICK1 and NSF." Neuron **45**(6): 903-915.

Gianotten, J., A. W. Schimmel, et al. (2004). "Absence of mutations in the PCI gene in subfertile men." Mol Hum Reprod **10**(11): 807-813.

Gilissen, C., H. H. Arts, et al. (2010). "Exome sequencing identifies WDR35 variants involved in Sensenbrenner syndrome." Am J Hum Genet **87**(3): 418-423.

Grant, S. F. and H. Hakonarson (2008). "Microarray technology and applications in the arena of genome-wide association." Clin Chem **54**(7): 1116-1124.

Gu, W., F. Zhang, et al. (2008). "Mechanisms for human genomic rearrangements." Pathogenetics **1**(1): 4.

Harbuz, R., R. Zouari, 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**(3): 351-361.

Hastings, P. J., G. Ira, et al. (2009). "A microhomology-mediated break-induced replication model for the origin of human copy number variation." PLoS Genet **5**(1): e1000327.

Hastings, P. J., J. R. Lupski, et al. (2009). "Mechanisms of change in gene copy number." Nat Rev Genet **10**(8): 551-564.

Hayashi, K., K. Yoshida, et al. (2005). "A histone H3 methyltransferase controls epigenetic events required for meiotic prophase." Nature **438**(7066): 374-378.

Heard, E. and J. Turner (2011). "Function of the sex chromosomes in mammalian fertility." Cold Spring Harb Perspect Biol **3**(10): a002675.

Heindryckx, B., S. De Gheselle, et al. (2008). "Efficiency of assisted oocyte activation as a solution for

failed intracytoplasmic sperm injection." Reproductive biomedicine online **17**(5): 662-668.

Heindryckx, B., J. Van der Elst, et al. (2005). "Treatment option for sperm- or oocyte-related fertilization failure: assisted oocyte activation following diagnostic heterologous ICSI." Human reproduction **20**(8): 2237-2241.

Heller, C. G. and Y. Clermont (1963). "Spermatogenesis in man: an estimate of its duration." Science **140**(3563): 184-186.

Hermo, L. and Y. Clermont (1976). "Light cells within the limiting membrane of rat seminiferous tubules." Am J Anat **145**(4): 467-483.

Hermo, L. and J. Dworkin (1988). "Transitional cells at the junction of seminiferous tubules with the rete testis of the rat: their fine structure, endocytic activity, and basement membrane." Am J Anat **181**(2): 111-131.

Hermo, L., R. M. Pelletier, et al. (2010). "Surfing the wave, cycle, life history, and genes/proteins expressed by testicular germ cells. Part 3: developmental changes in spermatid flagellum and cytoplasmic droplet and interaction of sperm with the zona pellucida and egg plasma membrane." Microsc Res Tech **73**(4): 320-363.

Heyer, W. D., K. T. Ehmsen, et al. (2010). "Regulation of homologous recombination in eukaryotes." Annu Rev Genet **44**: 113-139.

Heytens, E., J. Parrington, et al. (2009). "Reduced amounts and abnormal forms of phospholipase C zeta (PLCzeta) in spermatozoa from infertile men." Hum Reprod **24**(10): 2417-2428.

Hoischen, A., B. W. van Bon, et al. (2010). "De novo mutations of SETBP1 cause Schinzel-Giedion syndrome." Nat Genet **42**(6): 483-485.

Hu, Z., Y. Xia, et al. (2011). "A genome-wide association study in Chinese men identifies three risk loci for non-obstructive azoospermia." Nat Genet.

Huckins, C. and Y. Clermont (1968). "Evolution of gonocytes in the rat testis during late embryonic and early post-natal life." Arch Anat Histol Embryol **51**(1): 341-354.

Huynh, T., R. Mollard, et al. (2002). "Selected genetic factors associated with male infertility." Hum Reprod Update **8**(2): 183-198.

