DOCTORAL THESIS

GENETICS OF MALE INFERTILITY: MOLECULAR STUDY OF NON-SYNDROMIC CRYPTORCHIDISM AND SPERMATOGENIC IMPAIRMENT

Deborah Grazia Lo Giacco

November 2013

Genetics of male infertility: molecular study of non-syndromic cryptorchidism and spermatogenic impairment

Thesis presented by
Deborah Grazia Lo Giacco

To fulfil the PhD degree at
Universitat Autònoma de Barcelona

Thesis realized under the direction of Dr. Elisabet Ars Criach and Prof. Csilla Krausz at the laboratory of Molecular Biology of Fundació Puigvert, Barcelona

Thesis ascribed to the Department of Cellular Biology,
Physiology and Immunology, Medicine School of Universitat Autònoma de Barcelona
PhD in Cellular Biology

Dr. Elisabet Ars Criach	Prof. Csilla Krausz	Dr. Carme Nogués Sanmiquel
Director of the thesis	*Director of the thesis*	*Tutor of the thesis*

Deborah Grazia Lo Giacco
Ph.D Candidate

A mis padres

Agradecimientos

Esta tesis es un esfuerzo en el cual, directa o indirectamente, han participado varias personas, leyendo, opinando, corrigiendo, teniéndo paciencia, dando ánimo, acompañando en los momentos de crisis y en los momentos de felicidad.

Antes de todo quisiera agradecer a mis directoras de tesis, la **Dra Csilla Krausz** y la **Dra Elisabet Ars,** por su dedicación costante y continua a este trabajo de investigación y por sus observaciones y siempre acertados consejos. Gracias por haber sido mentores y amigas, gracias por transmitirme vuestro entusiasmo y por todo lo que he aprendido de vosotras.

Mi más profundo agradecimiento a la **Sra Esperança Marti** por haber creído en el valor de nuestro trabajo y haber hecho que fuera posible.

Quisiera agradecer al **Dr. Eduard Ruiz-Castañé** por su apoyo y ayuda constante, y a todos los médicos del Servicio de Andrología: Dr. **Lluis Bassas**, **Josvany Sanchez Curbelo**, **Joaquim Sarquella**, **Osvaldo Rajmil**, **Alvaro Vives, Saulo Camarena** y **Gerardo Ortiz** por tener su puerta siempre abierta para cualquier consulta y por el esfuerzo y el tiempo que me han dedicado. Sin ellos este trabajo no hubiese sido posible.

Deseo dar las gracias a todas las "moleculitas" del laboratorio de biología molecular, por haberme alegrado cada día y por su apoyo moral y material: a **Patricia** por encontrar siempre una solución para cada problema y por estar siempre dispuesta a ayudarme desde el día que llegué al laboratorio que a penas entendía lo que me explicabas. A **Ania,** que es una persona deliciosa, como sus pasteles. Mil gracias Ania por interesarte en mi trabajo por apoyarme y animarme. A **Paola** por sus consejos y por ser siempre tan positiva y enseñarme que la calma y el optimismo pueden con todo. A **Elena,** mi vecina de poyata, unas de las personas más creativas que he encontrado nunca, gracias por escucharme y aconsejarme, por aguantar mis momentos de crisis y por compartir conmigo los tuyos. A **Olga** por su alegria y su risa contagiosa, gracias por animarme siempre y por estar cada día a mi lado, en todos los sentidos…A **Irene,** gracias por todo lo que has hecho por mi en estos años, por estar siempre dispuesta a ayudarme y por haberlo hecho cuando más lo he necesitado…y por supuesto por acordarme cada día cuando es el momento de hacer una pausa. A **Lluisa** por ser siempre tan alegre y por encontrar el lado divertido de cada situatión. A **Gemma,** con la que he compartido las horas más solitarias del laboratorio y las alegrías y los problemas del doctorado. Que et vagi molt bé, i ànim que el pom de dalt del castell está molt aprop! A **Bea** y **Fede** con los que he compartido preocupaciones y dudas de la tesis pero también muy buenos momentos musicales….habría que "volver" a tocar juntos.. A **Sheila, Ana**, **Arantxa, Laia** y **Xenia** que me han acompañado en el primer tramo de este camino, el mas durillo.

Gracias al **Dr Artur Oliver** y a las **Dra Charo Montañes** y **Silvia Gracia** por su constante muestra de cariño.

Un agradecimiento especial a las Dras. **Ana Mata**, **Olga Martinez**, **Olga Lopez**, **Aurora Garcia** y a **Rosalia Gusta** del laboratorio de seminología, por su infinita paciencia y por acojerme siempre con una sonrisa….a mi y a mi aparatitos "marea espermatozoides".

Gracias a **los chicos de archivo** por ser siempre tan amables conmigo y a **Ricard** de la biblioteca por su ayuda y por su simpatía.

Gracias a la **Dra Silvia Mateu** por el cariño, la comprensión y su disponibilidad a auydarme siempre; a **Bea Rodrigo** con su italiano perfecto y a **las chicas de secretaría** por atenderme siempre con una sonrisa.

Mi más profundo agradecimiento **a todo el personal** y a **los pacientes de la Fundació Puivgvert** por haber hecho que este trabajo fuera posible y a la financiación del Ministerio de Salud (FIS) por haber garantizado su continuidad.

Agradezco a la **Dra Carme Nogués** por su trabajo como tutora y su ayuda preciosa con todo el papeleo y más....

Les doy las gracias a mi "familia española" y en particular a **Miquel** y **Cristina** por haberme hecho sentir como en mi casa desde el primer día, por haberme hecho entrar en sus vidas y compartir conmigo momentos felices y momentos tristes, por alegrase conmigo de mis éxitos y animarme a seguir en los momentos mas dificiles. Gracias a ti **Jordi** por tu infinita paciencia, por entender mis ausencias y mis malos momentos, por esperarme horas cuando te había dicho que saldría en minutos, por animarme, por creer en lo que hago y por hacerme creer que lo que hago es importante.

Chiara, io veramente pensavo che avrei trovato le parole giuste per ringraziarti per tutto quello che hai fatto per me, ancor prima di conoscerci di persona. E invece tutto quello che riesco a dirti e' un semplice "Grazie di cuore!!": per aiutarmi con la tesi, per ascoltarmi e consigliarmi, per essermi stata vicina in questi ultimi passi verso la meta.

Grazie a **Fabrice** e **Claudia** collaboratori e cari amici, grazie per esserci sempre!!

Un ringraziamento speciale **ai miei genitori e a mio fratello,** che porto sempre con me per quanto lontana io sia da casa e a cui devo tutto ciò che sono oggi e tutto ciò che di bello farò nella vita.

Mi más profundo agradecimiento a todos!

TABLE OF CONTENTS

1. SUMMARY

The present thesis explores the role of specific genetic variants in the etiology of two forms of disturbed male reproductive fitness: non-syndromic cryptorchidism and idiopathic spermatogenic failure.

The etiology of non-syndromic cryptorchidism, still largely unknown, is likely to be multifactorial reflecting the involvement of environmental and genetic factors. *RXFP2* and *ESR1* are interesting candidate genes due to the involvement of RXFP2 receptor in testis descent and the potential role of ESR1 receptor as mediator of substances able to interfere with the development of the male urogenital tract. The first part of the thesis presents the study of two genetic variants of *RXFP2* and *ESR1*, aimed to explore their contribution to non syndromic cryptorchidism. These are the missense substitution T222P in exon 8 of *RXFP2*, previously proposed as a pathogenic mutation for cryptorchidism, and the VNTR polymorphism $(TA)_n$ within the *ESR1* promoter, previously reported as a potential functional polymorphism in relation to bone mineralization and spermatogenesis and never studied so far in relation to cryptorchidism. Complessively 550 subjects from Italy and 570 from Spain (with and without history of cryptorchidism) were screened for each variant. The T222P was found at a similar frequency in both patients (1.6%) and controls (1.8%) in the Spanish population thus indicating that in this population it is a common polymorphism with no pathogenic effect. In the Italian study population the frequency of T222P was significantly higher in patients ($p = 0.031$) supporting a role for this variant as mild risk factor for cryptorchidism (OR= 3.17, 95% CI: 1.07–9.34). Nevertheless, the screening for this variant for diagnostic purposes is not advised because of the relatively high frequency of control carriers (1.4%). Two $(TA)_n$ genotypes (A and B) were defined based on the possible allelic combinations of high, medium or low number of repeats and their frequency was compared between cases and controls from the two study populations. Allelic distribution of $(TA)_n$ did not show significant differences between cases and controls. The frequency of genotype A, considered the functionally most active one, was also similar in cases and controls both in the Italian and Spanish study populations. These results indicate that the $(TA)_n$ within the *ESR1* promoter is not associated with non syndromic cryptorchidism neither in the Spanish nor in the Italian population.

The second part of the thesis explores the role of Y and X-linked Copy Number Variations (CNVs) in the etiology of spermatogenic impairment. The study of Y-linked CNVs includes: i) the retrospective analysis of 806 mainly Spanish infertile patients screened for Y microdeletions and ii) the study of partial AZFc deletions and duplications (the latter studied here for the first time in the Spanish population) including 330 idiopathic infertile patients and 385 controls from Spain. A total of 27/806 (3.3%) patients carried complete AZF deletions. All were azoo/cryptozoospermic, except for one whose sperm concentration was 1-$2x10^6$/ml. This finding integrated with the literature suggests that routine AZF microdeletion testing could eventually include only men with $\leq 2x10^6$/ml. In AZFc deleted men a lower sperm recovery rate was observed upon conventional TESE (9.1%) compared to the literature (60%-80% with microTESE) indicating that microTESE ensuring better outcomes, should be regarded as the best option. Haplogroup (hgr) E was the most represented among non-Spanish whereas hgr P among Spanish *AZF* deletion carriers supporting a potential contribution of Y background to the inter-population variability in deletion frequency. The *gr/gr* deletion was significantly associated with spermatogenic impairment, further supporting the inclusion of this genetic test in the work-up of infertile men, while partial AZFc duplications do not represent a risk for spermatogenic failure in the Spanish population. The present thesis addresses the topic of X-linked CNVs in spermatogenic defects by presenting the X-chromosome array-CGH (a-CGH) analysis of 199 men with different sperm count. A total of 73 CNVs were identified including 29 mostly rare losses and 44 gains. A significantly higher burden of deletions was found in patients compared to controls due to an excessive rate of

deletions/person (0.57 *versus* 0.21, respectively; p =8.78x10^{-6}) and to a higher mean sequence loss/person (11.79 Kb and 8.13 Kb, respectively; p = 3.43 x10^{-6}). This finding suggests that an increased X chromosome deletion burden may be involved in the etiology of spermatogenic impairment. X-linked Cancer Testis Antigen (CTA) genes resulted to be frequently affected, indicating that their dosage variation may play a role in X-linked CNV-related spermatogenic failure.
Three recurrent deletions, mapping to the Xq27.3 (CNV64) and Xq28 (CNV67 and CNV69), were considered of interest for their exclusive (CNV67) or prevalent (CNV64 and CNV69) presence in patients. These deletions were object of an in-depth analysis including: i) the screening of 627 idiopathic patients and 628 normozoospermic controls from two Mediterranean populations (Spanish and Italian); ii) the molecular characterization of deletions; iii) the exploration for functional elements. CNV64 and CNV69 were significantly more frequent in patients than controls (OR=1.9 and 2.2, for CNV64 and CNV69 respectively). CNV69 displayed at least two deletion patterns (type A and type B), of which type B, being significantly more represented in patients than controls, may account for the potential deleterious effect of CNV69 on sperm production. No genes have been identified inside CNV64 and CNV69, nevertheless a number of regulatory elements, have been found to be potentially affected. CNV67 deletion was exclusively found in patients at a frequency of 1,1% ($p<0.01$) thus resembling the *AZF* region of the Y chromosome. This deletion may involve the CTA gene *MAGEA9* and/or of its regulatory elements. It may also affect regulatory elements of *HSFX1/2* showing testis-specific expression. Pedigree analyses of two CNV67 carriers indicated that CNV67 deletion is maternally inherited. One of these families was especially informative since the pathological semen phenotype of the carrier versus his normozoospermic non-carrier brother was a strong indicator for a pathogenic effect of the deletion on spermatogenesis.

The results of the present thesis are in line with the prevalent view that the etiology of non-syndromic cryptorchidism is likely to be polygenic and in most of cases multifactorial. We are confident that high throughput technologies, especially next generation sequencing, will help to unravel the complexity of the etiology of this disease. We provide evidence for a significant association between recurrent X-linked deletions and impaired sperm production. Further studies in other ethnic-geographic groups are needed to confirm the association of CNV64 and CNV69 type B with spermatogenic failure. Due to the specific association of CNV67 with spermatogenic impairment a parallelism with *AZF* deletions of the Y chromosome is tempting. This CNV merits further investigation in order to provide a feasible substrate for fine molecular characterization, and to further evaluate its putative diagnostic value.