Ikawa, M., N. Inoue, et al. (2010). "Fertilization: a sperm's journey to and interaction with the oocyte." J Clin Invest **120**(4): 984-994.

Iliakis, G., H. Wang, et al. (2004). "Mechanisms of DNA double strand break repair and chromosome aberration formation." Cytogenet Genome Res **104**(1-4): 14-20.

Inoue, K. and J. R. Lupski (2002). "Molecular mechanisms for genomic disorders." Annu Rev Genomics Hum Genet **3**: 199-242.

Jamsai, D. and M. K. O'Bryan (2011). "Mouse models in male fertility research." Asian J Androl **13**(1): 139-151.

Jan, S. Z., G. Hamer, et al. (2012). "Molecular control of rodent spermatogenesis." Biochim Biophys

Acta.

Jefferys, A., D. Siassakos, et al. (2012). "The management of retrograde ejaculation: a systematic review and update." Fertil Steril **97**(2): 306-312.

Jin, J., N. Jin, et al. (2007). "Catsper3 and Catsper4 are essential for sperm hyperactivated motility and male fertility in the mouse." Biol Reprod **77**(1): 37-44.

Juneja, S. C. and J. M. van Deursen (2005). "A mouse model of familial oligoasthenoteratozoospermia." Hum Reprod **20**(4): 881-893.

Kamal, R., R. S. Gupta, et al. (2003). "Plants for male fertility regulation." Phytother Res **17**(6): 579-590.

Kang-Decker, N., G. T. Mantchev, et al. (2001). "Lack of acrosome formation in Hrb-deficient mice." Science **294**(5546): 1531-1533.

Kashir, J., B. Heindryckx, et al. (2010). "Oocyte activation, phospholipase C zeta and human infertility." Hum Reprod Update **16**(6): 690-703.

Kasturi, M., B. Manivannan, et al. (1995). "Changes in epididymal structure and function of albino rat treated with Azadirachta indica leaves." Indian J Exp Biol **33**(10): 725-729.

Kauppi, L., A. J. Jeffreys, et al. (2004). "Where the crossovers are: recombination distributions in mammals." Nat Rev Genet **5**(6): 413-424.

Kidd, J. M., G. M. Cooper, et al. (2008). "Mapping and sequencing of structural variation from eight human genomes." Nature **453**(7191): 56-64.

Kierszenbaum, A. L. and L. L. Tres (2004). "The acrosome-acroplaxome-manchette complex and the shaping of the spermatid head." Arch Histol Cytol **67**(4): 271-284.

Kilani, Z., R. Ismail, et al. (2004). "Evaluation and treatment of familial globozoospermia in five brothers." Fertil Steril **82**(5): 1436-1439.

Kimmins, S., C. Crosio, et al. (2007). "Differential functions of the Aurora-B and Aurora-C kinases in mammalian spermatogenesis." Mol Endocrinol **21**(3): 726-739.

Kobayashi, H., T. Sakurai, et al. (2012). "Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks." PLoS Genet **8**(1): e1002440.

Kolodner, R. D., C. D. Putnam, et al. (2002). "Maintenance of genome stability in Saccharomyces cerevisiae." Science **297**(5581): 552-557.

Kopera, I. A., B. Bilinska, et al. (2010). "Sertoli-germ cell junctions in the testis: a review of recent data." Philos Trans R Soc Lond B Biol Sci **365**(1546): 1593-1605.

Korbel, J. O., A. E. Urban, et al. (2007). "Paired-end mapping reveals extensive structural variation in the human genome." Science **318**(5849): 420-426.

Koscinski, I., E. Elinati, et al. (2011). "DPY19L2 deletion as a major cause of globozoospermia." Am J

Hum Genet **88**(3): 344-350.

Krachler, A. M., A. Sharma, et al. (2010). "Self-association of TPR domains: Lessons learned from a designed, consensus-based TPR oligomer." Proteins **78**(9): 2131-2143.

Kraus, E., W. Y. Leung, et al. (2001). "Break-induced replication: a review and an example in budding yeast." Proc Natl Acad Sci U S A **98**(15): 8255-8262.

Krausz, C. (2010). "Male infertility: pathogenesis and clinical diagnosis." Best Pract Res Clin Endocrinol Metab **25**(2): 271-285.

Krausz, C. (2011). "Male infertility: pathogenesis and clinical diagnosis." Best Pract Res Clin Endocrinol Metab **25**(2): 271-285.

Krausz, C. and G. Forti (2000). "Clinical aspects of male infertility." Results Probl Cell Differ **28**: 1-21.

Krawitz, P. M., M. R. Schweiger, et al. (2010). "Identity-by-descent filtering of exome sequence data identifies PIGV mutations in hyperphosphatasia mental retardation syndrome." Nat Genet **42**(10): 827-829.

Kumar, P., S. Henikoff, et al. (2009). "Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm." Nature Protocols **4**(7): 1073-1082.

Kurahashi, H., T. Shaikh, et al. (2003). "The constitutional t(17;22): another translocation mediated by palindromic AT-rich repeats." Am J Hum Genet **72**(3): 733-738.

Lander, E. S., L. M. Linton, et al. (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.

Leblond, C. P. and Y. Clermont (1952). "Definition of the stages of the cycle of the seminiferous epithelium in the rat." Ann N Y Acad Sci **55**(4): 548-573.

Lee, J. A., K. Inoue, et al. (2006). "Role of genomic architecture in PLP1 duplication causing Pelizaeus-Merzbacher disease." Hum Mol Genet **15**(14): 2250-2265.

Lee, J. Y., R. Dada, et al. (2011). "Role of genetics in azoospermia." Urology **77**(3): 598-601.

Li, Y. C., X. Q. Hu, et al. (2006). "Afaf, a novel vesicle membrane protein, is related to acrosome formation in murine testis." FEBS Lett **580**(17): 4266-4273.

Liao, H. F., K. Y. Tai, et al. (2012). "Functions of DNA methyltransferase 3-like in germ cells and beyond." Biol Cell.

Lieber, M. R., Y. Ma, et al. (2003). "Mechanism and regulation of human non-homologous DNA end-joining." Nat Rev Mol Cell Biol **4**(9): 712-720.

Lin, L., P. Philibert, et al. (2007). "Heterozygous missense mutations in steroidogenic factor 1 (SF1/Ad4BP, NR5A1) are associated with 46,XY disorders of sex development with normal adrenal function." J Clin Endocrinol Metab **92**(3): 991-999.

Liu, G., Q. W. Shi, et al. (2010). "A newly discovered mutation in PICK1 in a human with

globozoospermia." Asian J Androl **12**(4): 556-560.

Lopez-Correa, C., M. Dorschner, et al. (2001). "Recombination hotspot in NF1 microdeletion patients." Hum Mol Genet **10**(13): 1387-1392.

Lourenco, D., R. Brauner, et al. (2009). "Mutations in NR5A1 associated with ovarian insufficiency." N Engl J Med **360**(12): 1200-1210.

Luan, D. D., M. H. Korman, et al. (1993). "Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition." Cell **72**(4): 595-605.

Lupski, J. R. (1998). "Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits." Trends Genet **14**(10): 417-422.

Lupski, J. R. and P. Stankiewicz (2005). "Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes." PLoS Genet **1**(6): e49.

Machev, N., P. Gosset, et al. (2005). "Chromosome abnormalities in sperm from infertile men with normal somatic karyotypes: teratozoospermia." Cytogenet Genome Res **111**(3-4): 352-357.

Martinez-Perez, E. and M. P. Colaiacovo (2009). "Distribution of meiotic recombination events: talking to your neighbors." Curr Opin Genet Dev **19**(2): 105-112.