RESUMEN

La presente tesis es una aportación al conocimiento de las bases genéticas de la criptorquidia no sindrómica y de las formas idiopáticas de espermatogénesis anómala. *RXFP2* y *ESR1* son dos genes candidatos en la etiología de la criptorquidia no sindrómica debido al rol del receptor RXFP2 en el control hormonal del descenso testicular y al posible papel del receptor ESR1 como mediador de los efectos de substancias capaces de interferir con el desarrollo del tracto urogenital masculino. La primera parte de la tesis está enfocada en el estudio del papel de dos variantes de estos genes en la etiología de la criptorquidia: la variante de secuencia p.T222P del gen *RXFP2* (previamente descrita como una mutación patogénica causante de criptorquidia) y el microsatélite (TA)n del promotor del gen *ESR1* estudiado por primera vez en relación con la criptorquidia. Para cada una de las variantes hemos analizado 550 sujetos de origen italiana y 570 de origen española (entre pacientes y controles). En la población española la variante p.T222P se ha encontrado en una proporción parecida de casos (1.6%) y controles (1.8%). Estos resultados indican que dicha variante genética es un polimorfismo común sin significado clínico en esta población. En cambio en la población italiana la frecuencia de la p.T222P ha resultado ser significativamente más alta en pacientes que en controles, lo que sugiere un posible papel para dicha variante como factor de riesgo para la criptorquidia. No hemos observado diferencias significativas en la distribución de los alelos (TA)n entre casos y controles. La frecuencia del genotipo A (supuestamente asociado a mayor actividad del promotor de *ESR1*) ha resultado ser parecida en casos y controles en ambas poblaciones de estudio. Estos resultados indican que el polimorfismo (TA)n no está asociado con la criptorquidia no sindrómica ni en población española ni en población italiana.

La segunda parte de la tesis investiga el papel de variantes de número de copias (CNVs) en los cromosomas Y y X en la etiología de la espermatogenésis anómala. El estudio de las CNVs del cromosoma Y incluye: i) el análisis de más de 800 pacientes infértiles (mayoritariamente españoles) para las microdeleciones del cromosoma Y; ii) el estudio de deleciones y duplicaciones parciales de la región *AZFc* (estas últimas estudiadas por primera vez en población española) en 330 infértiles idiopáticos y 385 controles normozoospermicos de origen española. Nuestros datos basados en 27 portadores de microdeleciones clasicas del Y (3.3%) junto con una amplia revisión de la literatura indican que el diagnóstico genético de rutina para las microdeleciones del Y se podría limitar a varones con concentración espermática $\leq 2x10^6$ /ml; ii) la técnica de recuperación espermática testicular más adecuada para los pacientes azoospérmicos portadores de deleciones *AZFc* es la microTESE; iii) el background del Y podría explicar en parte que las microedeleciones del Y si encuentren en diferente frecuencia en distintas poblaciones. Con respecto al estudio de los reordenamientos parciales en *AZFc*, hemos corroborado que las deleciones *gr/gr* (deleciones parciales *AZFc*) representan un factor de riesgo para la espermatogénesis alterada en población caucásica y hemos descartado que las duplicaciones *AZFc* contribuyan a esta anomalía. Para estudiar el papel de las CNVs del cromosoma X en la etiología de la espermatogenésis anómala, hemos analizado 199 sujetos con diferente concentración espermática utilizando un array de CGH de alta resolución específico para el cromosoma X. Hemos identificado 73 CNVs entre las cuales 29 perdidas (mayoritariamente raras) y 44 ganancias. El número de deleciones por sujeto y la cantidad media de DNA delecionado han resultado ser significativamente mayores en pacientes que en controles, lo que sugiere que un aumento de la "carga de deleción" podría estar implicado en la etiología de la espermatogenésis anómala. Tres deleciones recurrentes localizadas en Xq27.3 (CNV64) y en Xq28 (CNV67 y CNV69) han sido consideradas de mayor interés por su presencia exclusiva (CNV67) o prevalente (CNV64, CNV69) en el grupo de los pacientes. Dichas deleciones han sido objeto de un estudio profundizado que incluye: i) el screening de 627 infértiles idiopáticos y 628 controles normozoospérmicos de dos poblaciones mediterráneas (española e italiana); ii) la caracterización molecular de las deleciones; iii) la identificación de elementos funcionales afectados por las deleciones. Las CNV64 y CNV69 se han encontrado más frecuentemente en pacientes que en controles. La CNV69 presenta al menos dos patrones de deleción (tipo A y tipo B) de los cuales solo el tipo B ha resultado ser significativamente más representado en pacientes, indicando que este tipo de deleción podría ser responsable del efecto deletéreo de la CNV69 en la espermatogenésis. Las CNV64 y CNV69 no contienen genes en su interior pero podrían afectar una serie de elementos reguladores localizados <0.5 Mb. La CNV67, identificada únicamente en un 1.1% de los pacientes ($p<0.01$), podría directamente afectar el gen *MAGEA9* y/o sus elementos reguladores. El análisis del pedigrí de dos portadores de CNV67 ha evidenciado que esta deleción es trasmitida por la madre. La condición de normozoospermico del hermano no portador de uno de los pacientes portador de CNV67 suporta el posible efecto patogénico de esta deleción en la espermatogenésis.

Los resultados presentados en esta tesis suportan la visión que la etiología de la criptorquidia es probablemente poligénica y en la mayoría de los casos multifactorial. Es plausible que la aplicación de nuevas tecnologías de secuenciación masiva, garantizando un análisis más a gran escala del background genético del individuo, contribuyan a desenredar la complejidad de la etiología de esta enfermedad. Futuros estudios serán necesarios para corroborar el significado clínico de la CNV67 y confirmar, en otras poblaciones, la significativa asociación entre CNV64, CNV69 tipo B y espermatogenésis anómala.

2. ABBREVIATIONS

A	Adenine
a-CGH	array-Comparative Genomic Hybridization
AMELY	Amelogenin, Y-linked
AMH	Anti-Mullerian hormone
AKAP	A kinase (PRKA) anchor protein
AR	Androgen receptor
AZF	AZoospermia factor
b	blue amplicon family of the AZFc region of the Y chromosome
BOULE	bol, boule-like (Drosophila)
BPY2	Basic charge, Y-linked, 2
C	Cytosine
cAMP	Cyclic Adenosine monophosphate
CBAVD	Congenital Bilateral Absence of the Vas Deferens
CDY	Chromodomain protein, Y-linked
CDYL	Chromodomain protein, Y-like
CFTR	Cystic fibrosis transmembrane conductance regulator
CGRP	Calcitonin gene-related peptide
CI	Confidence Interval
CNV	Copy Number Variation
CoA	Coenzyme A
CSL	Cranial suspensory ligament
CSPG4LYP1	Chondroitin sulfate proteoglycan 4 pseudogene 1, Y-linked
CSRP	Cysteine and glycine-rich protein
CTA	Cancer Testis Antigen
CYorf15A/B	Chromosome Y open reading frame 15A/B
CYP8A1	Cytochrome P450, family 8, subfamily A, polypeptide 1
DAZ	Deleted in AZoospermia
DAZL	Deleted in AZoospermia-like
DES	Diethylstilbestrol
DDX3Y	DEAD (Asp-Glu-Ala-Asp) box helicase 3, Y-linked
DNA	Desoxyribonucleic Acid
DSBs	Double Strand Breaks
EAA	European Academy of Andrology
EED	Environmental Endocrine Disruptors
EIF1AY	Eukaryotic translation initiation factor 1A, Y-linked
EMQN	European Molecular GeneticsQuality Network
ENCODE	Encyclopedia of DNA Elements
ESR1	Estrogen receptor 1
FATE	Fetal and adult testis expressed
FGFR1	Fibroblast growth factor receptor 1
FISH	Fluorescent in situ hybridization
FoSTeS	Fork Stalling and Template Switching
FSH	Follicle-stimulating hormone

FSHR	Follicle-stimulating hormone receptor
G	Guanine
g	green amplicon family of the AZFc region of the Y chromosome
GAGE	G antigen
GNF	Genitofemoral nerve
GOLGA2LY1	Golgi autoantigen, golgin subfamily a, 2-like, Y-linked 1
GWAS	Genome Wide Association Study
hCG	Human chorionic gonadotropin
HDAC	Histone deacetylase
HERV	Human endogenous retrovirus
hgr	Haplogroup
HIV	Human Immunodeficiency Virus
HSD17B4	Hydroxysteroid (17-beta) dehydrogenase 4
HSFY1/2	Heat shock transcription factor, Y-linked 1/2
HSFX1/2	Heat shock transcription factor family, X linked 1/2
H3K4	Lysine 4 of histone 3
ICSI	Intracytoplasmic sperm injection
idicY	Isodicentric Y chromosome
INSL3	Insulin-like factor 3
isoY	Isochromosome Y
KAL1/2	Kallmann syndrome 1/2 sequence
KDM5D	Lysine (K)-specific demethylase 5D
LCR	Low copy repeats
Ley-I-L	Leydig insulin-like petide
LGR8	Leucine-rich repeat-containing G protein-coupled receptor 8
LH	Luteinizing hormone
LHR	Luteinizing hormone receptor
LINE	Long Interspersed Nuclear Elements
LTR	Long Terminal Repeats
MAGEA9	Melanoma antigen family A, 9
MAOA	Monoamine oxidase A
MAP1S	Microtubule-associated protein 1S
MECP2	Methyl CpG binding protein 2
MSCI	Meiotic sex chromosome inactivation
MSH5	MutS homolog 5
MSY	Male specific region Y
MTHFR	5,10-Metylene-tetra-hidrofolate reductase
NAHR	Non-allelic homologous recombination
NHEJ	Non-homologous end joining
NLGN4Y	Neuroligin 4, Y-linked
NOA	Non-obstructive azoospermia
OR	Odds Ratio
PAIS	Partial androgen insensitivity syndrome
PAR1/2	Pseudo autosomal region 1/2
PCDH11Y	Protocadherin 11 Y-linked
PGD	Preimplantation genetic diagnosis

PRKX	Protein kinase, X-linked
PRKY	Protein kinase, Y-linked, pseudogene
PRY	Testis-Specific PTP-BL-Related Protein on Y
r	red amplicon family of the AZFc region of the Y chromosome
RBM	RNA binding motif
RBMXL	RNA binding motif protein, X-linked-like
RBMY1A1	RNA binding motif protein, Y-linked, family 1, member A1
RFLP	Restriction fragment length polymoprhism
RLF	Relaxin like factor
RPS4Y1/2	Ribosomal protein S4, Y-linked 1/2
RRM	RNA recognition motif
RXFP2	Relaxin/insulin-like family peptide receptor 2
SCOS	Sertoli cell only syndrome
SD	Segmental duplication
SINE	Short interspersed nuclear elements
SMCY	SMC (mouse) homologue, Y
SNP	Single nucleotide polymorphism
SOX3	SRY (sex determining region Y)-box 3
SRY	Sex determining region Y
STAR	Steroidogenic acute regulatory protein
STS	Sequence tagged site
STS-PCR	Sequence tagged site-Polymerase chain reaction
SV40	Simion virus 40
T	Thymine
TAF7L	TAF7-like RNA polymerase II, TATA box binding protein (TBP)-associated factor
TBL1Y	Transducin (beta)-like 1, Y-linked
TDS	Testicular dysgenesis syndrome
TESE	Testicular sperm extraction
TGIF2LY	TGFB-induced factor homeobox 2-like, Y-linked
TMSB4Y	Thymosin beta 4, Y-linked
TSHR	Thyroid stimulating hormone receptor
TSPY	Testis specific protein, Y-linked
TTTY	Testis-specific transcript, Y-linked
USP26	Ubiquitin specific peptidase 26
USP9Y	Ubiquitin specific peptidase 9, Y-linked
UTR	Untranslated region
UTY	Ubiquitously transcribed tetratricopeptide repeat containing, Y-linked
VCY	Variable charge, Y-linked
VNTR	Variable number of tandem repeats
WHO	World health organization
XAR	X-added region
XCR	X-conserved region
XKRY	XK, Kell blood group complex subunit-related, Y-linked
YAP	Y-chromosome Alu Polymorphism
YCC	Y chromosome consortium

yel	yellow amplicon family of the AZFc region of the Y chromosome
ZFY	Zinc finger protein, Y-linked

3. INTRODUCTION

3.1 Infertility

3.1.1 Definition and prevalence

Infertility is a couple-affecting problem defined as failure to achieve a clinical pregnancy after 12-24 months of regular unprotected sexual intercourse (Zegers-Hochschild et al. 2009). It is estimated that one out of seven European couples (15%) suffer from reproductive health disorders in the form of infertility. The difficulty/impossibility of conceiving children is ascribable to a pure female factor in 35% of cases, whereas a male factor is (co-)responsible for roughly half of the cases of involuntary childless (pure or contributory male factor). This implies that approximately 7% of all men in Western countries are confronted with infertility problems (Forti and Krausz 1998). Finally, in 15% of cases, neither of the two partners show reproductive abnormalities: i) normal ovulation and absence of genital tract occlusions in the female partner; ii) normal sperm parameters in the male partner. These cases of couple infertility are defined as "unexplained".

3.1.2 Etiology of male factor infertility

In the majority of the cases, male factor infertility consists of a decreased semen quality in terms of number, progressive motility and morphology of spermatozoa or in terms of abnormal physico-chemical characteristics of the semen (volume, pH, liquefaction time and viscosity). The World Health Organization (WHO) established the reference ranges for normal values of the three sperm parameters which are included in the recent up-dated version of the guidelines for examination and processing of human semen (WHO 2010). Considering the WHO reference values three types of anomalies can be distinguished affecting the three sperm parameters:

- **Abnormalities of sperm number,** represented by **azoospermia** (no spermatozoa detectable in the ejaculate even after centrifugation of the semen sample), **cryptozoospermia** (spermatozoa detectable only in cytocentrifuged at concentration below 1 million spermatozoa/ml) and **oligozoospermia** (sperm concentration below 15 million spermatozoa/ml) which can be in turn distinguished into **moderate** (sperm concentration within 5-10 million spermatozoa/ml) and **severe** (sperm concentration below 5 million spermatozoa/ml) **oligozoospermia.**
- **Asthenozoospermia,** with less than 32% of progressively motile spermatozoa in the ejaculate.
- **Terathozoospermia**, with less than 4% of morphologically normal forms in the ejaculate.

The majority of infertile patients display anomalies in all three sperm parameters simultaneously, a condition conventionally defined as **oligo-astheno-theratozoospermia.**

Overall, the etiology of male infertility can be related to a wide range of congenital and acquired factors acting at pre-testicular, testicular and post-testicular levels (Krausz 2011).

Pre-testicular causes accounting for 10% of forms are mainly represented by two types of pathological conditions: hypogonadotrophic hypogonadism and coital disorders (erectile dysfunction and ejaculatory disorders such as eiaculatio precox and retrograde ejaculation).

Primary testicular dysfunction is the most common cause of spermatogenic impairment (75% of cases) and is related to a number of acquired and congenital etiological factors (**testicular causes**). A large number of pathologies may lead to an acquired primary testicular failure. Among them are orchitis, testis trauma, torsions, iatrogenic forms (gonadotoxic medications, chemo/radiotherapy, previous inguinal surgery) and some systemic diseases. Anorchia, cryptorchidism (especially bilateral forms) and genetic abnormalities such as karyotype anomalies and Y chromosome microdeletions are well defined congenital testicular factors of male infertility.

Post-testicular causes, representing 15% of the cases of male infertility, include obstruction/subostruction of the seminal tract both congenital, such as congenital absence of the vas deferens (CBAVD), and acquired, such as secondary to infections and inflammatory diseases of accessory glands, or to autoimmune causes.

Despite major advances in the diagnostic workup of infertile males, the etiopathogenesis of testicular failure remains undefined in about 50% of cases and are referred to as "**idiopathic infertility"**(Krausz 2011)**.** These idiopathic cases are likely to be of genetic origin since the number of genes involved in human spermatogenesis is probably over thousand and only a small proportion of them has so far been identified and even fewer have been analyzed.
Thus, it is highly likely that mutations or polymorphisms in candidate genes involved in spermatogenesis are responsible for the majority of idiopathic forms.

3.2 Cryptorchidism

The terms "cryptorchidism", "maldescensus testis" and "undescended testis" are synonymously used to define the impairment of the physiological descent of the testis from an intra-abdominal position into the scrotal sac. Cryptorchidism is one of the main forms of primary testicular disfunction. Failure of the testes to normally descend can occur bilaterally in one third of cases and unilaterally in two thirds of cases. Cryptorchid testes are classified on the basis of their position along the normal route of descent (high/low abdominal, inguinal, suprascrotal, high scrotal) or as ectopic. In the clinical setting, however, a simple distinction between palpable and non-palpable and between unilateral and bilateral is most often used (Virtanen HE, Bjerknes R, Cortes D, Jørgensen N et al. 2007) (Fig.1). Retractile testis represents another condition of impaired testicular descent in which testis that could be easily brought into a stable position at the bottom of the scrotum, tends to ascend to the upper part of the scrotum or into the inguinal canal as a result of an active cremasteric reflex. It is important to distinguish between retractile and real cryptorchid testis since the former is often a normal gonad both from a morphological and functional point of view. In some cases, cryptorchid patients are found to have an absence of one or both testes, a condition better defined as anorchia or "vanishing testis syndrome" that can be observed in 3-5% of all patients undergoing laparoscopic exploration for cryptorchidism (Foresta et al. 2008).

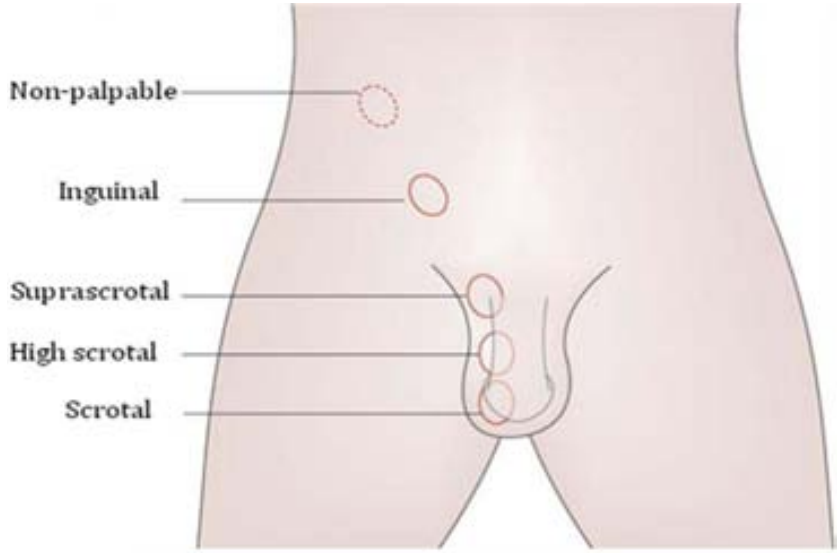


Figure 1. **Schematic illustration of the stages of cryptorchidism, according to the system proposed by Scorer. Figure of** (Bay et al. 2011)

Cryptorchidism is one of the most common congenital birth defects in male children, showing a 2-9% incidence in full-term male births with some geographical variation across Europe, which can be only partially due to genuine inter-population differences. In fact, the lack of a uniform definition of cryptorchidism might influence epidemiological studies reporting its prevalence. In about 50% of congenital cases of cryptorchidism, a spontaneous delayed descent can be observed by 3 months of age so that the prevalence of true cryptorchidism at 1 year of age decreases to 1-2%.

Cryptorchidism is not exclusively a congenital disorder since acquired forms diagnosed at infancy and childhood can also be found. The so-called secondary or acquired cryptorchidism consists of the ascent of testis into a cryptorchid position after normal descent into the scrotal sac at birth. Acquired cryptorchidism shows a cumulative incidence at 24 months of age, even higher than that observed at birth (in the UK congenital forms have a prevalence of 5.7% while acquired forms of 7%) (Acerini et al. 2009). Although cryptorchidism is often considered a mild malformation, it represents the best-characterized risk factor for infertility and testicular cancer in adulthood. Men with a history of cryptorchidism are frequently subfertile in adulthood due to spermatogenic impairment, which is most frequently observed in bilateral forms. In fact, abnormal sperm parameters are found in 80% of men with a history of bilateral cryptorchidism and in 50% of men with a history of unilateral cryptorchidism. Furthermore, cryptorchidism is one of the most common causes of secretory azoospermia and is accountable for 20% of all azoospermic patients (Hadziselimovic 2006). Moreover, cryptorchid boys have 4 times higher incidence of testis cancer than general population (Cortes et al. 2001) .

However, the mechanism by which cryptorchidism may contributes to carcinogenesis is largely unknown. The so-called Testicular Dysgenesis Syndrome (TDS) theory hypothesized the existence of common etiological factors for cryptorchidism, testis neoplasia, hypospadia and spermatogenic impairment. According to this theory, these pathological conditions may reflect an underlying gonadal disgenesis resulting from the combined action of genetic and intrauterine exposure to environmental factors, including endocrine disrupting chemicals. Therefore, cancer would be the result of such a disturbed gonadal development in which altered gonocytes fail to differentiate and progressively degenerate up to carcinoma *in situ*. This theory is supported by the evidence that the increased incidence of cryptorchidism, observed in some populations, has paralleled that in testicular cancer, hypospadia, and male infertility (Herrinton et al. 2003).

3.2.1 Physiology of testicular descent

Testicular descent is a complex process requiring the interaction of anatomical and hormonal factors. The most accepted theory describes the descent of the testis as a biphasic process consisting of a first transabdominal migration, that in human embryo occurs between the 10th and the 23rd gestational week, followed by an inguinoscrotal descent that starts at around the 26th week and ends between the 28th week and birth (Hutson et al. 1997). Nevertheless, a division into stages is both artificial and arbitrary, as overall the functional processes related to testicular descent can be regarded as continuous and spreading over the majority of intrauterine life (Bay et al. 2011).

Anatomical factors involved in testis descent

The entire process is guided by two mesenteric ligaments: the cranial suspensory ligament (CSL) and the caudal genitoinguinal ligament or the *gubernaculum*. An early inner descent of the undifferentiated gonad occurs in both sexes during gonadal development and the differential evolution of these two mesenteric ligaments is responsible for the sexually dimorphic position of gonads in the adults. Initially, CSL maintains the undifferentiated gonads attached to the posterior abdominal wall in a pararenal position. In the female fetus, ovaries do not descend further through the abdomen due to the regression of the *gubernaculum* and the development of the CSL. In the male fetus, an opposite phenomenon occurs. The *gubernaculum* connects the testis via the epididymis to the future intraabdominal inner ring of the inguinal canal (Fig. 2A). The transabdominal phase of testicular descent is the result of the vector sum of traction by the CSL and the *gubernaculum* (Hutson et al. 1997). The enlargement of the abdominal cavity and the pressure of the abdominal visceral growth, maintains testes close to the future inguinal region. The CSL regresses, whereas the *gubernaculum* develops its caudal segment into the so-called gubernacular bulb, a reaction called the "swelling reaction" or "gubernacular outgrowth," protruding into the forming scrotal sac. The swelling reaction of the *gubernaculum* holds testis very close to the future internal inguinal ring, and this causes the transabdominal migration of the testes into the inguinal region (Fig. 2B).

During the following inguinoscrotal phase, as consequence of the shortening of the gubernacular cord and the outgrowth of the gubernacular bulb (Fig. 2C and D) the testes move from the inguinal region to the bottom of the scrotum.

Hormonal regulation

A critical role in testicular descent is certainly played by testicular hormones (Hutson et al. 1997; Nef and Parada 2000; Kaleva and Toppari 2003) such as testosterone and insulin-like factor 3 (INSL3) (produced by Leydig cells), anti-Mullerian hormone (AMH) (produced by Sertoli cells)(Fig. 2). Testosterone acts both on the CSL and *gubernaculum* ligaments, stimulating at the same time the development of Wollfian derivatives; AMH causes involution of the Müllerian ducts, and INSL3 controls the *gubernaculum* differentiation. The two phases of testicular descent displayed a differential regulation with a first transabdominal phase, essentially INSL3-dependent, and a second inguinoscrotal phase, androgen (testosterone)-dependent. The role of INSL3 and of its receptor, relaxin family peptide receptor 2 (RXFP2) in the physiopathology of testicular descent, will be extensively described below. The Androgen Receptor (AR) mediates the physiological effect of androgens on the two mesenteric ligaments and its expression by mesenchimal cells of the *gubernaculum* is ligand-dependent (Bentvelsen et al. 1994) so that AR expression or AR stabilization increases in males and declines in females. An equally important role in the hormonal regulation of testicular descent is played by Sertoli cells, as responsible for the synthesis of AMH. This circulating hormone was considered as involved in the transabdominal phase. This was derived from observations

that in humans with genetic defects in the *AMH* gene or its receptor with the so-called persisting Müllerian duct syndrome, the testes are undescended and the *gubernaculum* is thin and elongated (Guerrier et al. 1989; Hutson et al. 1997). The latter defect suggested that the gubernacular swelling reaction fails to occur in persisting Müllerian duct syndrome, leading to cryptorchidism.