Massart, A., W. Lissens, et al. (2012). "Genetic causes of spermatogenic failure." Asian J Androl **14**(1): 40-48.

Matzuk, M. M. and D. J. Lamb (2008). "The biology of infertility: research advances and clinical challenges." Nat Med **14**(11): 1197-1213.

McVey, M. and S. E. Lee (2008). "MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings." Trends Genet **24**(11): 529-538.

Medicine, T. P. C. o. t. A. S. f. R. (2008). "Smoking and infertility." Fertil Steril **90**(5 Suppl): S254-259.

Meservy, J. L., R. G. Sargent, et al. (2003). "Long CTG tracts from the myotonic dystrophy gene induce deletions and rearrangements during recombination at the APRT locus in CHO cells." Mol Cell Biol **23**(9): 3152-3162.

Metzker, M. L. (2010). "Sequencing technologies - the next generation." Nat Rev Genet **11**(1): 31-46.

Miyamoto, T., S. Hasuike, et al. (2003). "Azoospermia in patients heterozygous for a mutation in SYCP3." Lancet **362**(9397): 1714-1719.

Moens, P. B., N. K. Kolas, et al. (2002). "The time course and chromosomal localization of recombination-related proteins at meiosis in the mouse are compatible with models that can resolve the early DNA-DNA interactions without reciprocal recombination." J Cell Sci **115**(Pt 8): 1611-1622.

Morel, F., N. Douet-Guilbert, et al. (2004). "Meiotic segregation of translocations during male gametogenesis." Int J Androl **27**(4): 200-212.

Morel, F., N. Douet-Guilbert, et al. (2004). "Meiotic segregation of a t(7;8)(q11.21;cen) translocation in two carrier brothers." Fertil Steril **81**(3): 682-685.

Moreno, R. D. and C. P. Alvarado (2006). "The mammalian acrosome as a secretory lysosome: new and old evidence." Mol Reprod Dev **73**(11): 1430-1434.

Moreno, R. D., J. Palomino, et al. (2006). "Assembly of spermatid acrosome depends on microtubule organization during mammalian spermiogenesis." Dev Biol **293**(1): 218-227.

Moreno, R. D., J. Ramalho-Santos, et al. (2000). "The Golgi apparatus segregates from the lysosomal/acrosomal vesicle during rhesus spermiogenesis: structural alterations." Dev Biol **219**(2): 334-349.

Moreno, R. D., J. Ramalho-Santos, et al. (2000). "Vesicular traffic and golgi apparatus dynamics during mammalian spermatogenesis: implications for acrosome architecture." Biol Reprod **63**(1): 89-98.

Musunuru, K., J. P. Pirruccello, et al. (2010). "Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia." N Engl J Med **363**(23): 2220-2227.

Myers, S., L. Bottolo, et al. (2005). "A fine-scale map of recombination rates and hotspots across the human genome." Science **310**(5746): 321-324.

Myers, S., C. Freeman, et al. (2008). "A common sequence motif associated with recombination hot spots and genome instability in humans." Nat Genet **40**(9): 1124-1129.

Ng, S. B., K. J. Buckingham, et al. (2010). "Exome sequencing identifies the cause of a mendelian disorder." Nat Genet **42**(1): 30-35.

Nielsen, R., J. S. Paul, et al. (2011). "Genotype and SNP calling from next-generation sequencing data." Nat Rev Genet **12**(6): 443-451.

O'Flynn O'Brien, K. L., A. C. Varghese, et al. (2010). "The genetic causes of male factor infertility: a review." Fertil Steril **93**(1): 1-12.

Oakberg, E. F. (1956). "Duration of spermatogenesis in the mouse and timing of stages of the cycle of the seminiferous epithelium." Am J Anat **99**(3): 507-516.

Oakberg, E. F. (1957). "Duration of spermatogenesis in the mouse." Nature **180**(4595): 1137-1138.

Oates, R. (2012). "Evaluation of the azoospermic male." Asian J Androl **14**(1): 82-87.

Ostertag, E. M. and H. H. Kazazian, Jr. (2001). "Biology of mammalian L1 retrotransposons." Annu Rev Genet **35**: 501-538.

Page, S. L. and R. S. Hawley (2003). "Chromosome choreography: the meiotic ballet." Science **301**(5634): 785-789.

Parvanov, E. D., P. M. Petkov, et al. (2010). "Prdm9 controls activation of mammalian recombination hotspots." Science **327**(5967): 835.

Patterson, T. R., J. D. Stringham, et al. (1990). "Nicotine and cotinine inhibit steroidogenesis in mouse

Leydig cells." Life Sci **46**(4): 265-272.

Pawlowski, W. P. and W. Z. Cande (2005). "Coordinating the events of the meiotic prophase." Trends Cell Biol **15**(12): 674-681.

Peoples-Holst, T. L. and S. M. Burgess (2005). "Multiple branches of the meiotic recombination pathway contribute independently to homolog pairing and stable juxtaposition during meiosis in budding yeast." Genes Dev **19**(7): 863-874.

Phadnis, N., R. W. Hyppa, et al. (2011). "New and old ways to control meiotic recombination." Trends Genet **27**(10): 411-421.

Pierre, V., G. Martinez, et al. (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.

Pirrello, O., N. Machev, et al. (2005). "Search for mutations involved in human globozoospermia." Hum Reprod **20**(5): 1314-1318.

Poongothai, J., T. S. Gopenath, et al. (2009). "Genetics of human male infertility." Singapore Med J **50**(4): 336-347.

Pradillo, M. and J. L. Santos (2011). "The template choice decision in meiosis: is the sister important?" Chromosoma **120**(5): 447-454.

Ptak, S. E., D. A. Hinds, et al. (2005). "Fine-scale recombination patterns differ between chimpanzees and humans." Nat Genet **37**(4): 429-434.

Qi, H., M. M. Moran, et al. (2007). "All four CatSper ion channel proteins are required for male fertility and sperm cell hyperactivated motility." Proc Natl Acad Sci U S A **104**(4): 1219-1223.

Ramadan, W. M., J. Kashir, et al. (2012). "Oocyte activation and phospholipase C zeta (PLCzeta): diagnostic and therapeutic implications for assisted reproductive technology." Cell Commun Signal **10**(1): 12.

Ramalho-Santos, J., R. D. Moreno, et al. (2001). "Membrane trafficking machinery components associated with the mammalian acrosome during spermiogenesis." Exp Cell Res **267**(1): 45-60.

Ramalho-Santos, J., G. Schatten, et al. (2002). "Control of membrane fusion during spermiogenesis and the acrosome reaction." Biol Reprod **67**(4): 1043-1051.

Reiter, L. T., P. J. Hastings, et al. (1998). "Human meiotic recombination products revealed by sequencing a hotspot for homologous strand exchange in multiple HNPP deletion patients." Am J Hum Genet **62**(5): 1023-1033.

Reiter, L. T., T. Murakami, et al. (1996). "A recombination hotspot responsible for two inherited peripheral neuropathies is located near a mariner transposon-like element." Nat Genet **12**(3): 288-297.

Ren, D., B. Navarro, et al. (2001). "A sperm ion channel required for sperm motility and male

fertility." Nature **413**(6856): 603-609.

Richardson, C., N. Horikoshi, et al. (2004). "The role of the DNA double-strand break response network in meiosis." DNA Repair (Amst) **3**(8-9): 1149-1164.

Robey, P. G. (2000). "Stem cells near the century mark." J Clin Invest **105**(11): 1489-1491.

Robichon, C., M. Varret, et al. (2006). "DnaJA4 is a SREBP-regulated chaperone involved in the cholesterol biosynthesis pathway." Biochim Biophys Acta **1761**(9): 1107-1113.