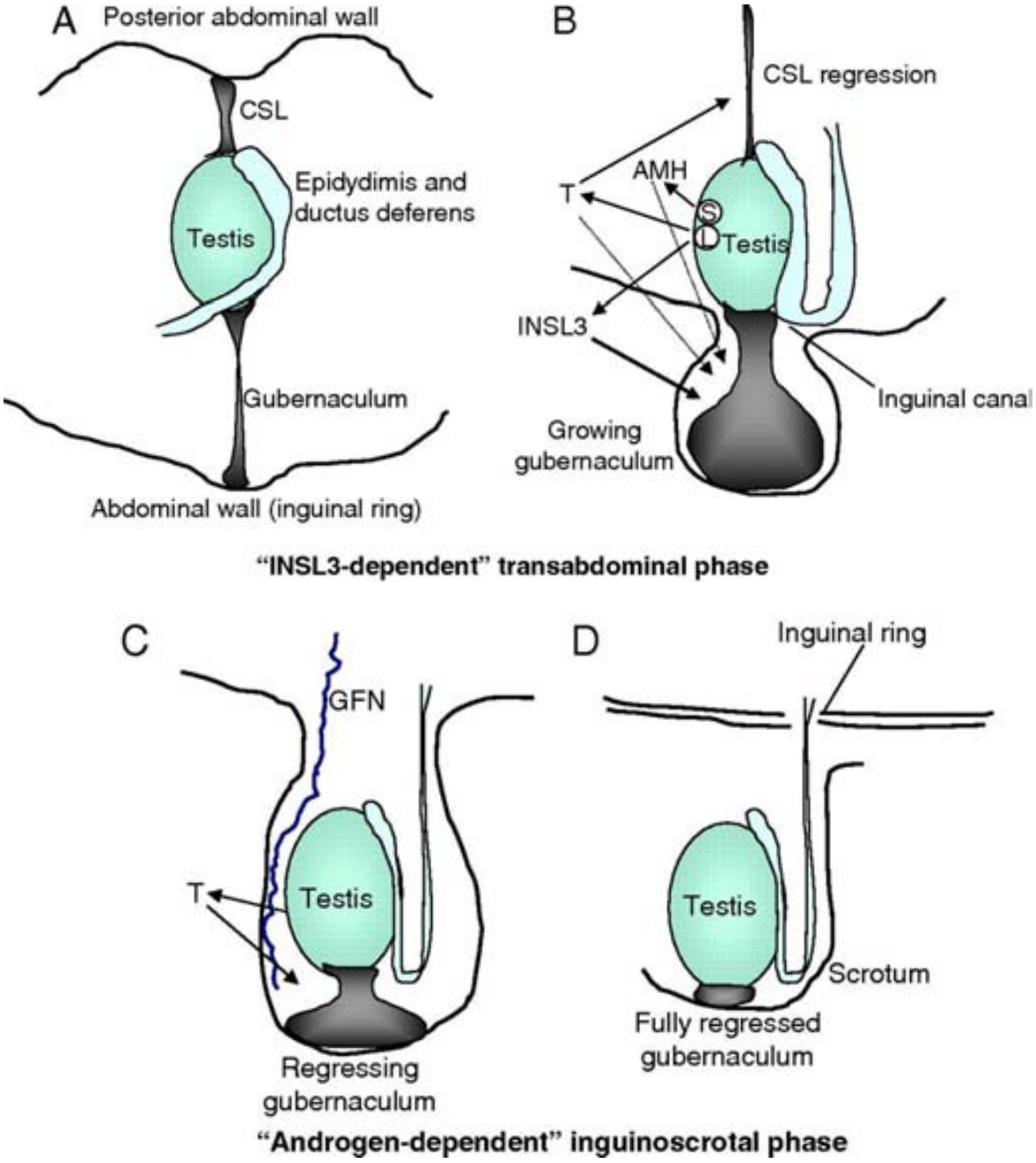


Figure 2. Model of testicular descent in humans, showing the INSL3-dependent transabdominal phase and the androgen-dependent inguinoscrotal phase. The major structures and the roles of hormones are shown. Testicular differentiation from the ambisexual gonad in the presence of the Y chromosome (A) led to the production of AMH from the developing Sertoli cells (S) and production of testosterone (T) and INSL3 from the Leydig cells (L) (B). The direct and indirect (via the Genitofemoral nerve –GFN- and the Calcitonin gene-related peptide-CGRP) effects of these two hormones principally on the CSL and *gubernaculum* cause the two-step process of testicular descent. Regression of CSL is mainly under the control of testosterone (B). Masculinization of the *gubernaculum* is under the major control of INSL3 (B) whereas minor roles seem to be exerted by AMH and androgens, possibly via the GFN and CGRP (C). Figure of *Foresta et al 2009*

3.2.2 Etiopathogenesis of non-syndromic cryptorchidism

Cryptorchidism can be accompanied by other congenital anomalies or included in complex syndromes. These forms of cryptorchidism, defined as syndromic, often result from chromosomal aberrations, as in the case of Klinefelter syndrome, Noonan Syndrome or Down syndrome or from single gene mutations. Undescended testis is indeed a typical feature of central congenital hypogonadism and can be observed also in the Partial forms of Androgen Insensitivity Syndrome (PAIS) caused by mutations in *AR* gene.
Nevertheless, hypogonadotropic hypogonadism and syndromic forms can explain only a small part of the cases of cryptorchidism.
The etiology of the most common forms of cryptorchidism, defined as "non-syndromic" or "isolated", remains largely unknown. Nevertheless, it is likely to be complex and multifactorial, reflecting the involvement of endocrine, environmental and genetic factors.
Due to the pivotal role played by testosterone/AR and INSL3/RXFP2 systems in controlling the physiological process of testis descent, it was predicted that mutations in genes coding for the two "hormone/receptor" systems might potentially cause cryptorchidism.
Mutations in *AR* are clearly involved in the etiology of certain forms of syndromic cryptorchidism (PAIS), whereas they appear not to be a frequent cause of non-syndromic forms. In fact, to date *AR* has been fully sequenced in a total of 792 cryptorchid patients (Ferlin et al. 2006c; Ferlin et al. 2008) and mutations have been found only in 4 individuals.
Concerning *AR* gene polymorphisms, two polymorphic sites in exon 1 of the *AR* gene have been studied in relation to non-syndromic cryptrochidism: the CAG and GGN repeats encoding an uninterrupted polyglutamine and polyglycine tract, respectively, in the N-terminal transactivation domain receptor. As for the CAG repeat polymorphism, the normal distribution of the $(CAG)_n$ is reported as 6–39 repeats, with a median of 21–22 in White Caucasian, 19–20 in African-American, 22–23 in Asian, and 23 in Hispanic populations. It is commonly accepted that the length of the polyglutamine tract influences the transactivation capacity of the receptor in an inverse manner; that is, the longer the tract, the lower the activity. To support this hypothesis, a clear negative impact on AR activity is documented in relationship with pathological expansions of the repeat length (40 or more), known as the Kennedy syndrome (La Spada et al. 1991). This syndrome is characterized by spinobulbar muscular atrophy and hypoandrogenism due to partial androgen insensitivity. Multiple studies have investigated a possible association between CAG and GGN repeat lengths within the normal range and the risk of isolated cryptorchidism. In humans, some (Aschim et al. 2004; Castro-Nallar et al. 2010) but not all, (Ferlin et al. 2005) studies reported that a GGN repeat length of 24 is more prevalent in cryptorchid men. Conversely, literature had unanimously dismissed long CAG repeat length as a risk factor for cryptorchidism (Krausz 2012) although specific GGN and CAG repeat combinations have been implicated (Ferlin et al. 2005; Castro-Nallar et al. 2010). Recently, the reported association between shorter CAG repeats (CAG <=19 *vs.* CAG>=20) and cryptorchidism risk, has re-opened the debate on the clinical relevance of this polymorphism in relation to the etiology of isolated disorder of testicular descent (Davis-Dao et al. 2012).

3.2.3 Importance of INSL3-RXFP2 system in the physiopathology of testicular descent

INSL3, initially named Leydig insulin-like peptide (Ley-I-L) (Adham et al. 1993) and also known as relaxin-like factor (RLF), is a peptidic hormone member of the relaxin-like hormone family. It is produced by pre- and post-natal Leydig cells under the differentiation action of the luteinizing hormone (LH) and it is therefore regarded as a sensitive marker of Leydig cell function and differentiation status (Foresta et al. 2004; Bay et al. 2005; Bay et al. 2006; Ferlin

et al. 2006a). INSL3 is synthesized as an immature pre-prohormone, composed of A and B chains connected by a C-peptide. During the processing of the inactive hormone, the N-terminus and the connecting C-peptide are removed and two inter-chain and one intra-chain disulfide bond are formed in the active hormone.

The dynamic of INSL3 production is well known. INSL3 is produced from the developing testis under the influence of maternal human chorionic gonadotropin (hCG) and/or fetal LH to control the first phase of testicular descent. Indeed, hormone concentration can be measured in amniotic fluid surrounding male fetuses in gestational week 15–21. INSL3 levels, detectable in cord blood, undergo a transient increase during minipuberty in early postnatal life, presumably stimulated by the transient postnatal LH peak (Bay et al. 2007); a decline is then observed until puberty, when INSL3 is again produced under the influence of LH secretion typical of pubertal development (Wikström et al. 2006; Ferlin et al. 2006b); INSL3 concentration is maintained during adulthood for yet unknown endocrine and paracrine roles (Foresta et al. 2004; Bay et al. 2005) and finally, it decreases in advanced age (Anand-Ivell et al. 2006).

INSL3 has been shown to act through a G-protein-coupled receptor called RXFP2, also known as leucine-rich repeat-containing G protein-coupled receptor 8, LGR8), which belongs to the super-family of the glycoprotein hormone receptors FSHR (Follicle-stimulating hormone receptor), LHR (Luteinizing hormone receptor) and TSHR (Thyroid stimulating hormone receptor). These receptors display a common structure characterized by an extracellular N-terminus domain containing several Leu/Ile-rich regions, seven transmembrane α-helices connected by three intracellular and three extracellular loops, and finally an intracellular C-terminus responsible for the intracellular signal transduction of the cAMP.

Evidence of the role of INSL3-RXFP2 in the physiopathology of testis descent is based on mouse models suggesting that this hormone-receptor system may be responsible for gubernacular differentiation during the transabdominal phase of testis descent. Indeed, INSL3 has been shown to act directly on the *gubernaculum* stimulating the mesenchymal gubernacular cells proliferation (Kubota et al. 2002) via its specific receptor RXFP2, the expression of which is restricted to testis, brain and skeletal muscles, but reaches the highest levels in the *gubernaculum* (Hsu et al. 2002; Bay et al. 2005). The *Rxfp2* gene was identified for the first time in a transgenic mouse model (Csrp model) characterized by an insertional mutation leading to a 550 Kb deletion in the proximal part of chromosome 5. This mutation affected several genes, among them the *Rxfp2* locus, and caused high intra-abdominal cryptorchidism in homozygous males.

Insl3$^{-/-}$ and *Rfxp2*$^{-/-}$ male mice exhibit bilateral cryptorchidism, with testis located high in the abdominal cavity close to the kidney due to impaired development of the *gubernaculum* (Nef and Parada 1999; Zimmermann et al. 1999; Kubota et al. 2001; Gorlov et al. 2002). Transgenic *Insl3*$^{-/-}$ male mice overexpressing *Insl3* in pancreatic β-cells showed normal transabdominal descent of the testes. It is worth noting that the overexpression of *Insl3* in female mice is sufficient to cause ovarian descent (Adham et al. 2002). Furthermore, in male mice heterozygotes for the *Insl3* deletion there was delayed gubernacular regression and delayed testicular descent, suggesting that the descent of the testis may be dependent on the dosage of *Insl3* (Adham and Agoulnik 2004).

Evidence exists supporting the involvement of the INSL3-RXFP2 systems in testis descent also in human: for instance, INSL3 is present in amniotic fluid surrounding human male fetuses at the gestational period of gubernacular swelling and outgrowth (Anand-Ivell et al. 2008) and the INSL3 levels in cord blood in persistent cryptorchid boys are significantly lower than in healthy boys (Bay et al. 2007).