Roosen-Runge, E. C. (1952). "Kinetics of spermatogenesis in mammals." Ann N Y Acad Sci **55**(4): 574-584.

Rowold, D. J. and R. J. Herrera (2000). "Alu elements and the human genome." Genetica **108**(1): 57-72.

Sadeghi-Nejad, H. and F. Farrokhi (2006). "Genetics of azoospermia: current knowledge, clinical implications, and future directions. Part I." Urol J **3**(4): 193-203.

San Filippo, J., P. Sung, et al. (2008). "Mechanism of eukaryotic homologous recombination." Annu Rev Biochem **77**: 229-257.

Sasaki, M., J. Lange, et al. (2010). "Genome destabilization by homologous recombination in the germ line." Nat Rev Mol Cell Biol **11**(3): 182-195.

Sayers, E. W., T. Barrett, et al. (2011). "Database resources of the National Center for Biotechnology Information." Nucleic Acids Res **40**(Database issue): D13-25.

Segurel, L., E. M. Leffler, et al. (2011). "The case of the fickle fingers: how the PRDM9 zinc finger protein specifies meiotic recombination hotspots in humans." PLoS Biol **9**(12): e1001211.

Seminara, S. B., L. M. Oliveira, et al. (2000). "Genetics of hypogonadotropic hypogonadism." J Endocrinol Invest **23**(9): 560-565.

Sermondade, N., E. Hafhouf, et al. (2011). "Successful childbirth after intracytoplasmic morphologically selected sperm injection without assisted oocyte activation in a patient with globozoospermia." Hum Reprod **26**(11): 2944-2949.

Setchell, B. P., L. G. Sanchez-Partida, et al. (1993). "Epididymal constituents and related substances in the storage of spermatozoa: a review." Reprod Fertil Dev **5**(6): 601-612.

Sharpe, R. M. and N. E. Skakkebaek (1993). "Are oestrogens involved in falling sperm counts and disorders of the male reproductive tract?" Lancet **341**(8857): 1392-1395.

Shaw, C. J. and J. R. Lupski (2004). "Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease." Hum Mol Genet **13 Spec No 1**: R57-64.

Shaw, C. J. and J. R. Lupski (2005). "Non-recurrent 17p11.2 deletions are generated by homologous and non-homologous mechanisms." Hum Genet **116**(1-2): 1-7.

Simorangkir, D. R., G. R. Marshall, et al. (2005). "Prepubertal expansion of dark and pale type A spermatogonia in the rhesus monkey (Macaca mulatta) results from proliferation during

infantile and juvenile development in a relatively gonadotropin independent manner." Biol Reprod **73**(6): 1109-1115.

Singh, K. and D. Jaiswal (2011). "Human male infertility: a complex multifactorial phenotype." Reprod Sci **18**(5): 418-425.

Singleton, A. B. (2011). "Exome sequencing: a transformative technology." Lancet Neurol **10**(10): 942-946.

Skakkebaek, N. E., E. Rajpert-De Meyts, et al. (2001). "Testicular dysgenesis syndrome: an increasingly common developmental disorder with environmental aspects." Hum Reprod **16**(5): 972-978.

Skaletsky, H., T. Kuroda-Kawaguchi, et al. (2003). "The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes." Nature **423**(6942): 825-837.

Slack, A., P. C. Thornton, et al. (2006). "On the mechanism of gene amplification induced under stress in Escherichia coli." PLoS Genet **2**(4): e48.

Smagulova, F., I. V. Gregoretti, et al. (2011). "Genome-wide analysis reveals novel molecular features of mouse recombination hotspots." Nature **472**(7343): 375-378.

Smith, C. E., B. Llorente, et al. (2007). "Template switching during break-induced replication." Nature **447**(7140): 102-105.