3.2.4 Genetic variants in the *INSL3* and *RXFP2* genes in human non-syndromic cryptorchidism

The identification of the fundamental role of the INSL3/RXFP2 system in the *gubernaculum* development, testicular descent, and cryptorchidism in rodents and other animals encouraged the search for mutations in the *INSL3* (16p14) and *RXFP2* (13q12.3) genes in patients affected by cryptorchidism. *INSL3* has been extensively screened for mutations and more than 1500 cryptorchid patients have been analyzed so far. Since the first mutation screening in cryptorchid men (Krausz et al. 2000a), a number of genetic variants have been reported, most of which were found both in patients and controls and were therefore considered common polymorphisms. Other *INSL3* variants, such as W69R, R73X, P93L, R102C, R102H, R105H, although found exclusively in cryptorchid patients, are unlikely to be of clinical relevance because of their location in the C-peptide, which is removed during the pre-pro-hormone processing to active hormone. Only two missense variations, the P49S and V18M, have so far been demonstrated to have a deleterious effect on the ability of INSL3 to activate its receptor *in vitro* (Bogatcheva et al. 2003; Adham and Agoulnik 2004; Ferlin et al. 2006a; El Houate et al. 2007). However, the role of these two genetic variants in the pathogenesis of cryptorchidism remains to be established.
Unlike *INSL3*, the *RXFP2* gene has been entirely sequenced in a small number of cryptorchid patients (aprox. 240 to date) (Roh et al.; Gorlov et al. 2002; Ferlin et al. 2003b; Feng et al. 2004; Yamazawa et al. 2007) and *RXFP2* genetic variants reported to date are almost exclusively represented by two missense substitutions: the T222P, that will be discussed in detail in the next paragraph, and the R223K described in one single cryptorchid patient (Bogatcheva et al. 2007).
A number of issues concerning *INSL3* and *RXFP2* genetic variants detected in cryptorchid patients limit their clinical relevance. First, all variants described in humans are heterozygous, whereas in knock out mouse model cryptorchid phenotype is associated with the ablation of both alleles. Furthermore, the phenotypes associated with *INSL3* and *RXFP2* genetic variants vary from bilateral cryptorchidism, unilateral cryptorchidism, failure of the testes to descend normally in the scrotum at birth with spontaneous descent during the first years of life and, even, a normal testis descent.

The T222P variant

Studies investigating *RXFP2* mutations in relation to cryptorchidism have primarily focused on one specific screening site, represented by an "A to C" substitution in the exon 8 (g.42857A>C; c.664A>C) of the gene, responsible for the replacement of Threonine 222 with a Proline in the ectodomain of the receptor (p.Thr222Pro). This mutation was first described by Gorlov et al (Gorlov et al. 2002) in a European patient with bilateral cryptorchidism and represents the most frequent and, at the same time, the only variant with a potential clinical interest reported so far. Based on the Gorlov's study, all the subsequent studies have focused specifically on the screening of the T222P variant with more than 4000 subjects screened until now As all the other variants described in *INSL3* and *RXFP2* genes, the T222P has always been found in heterozygosity and associated with a wide range of phenotypes: from bilateral to unilateral cryptorchidism and retractile testis. According to the first two studies, this mutation was present in heterozygosity exclusively in men with history of testicular maldescent (Gorlov et al., 2002; Ferlin et al., 2003). In addition, *in vitro* studies have demonstrated that the T222P variant severely compromise the INSL3 signaling. In fact, the INSL3/RXFP2-mediated cAMP production in cells transfected with a T222P mutant receptor is strongly decreased due to a reduction of the receptor surface expression that renders the protein functionally inactive (Bogatcheva et al. 2007). These findings led initially to suggest a clear-cut cause-effect relationship between the heterozygous mutation and cryptorchidism that was though questioned by a later multicenter study, which clearly demonstrated that this

genetic variant might not be considered further as a pathogenic mutation as it can also be found in men with no history of testicular maldescent (Nuti et al. 2008). Another study from Morocco also reported a high frequency for this variant in the range of polymorphism in both cryptorchid and non-cryptorchid men (El Houate et al. 2008). Despite these controversies, the T222P variant has been reported as a genetic alteration associated with persistent bilateral cryptorchidism in a study performed on newborn children from Italy (Ferlin et al. 2008). The role of T222P variant in Spain has not been fully investigated and only a small study population (77 patients and 48 controls) was screened in the multicenter study (Nuti et al. 2008). In the aforementioned study, the frequency of T222P mutation was 0% in the patient and 4% in the control groups, showing an inverse situation to that observed in the Italian population in which 5% of patients had the mutation *vs.* 1.8% of the controls. The distribution of T222P shows marked geographic differences and since the first studies it became evident that it is prevalent in the Mediterranean area (only two carriers have been described from Hungary) (Ferlin et al. 2003b; El Houate et al. 2008; Nuti et al. 2008), and it is completely absent in other geographical areas such as Northern Europe and Asia. Previous studies suggested that all (except three) T222P carriers share a common inferred haplotype: C-C-G-A-13 (Ferlin et al. 2003b; El Houate et al. 2008; Nuti et al. 2008). This finding leads to the hypothesis about a common ancestor with the C-C-G-A-13 haplotype and provides explanation for the prevalence of T222P in Mediterranean Countries as due to a founder effect.

3.2.5 Genetic-environmental factors interaction: xenoestrogens and polymorphisms of *ESR1*

The role of estrogens in testicular descent is still unclear. However, the identification of environmental factors with estrogenic activity able to disrupt the endocrine homeostasis and hence the normal development of the male urogenital tract (Toppari et al. 1996; Olesen et al. 2007), led to an increased interest on investigating the possible role of such substances as "endocrine disruptors" of testicular descent (Foresta et al. 2008) and thus hypothesize their implication in the etiology of human cryptorchidism.

The so-called "Environmental Endocrine Disruptors" (EEDs) include xenoestrogens (industrial chemicals), synthetic and natural hormones, phyto- and mycoestrogens, organohalogen pollutants (which accumulate in the food chain and are assumed by humans through diet) and phthalates contained in industrial products, which are considered to be ubiquitous environmental contaminants (Giwercman et al. 2007). As already mentioned, according to the TDS theory proposed by Skakkebaek (Skakkebaek et al. 2001), cryptorchidism, hypospadias, testicular cancer, and spermatogenic impairment may all be symptoms of the same entity called TDS, which origins during fetal life due to the combined action of genetic and environmental factors, such as the exposure to substances affecting endocrine signalling (Skakkebaek et al., 2001). This hypothesis is supported by evidence that exposure to exogenous estrogens and antiandrogens cause disorders of genital development in animals (Olesen et al. 2007). Furthermore, the so-called "estrogen hypothesis" postulated that the apparent increase in TDS incidence observed in the last decades, may be related to increased estrogen exposure of the human fetus (Sharpe and Skakkebaek 1993). The formulation of this theory was prompted by the evidence of an association between maternal exposure to Diethylstilbestrol (DES, a synthetic nonsteroidal estrogen) during early pregnancy and increased incidence of reproductive malformations, among which maldevelopment of the *gubernaculum* and cryptorchidism, in the male offspring (Nomura and Kanzaki 1977; Emmen et al. 2000). However, it is unlikely that any of the identified environmental compounds could individually induce cryptorchidism or other signs of TDS in humans since all of them display weak or very weak endocrine activity when compared with DES (Sharpe 2003). Furthermore, experimental doses used by *in vitro* studies are generally

much higher than those to which mother and fetus may be exposed. A synergic action of several different low-dose xenobiotic compounds may be a more realistic scenario (Ivell and Hartung 2003).
Environmental compounds supposed to play a role in cryptorchidism mimic the action of hormones involved in testicular descent, acting therefore mainly as estrogens or anti-androgens. The most plausible mechanisms by which altered estrogen exposure could induce cryptorchidism and other male reproductive disorders are related to the disruption of fetal Leydig cell endocrine function, more precisely the suppression of steroidogenic and INSL3 synthesis activities, and the suppression of androgen production (Sharpe 2003). Among these mechanisms, the suppression of INSL3 production is probably the best characterized in relation to cryptorchidism, although the exact mechanism by which estrogens may modulate fetal Leydig cell INSL3 transcription remains unclear. The physiological responses to estrogens are known to be mediated by at least two functional isoforms of Estrogen receptor (ER), namely ERα and ERβ, encoded by two different genes in different chromosomes (6q25 and 14q23-24, respectively). Both receptors share the common structure of steroid/thyroid hormone nuclear receptor, differing in the C-terminal ligand-binding domain and in the N-terminal trans-activation domain (O'Donnell et al. 2001)
It is logical to hypothesize that the same receptors may also mediate the estrogenic effects of environmental endocrine disruptors and that (Toppari et al. 1996; McLachlan 2001) xenoestrogens may affect fetal Leydig cell endocrine functions and hence testicular descent via an ER-dependent mechanism (Cederroth et al. 2007). The detrimental effect of EEDs on testicular descent may certainly depend on the interaction with genetic factors and therefore it is plausible that different individuals display a different susceptibility to endocrine disrupting chemicals, due to their specific genetic background. This hypothesis increased the interest in investigating polymorphism in *ESR1* and *ESR2* genes for their potential involvement in the etiology of isolated cryptorchidism.
Although EEDs may bind both ERα and ERβ (Hrabovszky and Hutson 1999), *ESR1* has received more attention because ERα seems to play a clearer role in male genital and reproductive health than ERβ. In humans the only man with a homozygous *ESR1* mutation has normally descended testes (Smith et al. 1994; O'Donnell et al. 2001). Similarly, *Esr1* knockout male mice displayed a normal testicular descent but defects in cremaster muscle development were noted indicating a role for ERα in some aspects of male reproductive tract development and testicular descent (Donaldson et al. 1996). Genetic screening of the human *ESR1* gene locus has revealed the existence of several polymorphic sites (Castagnoli et al. 1987; Coleman et al. 1988; Zuppan PJ, Hall JM, Ponglikitmongkol M 1989; del Senno et al. 1992; Gennari et al. 2005). Recently, a haplotype called AGATA and its tag single nucleotide polymorphism (SNP), SNP12, was proposed as being of potential clinical importance, and it was first reported as a significant risk factor for cryptorchidism in the Japanese population (Yoshida et al. 2005). However, data on the same polymorphism in the Italian population (Galan et al. 2007) gave contradictory results, showing a higher incidence of SNP12 in controls and thus defining it as a ''protective'' factor against cryptorchidism. Although the literature on *ESR1* polymorphisms and cryptorchidism is limited to the AGATA haplotype (SNP12), *ESR1* polymorphisms have been extensively analyzed in relationship to male infertility/spermatogenic impairment or hypospadias (Galan et al. 2005; Yoshida et al. 2005; Guarducci et al. 2006; Galan et al. 2007; Watanabe et al. 2007; Galan et al. 2008).

The $(TA)_n$ Variable Number of Tandem Repeats polymorphism

One of the most promising genetic variants of *ESR1* studied in relation to disturbances of male reproductive health is represented by a $(TA)_n$ variable number of tandem repeats (VNTR) polymorphism located 1.2 Kb upstream to the first exon, in the *ESR1* promoter region. This polymorpshism is in linkage *disequilibrium* with two other *ESR1* polymorphisms suggested to be involved in the etiology of severe spermatogenic impairment (Kukuvitis et al. 2002): the

variant C351G, a Restriction Fragment Length Polymorphism (RFLP) located in the *XbaI* restriction site and the T397C RFLP located in the *PvuII* restriction site within the intron 1 of *ESR1*.
The location of the $(TA)_n$ in *ESR1* promoter implies a potential regulatory function for this polymorphism. Indeed, a number of evidence exists indicating that VNTR polymorphisms can influence the promoter activity and therefore transcription of the downstream gene. For instance, the uVNTR polymorphism located upstream of the monoamine oxidase A (*MAOA*) gene has been shown to affect the transcriptional activity of the *MAOA* gene promoter *in vitro* (Sabol et al. 1998; Denney et al. 1999). Similarly, a modulator effect on the promoter activity has been demonstrated *in vitro* for another VNTR polymorphism consisting of different alleles with 4-6 tandem repeats in the promoter region of the *CYP8A1* gene (Chevalier et al. 2001). Finally, Iwashita and colleagues (2001) demonstrated that a VNTR sequence located upstream a gene corresponding to an expressed sequence tag in 11p15 (DKFZp434A0527), can affect the activity of the Simion virus 40 (SV40) promoter on the transcription of the firefly luciferase reporter gene.
Therefore, it is plausible that variations in the $(TA)_n$ tract length have a regulatory potential on transcription of *ESR1*. Based on this hypothesis, the VNTR polymorphism has been analyzed in relation to two different estrogens-dependent processes: bone mineralization and spermatogenesis (Becherini et al. 2000; Gennari et al. 2005; Guarducci et al. 2006). In particular, it was reported a significant association between increased $(TA)_n$ tract length (i.e a greater number of TA repeats) and higher bond mineral density values. This finding corroborates the hypothesis of a regulatory function for $(TA)_n$ polymorphism, suggesting that a greater number of $(TA)_n$ repeats may be associated with a higher *ESR1* expression and consequently a higher estrogen responsivity. On the other hand the distribution of the $(TA)_n$ genotype supposed to be the functionally most active, resulted to be not significantly different between normozoospermic controls and patients with impaired sperm production. Nevertheless, subjects with higher $(TA)_n$ repeats alleles displayed a significantly lower sperm count compared to men bearing low $(TA)_n$ repeats genotypes both in patients and control group, suggesting that $(TA)_n$ repeats influences spermatogenic efficiency.
Whether the $(TA)_n$ polymorphism is associated also with non-syndromic cryptorchidism has not been explored so far.