Sofikitis, N., N. Giotitsas, et al. (2008). "Hormonal regulation of spermatogenesis and spermiogenesis." J Steroid Biochem Mol Biol **109**(3-5): 323-330.

Stankiewicz, P. and J. R. Lupski (2002). "Genome architecture, rearrangements and genomic disorders." Trends Genet **18**(2): 74-82.

Stankiewicz, P., C. J. Shaw, et al. (2003). "Genome architecture catalyzes nonrecurrent chromosomal rearrangements." Am J Hum Genet **72**(5): 1101-1116.

Steinberg, J. P., K. Takamiya, et al. (2006). "Targeted in vivo mutations of the AMPA receptor subunit GluR2 and its interacting protein PICK1 eliminate cerebellar long-term depression." Neuron **49**(6): 845-860.

Stouffs, K., W. Lissens, et al. (2005). "SYCP3 mutations are uncommon in patients with azoospermia." Fertil Steril **84**(4): 1019-1020.

Tang, C. J., C. Y. Lin, et al. (2006). "Dynamic localization and functional implications of Aurora-C kinase during male mouse meiosis." Dev Biol **290**(2): 398-410.

Thielemans, B. F., C. Spiessens, et al. (1998). "Genetic abnormalities and male infertility. A comprehensive review." Eur J Obstet Gynecol Reprod Biol **81**(2): 217-225.

Toshimori, K. and C. Ito (2003). "Formation and organization of the mammalian sperm head." Arch Histol Cytol **66**(5): 383-396.

Vine, M. F., B. H. Margolin, et al. (1994). "Cigarette smoking and sperm density: a meta-analysis."

Fertil Steril **61**(1): 35-43.

Virtanen, H. E. and J. Toppari (2008). "Epidemiology and pathogenesis of cryptorchidism." Hum Reprod Update **14**(1): 49-58.

Viville, S., R. Mollard, et al. (2000). "Do morphological anomalies reflect chromosomal aneuploidies?: case report." Hum Reprod **15**(12): 2563-2566.

Wang, D. G., J. B. Fan, et al. (1998). "Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome." Science **280**(5366): 1077-1082.

Webster, K. E., M. K. O'Bryan, et al. (2005). "Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis." Proc Natl Acad Sci U S A **102**(11): 4068-4073.

Westerveld, G. H., S. Repping, et al. (2005). "Mutations in the human BOULE gene are not a major cause of impaired spermatogenesis." Fertil Steril **83**(2): 513-515.

Westerveld, G. H., S. Repping, et al. (2005). "Mutations in the chromosome pairing gene FKBP6 are not a common cause of non-obstructive azoospermia." Mol Hum Reprod **11**(9): 673-675.

Woodward, K. J., M. Cundall, et al. (2005). "Heterogeneous duplications in patients with Pelizaeus-Merzbacher disease suggest a mechanism of coupled homologous and nonhomologous recombination." Am J Hum Genet **77**(6): 966-987.

Xiao, N., C. Kam, et al. (2009). "PICK1 deficiency causes male infertility in mice by disrupting acrosome formation." J Clin Invest **119**(4): 802-812.

Xu, M., J. Xiao, et al. (2003). "Identification and characterization of a novel human testis-specific Golgi protein, NYD-SP12." Mol Hum Reprod **9**(1): 9-17.

Yan, W. (2009). "Male infertility caused by spermiogenic defects: lessons from gene knockouts." Mol Cell Endocrinol **306**(1-2): 24-32.

Yanagida, K., H. Katayose, et al. (1999). "Successful fertilization and pregnancy following ICSI and electrical oocyte activation." Human reproduction **14**(5): 1307-1311.

Yanagida, K., K. Morozumi, et al. (2006). "Successful pregnancy after ICSI with strontium oocyte activation in low rates of fertilization." Reproductive biomedicine online **13**(6): 801-806.