3.3 Genetic factors influencing spermatogenesis

The progress of molecular genetics in the last 30 years and the advent of new diagnostic tools allowed the identification of genetic anomalies responsible for spermatogenic impairment. Nevertheless, only a limited suite of tests is currently considered as essential in the evaluation of the infertile male. These include: i) The analysis of the *AR* gene in male with suspected mild form of androgen insensitivity; ii) Mutation analysis of candidate genes (such as *KAL1*, *FGFR1,KAL2*, etc.) in case of congenital hypogonadotropic hypogonadism; iii) The screening for *CFTR* gene mutations in men with CBAVD without kidney malformations.
Apart from the above mentioned genetic analyses performed only in selected cases when clear evidence of the associated phenotype exists, only two genetic tests are performed in the diagnostic workup of oligo/azoospermic men. These are the karyotype analysis and Y chromosome microdeletion screening, which are described in details below (paragraph 3.3.1).
It is worth noting that known genetic factors collectively account for only a minority of cases of male infertility (10%-15% of all infertile males) and despite major advances in the diagnostic workup of infertile males, the etiopathogenesis of testicular failure remains undefined in about 50% of cases referred to as “idiopathic”. Based on animal models the number of candidate genes involved in the regulation of a highly complex process such as spermatogenesis, is estimated to be over thousands (Hochstenbach and Hackstein 2000) and

only a small proportion of them has so far been identified and even fewer has been analyzed. Therefore, it is plausible that a wide majority of idiopathic forms of spermatogenic disturbance has a genetic origin. Nevertheless, the efforts aimed at the identification of new genetic factors potentially involved in the etiology of male infertility, have been mostly ineffective. An association, although weak, has been reported between a number of SNPs in genes with putative function in spermatogenesis but replication studies often failed to confirm the initial findings. Mutational analyses of spermatogenesis candidate genes have been mainly performed in research context and translation of results into clinical practice is lacking (Matzuk and Lamb 2008; Nuti and Krausz 2008). The paucity of gene mutations/polymorphisms raised questions about the appropriateness of the used classic screening approach (Nuti and Krausz 2008). The analysis of sequence variants on a genome-wide scale in exceptionally large study populations (the so called "Genome-Wide Association Study", GWAS approach) has been used for the identification of genetic factors in several other complex diseases. In the field of male infertility, four genome-wide SNP association studies have been performed so far (Aston and Carrell; Hu et al. 2012; Kosova et al. 2012; Zhao et al. 2012). The first pilot study, based on a small number of non-obstructive azoospermic (NOA) men and normozoospermic controls of European descent (Aston and Carrell), and the extended follow up study on 172 SNPs performed by the same group (Aston et al. 2010), provided evidence for some SNPs as potential risk factors and new candidate genes for impaired sperm production. However, these findings have not been confirmed by two subsequent GWA studies based on exceptionally larger series of NOA subjects (about 1000 cases) and controls (more than 1500 subjects) from the Han Chinese population (Hu et al. 2012; Zhao et al. 2012). Surprisingly, the results from these two studies, based on the same population, do not display any overlap, since different genomic regions have been reported as significantly associated with NOA phenotype by Zhao et al (1p13.3, 1p36.32 and 12p12.1) and Hu et al (6p22). Finally, a third GWA study based on a cohort of men (269 individuals) from a founder population of European descent (the Hutterties) reported other SNPs associated with reduced fertility in the Hutterties and confirmed as significantly associated with reduced sperm quality and/or function in 123 ethnically diverse infertile men. The lack of consistency between the above GWA studies may be only partially related to genuine ethnic differences since not even the two large GWA Chinese studies reported overlapping candidate SNPs. The failure of the GWAS approach in identifying relevant SNPs involved in spermatogenic failure may simply reflect a problem with the working hypothesis "common disease, common variants". At this regard, the Aston and Carrel study (2009) has predicted that it is more likely that "rare" variants rather than common polymorphisms are involved in the etiology of spermatogenic impairment. This would also explain the difficulty to detect these variants by SNP arrays, which are based on common genetic variants. Another explanation for the inconsistency and unsuccessful outcome of GWA studies may be that the pathogenic effect of SNPs is related to the combination of low size effect SNPs or their interaction with the environment.

In this context, a significant association with male infertility has been clearly demonstrated for the 677C>T polymorphism in the *MTHFR* (5,10-Metylene-tetra-hidrofolate reductase) gene in populations with low folate intake (Nuti and Krausz 2008).

The study of structural variations such as Copy Number Variations (CNVs) in relation to multifactorial complex diseases represented another interesting research field that allowed the identification of novel genetic factors involved in the etiology of some type of cancer, neurological and autoimmune diseases or susceptibility to human immunodeficiency virus (HIV-1) infections. It is plausible that CNVs affecting regions or multicopy genes relevant to spermatogenesis may also contribute to infertility. To date, only two Y-linked CNVs have been correlated to male infertility: the testis-specific protein Y-linked (*TSPY1*) gene copy number variation on Yp and the *AZFc* gene dosage variation due to complete/partial removal or duplication of multicopy *AZFc* genes that are object of study of the present thesis.

3.3.1 Karyotype anomalies

Chromosomal anomalies can affect both number and structure of chromosomes and arise mainly during meiosis. Meiotic errors are indeed extraordinarily common in humans, being the frequency of chromosome abnormalities at least an order of magnitude higher than in other mammals: approximately 21% of oocytes and 9% of spermatozoa display abnormal chromosomal complements (Martin 2008). Chromosomal aberrations, either numerical or structural in nature, have an approximately 0.4% incidence in the general population and can have profound effects on male fertility (Harton and Tempest 2012). In male with disturbance of spermatogenesis, karyotype aberrations display a 4%–16% incidence. This figure has been clearly demonstrated to increase proportionally with increasing severity of testicular phenotype (Johnson 1998; Shi and Martin 2001). Patients with less than 10 million spermatozoa/ml show 10 times higher incidence of karyotype anomalies (4%) compared to the general population. Among severe oligozoospermic men (less than 5 million spermatozoa/ml) the frequency increases to 7%-8% whereas in non-obstructive azoospermic men it reaches the highest values, 15%-16%.

Numerical anomalies: Klinefelter syndrome

Klinefelter syndrome (47,XXY) is a common chromosomopaty affecting 1:600 newborn males in the general population (Bojesen et al. 2003) and represents, at the same time, the most common genetic cause of secretory azoospermia. Men with Klinefelter syndrome, in fact, are thought to make up 3% of infertile men and 11% of men with azoospermia (Foresta et al. 1999). About 80% of patients bear a pure 47,XXY karyotype whereas the other 20% is represented either by 47,XXY/46,XY mosaics or higher grade sex chromosomal aneuploidy and structural abnormal X chromosome (Krausz 2011). The extra chromosome is inherited either from the mother or the father at an approximately equal ratio (Thomas and Hassold 2003). Although the high prevalence in the general population, Klinefelter Syndrome is a profoundly under diagnosed condition. Epidemiological studies have shown that only 25% of adult males with Klinefelter Syndrome are ever diagnosed, and diagnosis is rarely made before the onset of puberty (Bojesen et al. 2003). The syndrome is generally diagnosed at three main stages in life: prenatally; around school age mainly because of tall stature, learning and behavioral disabilities; or in adulthood mainly because of infertility (Bojesen et al. 2003; Aksglaede et al. 2011b). The phenotypic appearance of adult Klinefelter patients varies widely; nevertheless gynecomastia, sparse facial and body hair as well as small firm testes with a mean volume of 3.0 ml (range 1.0-7.0 ml) are common features. Adults with Klinefeter Syndrome are characterized by hypergonadotropic hypogonadism with highly elevated serum concentrations of follicle-stimulating hormone (FSH) and luteinizing hormone (LH). Testosterone serum concentration is most often within the lower half the reference range of healthy males, and rarely below the reference range. Inhibin B is below the detection limit in the vast majority of Klinefelter adults reflecting the absent spermatogenesis (Aksglaede et al. 2008), whereas the circulating concentrations of AMH and INSL3 are significantly reduced compared with healthy males (Foresta et al. 2004; Bay et al. 2005; Aksglaede et al. 2011a).

The degeneration of germ cells, through a not fully understood mechanism underlies the infertility status of Klinefelter patients. This testicular deterioration starts already *in utero*, progresses during infancy and early childhood, and accelerates during puberty and adolescence, eventually resulting in extensive fibrosis and hyalinization of the seminiferous tubules and hyperplasia of interstitium in the adult patient (Aksglaede et al. 2006). The development of new advanced assisted reproductive techniques such as testicular sperm extraction (TESE) combined with intracytoplasmic sperm injection (ICSI) had increased the chance of fatherhood among Klinefelter patients although the large majority of them are azoospermic. Based on the existing literature, (Aksglaede and Juul 2013)reported an average

sperm retrieval rate of 50%, ranging from an average of 42% by the use of TESE to an average of 57% by the use of micro-TESE. Moreover, spermatozoa can be even found in the ejaculate of mainly mosaic patients or in non-mosaic but young patients indicating the potential importance of an early diagnosis that would allow a preventive cryopreservation of ejaculated spermatozoa to preserve fertility. An extensively debated issue in this context concerns the genetic integrity of these gametes. Fluorescent in-situ hybridization (FISH) analysis of ejaculated or testicular spermatozoa in Klinefelter Syndrome have shown varying frequencies of normal spermatozoa ranging from 50.0 to 93.7% (Guttenbach et al. 1997; Estop et al. 1998; Levron et al. 2000). Accordingly, it has been proposed that adults with KS have a substantially higher proportion of hyperhaploid spermatozoa (46,XY and 46,XX) than healthy males (Foresta et al. 1998; Hennebicq et al. 2001), giving these males a theoretically increased risk of fathering a child with 47,XXY or 47,XXX (for review see (Maiburg et al. 2012). Furthermore, an increased frequency of autosomal aneuploidy 13, 18, and 21 in spermatozoa from Klinefelter Syndrome has been proposed (Hennebicq et al. 2001; Morel et al. 2003; Staessen et al. 2003). A study based on ICSI combined with preimplantation genetic diagnosis (PGD) reporting a significant fall in the rate of normal embryos for couples with Klinefelter syndrome compared to controls (54% versus 77,2%) (Staessen et al. 2003). However, according to recent reviews children born from Klinefelter fathers are healthy and only one 47,XXY fetus has been reported so far (for review see (Lanfranco et al.; Staessen et al. 2003; Fullerton et al. 2010)and references therein). Despite this encouraging data, due to the significant increase of sex chromosomal and autosomal abnormalities in the embryos of Klinefelter patients, ICSI+PGD should be an appropriate preventive option (Staessen et al. 2003).

Structural anomalies: autosomal abnormalities

Oligozoospermic men with <10 million spermatozoa display a higher risk to have autosomal abnormalities than azoospermic men (Vincent et al. 2002). Reciprocal translocations, Robertsonian translocations, paracentric inversions and marker chromosomes are the most frequently found abnormalities.