Yao, R., C. Ito, et al. (2002). "Lack of acrosome formation in mice lacking a Golgi protein, GOPC." Proc Natl Acad Sci U S A **99**(17): 11211-11216.

Yu, J., Z. Chen, et al. (2012). "CFTR mutations in men with congenital bilateral absence of the vas deferens (CBAVD): a systemic review and meta-analysis." Hum Reprod **27**(1): 25-35.

Zenzes, M. T., L. A. Puy, et al. (1999). "Detection of benzo[a]pyrene diol epoxide-DNA adducts in embryos from smoking couples: evidence for transmission by spermatozoa." Mol Hum Reprod **5**(2): 125-131.

Zickler, D. and N. Kleckner (1998). "The leptotene-zygotene transition of meiosis." Annu Rev Genet **32**: 619-697.

Zickler, D. and N. Kleckner (1999). "Meiotic chromosomes: integrating structure and function." Annu Rev Genet **33**: 603-754.

Elias ELINATI

Génétique de l'infertilité masculine

Résumé

Le génotypage d'une famille jordanienne consanguine constituée de 5 frères globozoospermiques et de 3 frères fertiles sur puce Affymetrix, a permis d'identifierun nouveau gène responsable de la globozoospermie situé dans un intervalle de 6.4Mb en 12q14.2. Au regard de son expression prédominante dans le testicule et l'implication de son orthologue, chez C. elegans, dans la polarisation cellulaire, le gène *DPY19L2* est un gène candidat parfait. Le gène, codant pour une protéine transmembranaire, est flanqué par deux séquences répétées (LCRs) qui partagent 96,5% d'identité. Dans une première étude, une délétion de 200Kb englobant l'ensemble du gène a été mise en évidence chez les 4 frères infertiles de cette famille jordanienne ainsi que chez 3 autres patients non apparentés. Nous avons ensuite recruté une plus grande cohorte de 54 patients. Parmi ces patients, 20 sont homozygotes pour la délétion de *DPY19L2* et 7 sont hétérozygotes composites associant la délétion hétérozygote et une mutation ponctuelle. En outre, nous avons identifié, 4 patients avec des mutations ponctuelles homozygotes. Par conséquent, la fréquence d'implication de *DPY19L2* s'élève à 66.7%. En tout, 9 points de cassures, regroupés en deux hotspots au sein des LCRs, ont pu être mis en évidence. Ceci confirme que le mécanisme sous-jacent de la délétion est une recombinaison homologue non allélique (NAHR) entre les LCRs. En conclusion, nous confirmons que *DPY19L2* est le principal gène de la globozoospermie et nous élargissons le spectre des mutations possible dans ce gène.

Mots Clés : globozoospermie, *DPY19L2*, NHAR, LCR, hotspots.

Résumé en anglais

Performing a genome wide scan by SNP microarray on a Jordanian consanguineous family where five brothers were diagnosed with complete globozoospermia, we show in a first study that the four out of five analysed infertile brothers carried a homozygous deletion of 200 kb on chromosome 12 encompassing only *DPY19L2*. The gene encodes for a transmembrane protein and is surrounded by two low copy repeats (LCRs). Very similar deletions were found in three additional unrelated patients. Later, we have pursued our patient screen by recruiting a largest cohort of patients. Out of a total of 54 patients analysed, 36 (66.7%) showed a mutation in *DPY19L2*. Out of 36 mutated patients, 20 are homozygous deleted, 7 heterozygous composite and 4 showed a homozygous point mutation. We characterized a total of nine breakpoints that clustered in two recombination hotspots, both containing direct repeat elements. These findings confirm that the deletion is due to a nonallelic homologous recombination (NAHR) between the two LCRs. Thus, Globozoospermia can be considered as a new genomic disorder. This study confirms that *DPY19L2* is the major gene responsible for globozoospermia and enlarges the spectrum of possible mutations in the gene.

Keywords: globozoospermia, DPY19L2, NHAR, LCR, hotspots.