Reciprocal translocations are the most common type of translocation (1:930 incidence in the general population) and can involve any of the chromosomes. These structural anomalies occur as consequence of the formation of breaks in two non-homologous chromosomes followed by the abnormal repair of the chromosomal fragments. The result is the transposition of genetic material from one chromosome to the other that can occur with or without loss/gain of genetic material and are hence defined as unbalanced and balanced translocations, respectively. Phenotypic effects of reciprocal translocation may be related to the deregulation of gene expression both because one of the translocation breakpoint can interrupt gene sequence, or via position effects (translocation of the gene into a region in which expression is either up or down-regulated). Furthermore, in order to pair during meiosis, the translocated chromosome and its non-translocated homologous are forced to align themselves in a cross shape, forming a structure known as a quadrivalent. This phenomenon can affect meiosis in several ways. First, the mechanics and time constraints imposed on the formation of such a structure can trouble the normal progress of meiosis. Secondly, asynaptic regions are common within the pairing cross and can lead to meiosis failure. Moreover, there is evidence that translocated segments of chromosomes attempt non-homologous pairing with X and Y chromosomes during meiosis I, which interferes with X inactivation, resulting in a lethal gene dosage effect on the germ cells. Finally, the interactions of the translocation chromosomes with other parts of the nucleus may produce errors in meiosis and cell death.

Robertsonian translocations are structural chromosomal aberrations involving the acrocentric chromosomes, specifically chromosomes 13, 14, 15, 21 and 22. In this instance,

the translocation arises as the result of a centromeric fusion of two acrocentric chromosomes. This type of translocation can occur between homologous as well as between non-homologous chromosomes. Their frequency is 0.1% in the general population, 1.1% in couples with recurrent fetal loss and 2–3% in infertile men. The most frequent Robertsonian translocation is the one involving chromosomes 13 and 14 (der (13;14) (q10;q10)) which accounts for about 75% of all Robertsonian translocations (Keymolen et al. 2009). Also in this case, the translocated chromosomes are forced to synapse through a pairing cross (trivalent structure) with all the abovementioned deleterious effect on meiosis.
A chromosomal inversion occurs when a segment of a chromosome is excised, inverted of 180°, and reintegrated into the same chromosome. These structural anomalies are classified into pericentric and paracentric depending on whether the centromeric region is involved or not in the inverted segment. Chromosomal inversions are found in 0.02% of newborns, with the exception of inversions affecting the heterochromatic region of chromosome 1, 9 and 16, which are considered as common polymorphisms (frequency >1%). Phenotypic effects of chromosomal inversions may be related to the deregulation of gene expression when the excision site is within the regulatory or structural region of a gene. Furthermore, similarly to chromosomal translocations, during meiosis inverted chromosomes are forced to form specialized structures (inversion loops) to enable homologous pairing. The mechanics and time constraints associated with the formation of the inversion loop may prevent the normal progression of meiosis, furthermore it has been demonstrated that recombination within these loops is reduced which can lead to a breakdown in meiosis and hence to the germ cells apoptosis (Brown et al. 1998).
The importance in detecting these structural chromosomal anomalies is mainly related to the increased risk of aneuploidy or unbalanced chromosomal complements in gametes and hence in fetus. Carriers of balanced chromosomal translocations, although phenotipically normal in the vast majority of the cases, may in fact experiment spontaneous abortions and birth defects in the offspring because the normal meiotic segregation of these translocations in the gametes leads to duplication or deletion of the chromosomal regions involved in the translocation. In case of Robertsonian translocations a special risk is represented by uniparental disomies, which are generated through a mechanism called “trisomy rescue” (repairing of trisomy) during the first division of the zygote. For chromosome 14 (the most frequently involved chromosome) and 15, both paternal and maternal uniparental disomies are pathological and give rise to severe disease such as Angelman or Prader–Willi syndromes despite an apparently normal or balanced karyotype.

Y chromosome terminal deletions (Yq-)

Terminal deletions of the long arm of the Y chromosome, including the terminal heterochromatic band Yq12, are visible at the karyotype analysis and represent a relatively frequent cause of azoospermia. Such large terminal deletions of the Yq can also results from complex structural abnormalities of the Y chromosome such as isodicentric (idicYp) and isochromosomic (isoYp) Y chromosome. The idicYp is characterized by the duplication of the Yp and of the most proximal region of the Yq, including the centromer, while showing the deletion of the terminal part of the Yq. The isoYp is a monocentric Y chromosome (only one centromer is present) showing two Yp and lacking any Yq material.
IdicYp and isoYp chromosomes are among the more common genetic causes of severe spermatogenic failure in otherwise healthy men. IdicYp or isoYp formation likely interferes with sperm production via several distinct mechanisms. First, many idicYp and all isoYp chromosomes lack distal Yq genes that play critical roles in spermatogenesis (Skaletsky et al. 2003) Secondly, idicYp or isoYp formation results in duplication of the Yp pseudoautosomal region and deletion of the Yq pseudoautosomal region, outcomes that may disrupt meiotic pairing of chrX and chrY and thereby preclude progression through meiosis (Mohandas et al. 1992).

The presence of two centromeric regions makes idicY chromosomes mitotically instable. As observed in many human dicentric chromosomes, the mitotic stability of idic Yp, especially those with greater intercentromeric distances, is likely to rely upon the functional inactivation of one of the two centromeric regions. Notwhitstanding, these chromosomes tend to be lost during mitosis leading to the generation of 45,X cell lines (45, X mosaicism).

3.3.2 Copy Number Variations (CNVs)

Definition of CNV

A Copy Number Variation (CNV) is conventionally defined as a DNA segment of at least 1 Kb in length that is present in a variable number of copies in the genome (Fanciulli et al. 2010).
The term "variation" instead of "polymorphism" is used because the relative frequency of most CNVs in the general population have not yet been well defined and the term polymorphism is reserved for genetic variants that have a minor allele frequency of ≥1% in a given population.
CNVs belong to a category of genomic alterations defined as "structural variants" (involving segments of DNA of 1 kb or larger) which includes also balanced alterations regarding position and orientation of genomic segments defined as translocations and invertions, respectively. The term CNV is not generally used to indicate variations caused by insertion/deletion of transposable elements.
These unbalanced quantitative variants can be classified into:

- **gains** when an increased number of DNA copies is observed with respect to the reference genome as consequence of duplication/amplification or insertion events. The amplified DNA fragments can be found adjacent to (tandem duplication) or distant from each other and even on different chromosomes.

- **losses** when a reduction or no DNA copies are observed with respect to the reference genome as consequence of deletion events. In the present thesis the terms "loss" and "deletion" will be used to indicate the reduction and the complete loss (null genotype) of a given DNA sequence compared to the reference genome, respectively

A CNV can be simple in structure or may involve complex gains or losses of homologous sequences at multiple sites in the genome (Fig.3).

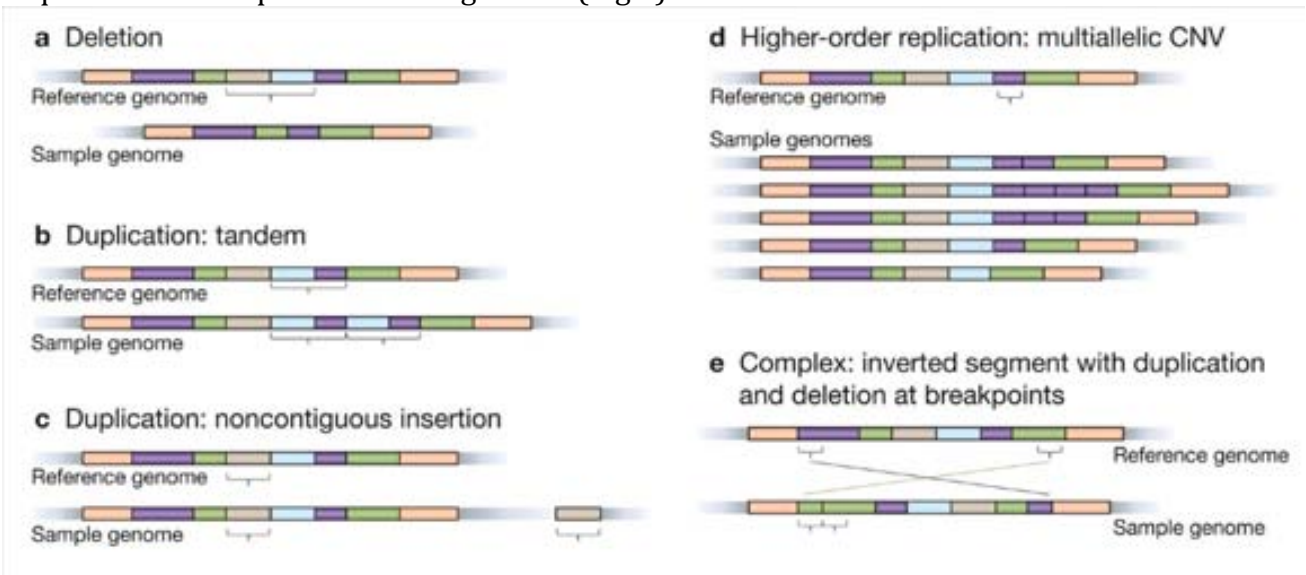


Figure 3. Different types of CNV. CNVs (in the sample genome) are defined by comparison with a reference genome. DNA blocks displaying sequence identity are represented with the same color. a) deletion of two contiguous fragments (deletion); b) tandem duplication (gain); c)

duplication (gain) with insertion of the duplicated sequence far from the origin; d) multiallelic gain produced by multiple duplication event; e) Complex CNVs resulting from inversion, duplication and deletion events. Figure of (Lee and Scherer 2010)

In 2006, Redon and colleagues (Redon et al. 2006) assembled the first comprehensive CNV map of the human genome and pointed out some interesting aspects concerning these structural variants. Based on the analysis of 270 individuals (the HapMan Collection), they found that 12% of the human genome (approximately 360 Mb pairs) results to be covered by CNVs with a preponderance of smaller size rearrangements (<20 kb). The genomic regions encompassed by these CNVs contain hundreds of genes and functional elements and many CNVs reached a population frequency greater than 1% and (Copy Number Polymorphisms).
With the growth of information on cnvs in the human genome, the accurate annotation of these structural variations has become increasingly important. at present, several databases are currently available for genome-wide investigation of genomic variants the most important of which is the "data base of genomic variants" (http://dgv.tcag.ca/dgv/app) that provides a comprehensive catalog (continuously up dated) of the structural variation identified the human genome

Functional effect and role of CNVs in complex diseases
CNVs are a widespread feature of the genomes of all healthy humans thus being mostly neutral or having only subtle influence on phenotype. beside having an important role in evolution and adaptation to different environments, as major source of genetic inter-individual variability (more than SNPs) (Iafrate et al. 2004; Sebat et al. 2004), CNVs may also have important functional effects depending on the presence of functional elements, such as genes or regulatory sequences, within or surrounding the affected genomic region. In this view, these sub-microscopic structural variants have received considerable attention by the scientific community during the last decade, for their potential role in modulating complex phenotypes and thus in the etiopathogenesis of human diseases.

CNVs may influence gene function directly by altering gene copy number (in case of dosage-sensitive genes) or by disrupting gene-coding sequence: CNVs breakpoints can disrupt gene structures leading to the formation of new transcripts through gene fusion or exon shuffling. furthermore, cnvs can influence gene expression indirectly through position effect, i.e. altering gene location with respect to regulatory elements as promoters, enhancers and silencer or disrupting the function of these, leading to changes in sequence or in location with respect to target gene (Hurles et al. 2008)(Fig.4).

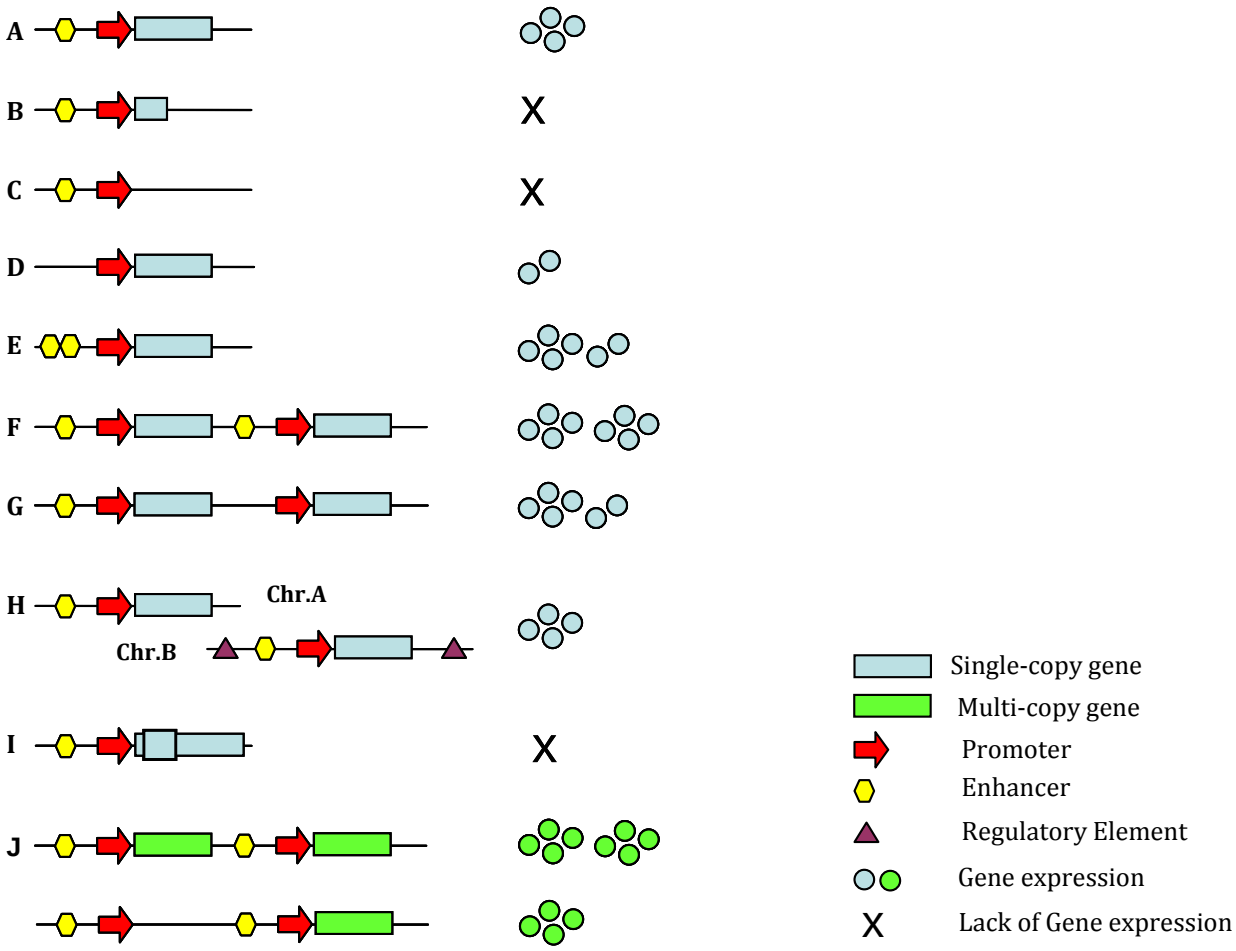


Figure 4. Impact of CNVs on gene expression. A. Single copy dosage-sensitive gene (reference genome): promoter, upstream enhancer element and coding sequence are represented; partial and complete deletion affecting coding sequence (B and C); Deletion and duplication affecting enhancer (D and E); Complete and partial (not involving enhancer) tandem duplication (F and G); Complete tandem inter-chromosome duplication involving a regulatory element inhibiting gene expression (H); Partial tandem duplication disrupting coding sequence (I); multicopy gene loss (upper part).

It is worth noting that CNVs affecting gene-lacking regions may also have phenotypic effects since they can disrupt the function of genes located even far from the genomic region directly involved in cnv. Accordingly, a number of studies reported that the effects of many of known cnvs are related to the disruption of functional elements (genes or regulatory elements) located at some mega bases of distance (Naranjo et al. 2010; Benko et al. 2011). CNV size is poorly associated to the severity of phenotype since several benign cnvs may display a 2-10 mb length (Hansson et al. 2007; Redon et al. 2006).

The number of studies dealing with CNVs and common diseases has increased markedly over the past decade (Fig. 5) Thanks to the development of new tools as genome-scanning array technologies and comparative DNA-sequence analyses, which allowed human structural genetic variation to be analyzed at a high resolution, CNVs have been associated with a growing number of common complex diseases, including HIV, autoimmune diseases such as Chron disease, psoriasis, systemic lupus erythematosus (Aitman et al. 2006; Fanciulli et al. 2007; Willcocks et al. 2008; Bassaganyas et al. 2013), a spectrum of neuropsychiatric disorders as autism, schizophrenia (Cook and Scherer 2008; Rodríguez-Santiago et al. 2010; Saus et al. 2010) and some type of cancer (neuroblastoma, breast and prostate cancer).

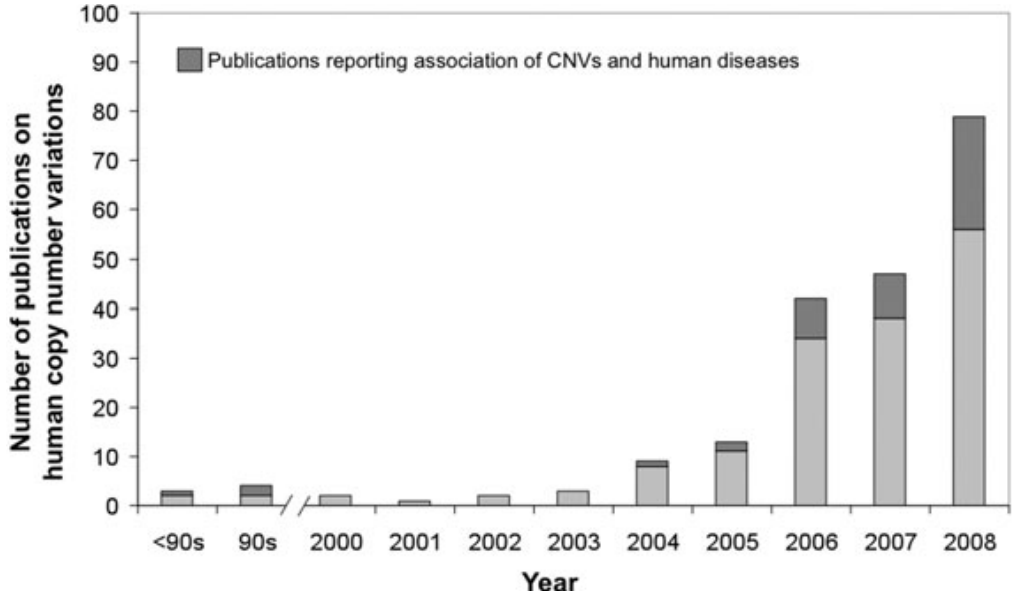


Figure 5. Number of publications reporting CNVs in humans. Bars represent cumulative data for human CNVs, dark bar indicate publications reporting association of CNVs and human diseases. Figure by *Fanciulli et al 2010*.

Mechanisms for CNV generation

Three major mechanisms, all involved in DNA double strand break repair process, are thought to account for the majority of genomic rearrangements in humans and thus represent the main molecular mechanisms for the formation of CNVs. These can be divided into two major classes:

- **Homologous recombination-based pathways** including the non-allelic homologous recombination mechanism (NAHR),
- **Non-homologous recombination-based pathways** including non-homologous end-joining (NHEJ) and the Fork Stalling and Template Switching (FoSTeS) models.

Non-allelic homologous recombination

Most recurrent CNVs occur in regions reported to contain, or be flanked by, large homologous repeats named low copy repeats (LCR) or segmental duplications (SDs) (Shaw and Lupski 2004; Stankiewicz and Lupski 2010) . SDs are defined as region-specific DNA blocks of >1 kb in size (usually 10 to 300 Kb) and of > 95%-97% sequence identity. They could arise by tandem repetition of a DNA segment followed by subsequent rearrangements that place the duplicated copies at different chromosomal loci. Alternatively, SD could occur via a duplicative transposition-like process: copying a genomic fragment while transposing it from one location to another.

The presence of SDs with the same orientation placed at less than 10 Mb from each other, is likely to account for the high susceptibility to misalignment events displayed by specific sites of the human genome. In fact, these highly homologous DNA blocks can act as substrate for the activation of NAHR mechanisms resulting in unequal crossing-over events and hence they represent hot-spot points for the generation of CNVs.

Homologous recombination is the basis of several mechanisms of accurate DNA repair that use another identical sequence to repair a damaged sequence. There will be no structural variation if a damaged sequence is repaired using homologous sequence in the same chromosomal position in the sister chromatid or in the homologous chromosome (allelic

homologous recombination). On the other hand, if a crossover forms when the interacting homologies are in non-allelic positions on the same chromosome or even on different chromosomes this will result in duplication and deletion of sequence between the repeats owing to unequal crossing over. More specifically, interchromosomal and interchromatid NAHR between LCRs with the same orientation results in reciprocal duplication and deletion, whereas intrachromatid NAHR creates only deletions (Fig. 6).

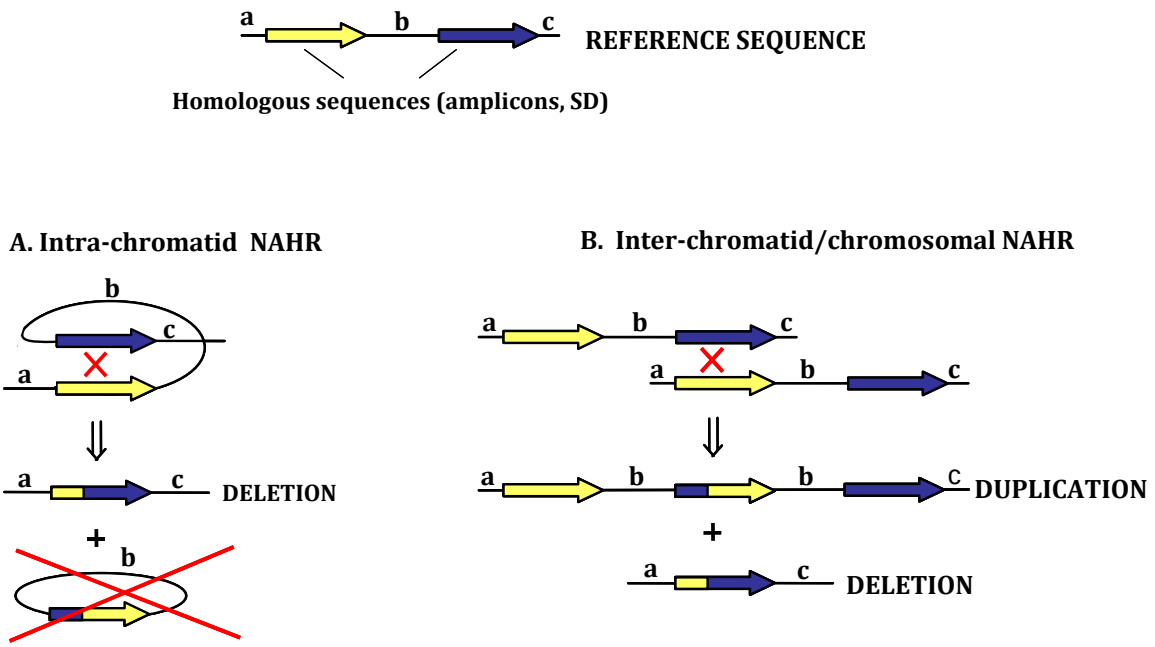


Figure 6. NAHR mechanisms. The substrates of recombination are two directly oriented SDs represented by yellow and blue arrows. Two scenarios are represented: A. Intra-chromatid NARH: recombination between two homologous sequences on the same chromatid results in the deletion of the interposed DNA segment. B. Interchromatid or interchromosome NAHR: two non allelic homologous sequence on sister chromatids or chromosomes are involved in recombination leading to a deletion and the reciprocal duplication.

Thus, theoretically, the frequency of deletions should be always higher than duplications. However, if deleterious deletions underwent negative selection, this would lead to a higher frequency of duplications (Turner et al. 2008). Therefore, duplication frequency should not exceed deletion frequency, unless negative selection in both germ cells and somatic cells makes deleterious deletions very rare or not represented.

Not all CNVs appear to be associated with SDs. It is possible that subsets of CNVs may be formed or maintained by non-homology-based mutational mechanisms. In addition to homologous recombination pathways, there are indeed mechanisms of DNA repair that use very limited or no homology. Certain CNVs may be found to be associated with non- β DNA structures (DNA regions that differ in structure from the canonical right-handed β -helical duplex, including left-handed Z-DNA and cruciforms). Such DNA structures are believed to promote chromosomal rearrangements (Bacolla et al. 2004; Kurahashi et al. 2005) and may also theoretically contribute to the genesis and maintenance of certain CNVs.

Non-homologous end-joining (NHEJ)

NHEJ is one of the two major mechanisms used by eukaryotic cells to repair DNA double strand breaks (DSBs) without involving a template DNA sequence. This non-homologous DNA repair pathway has been described in organisms from bacteria to mammals and is routinely

used by human cells to repair both 'physiological' DSBs, such as in V(D)J recombinations, and 'pathological' DSBs, such as those caused by ionizing radiation or reactive oxygen species. NHEJ proceeds in four steps (Fig. 7): detection of DSB; molecular bridging of both broken DNA ends; modification of the ends to make them compatible and ligatable; and the final ligation step (Weterings and van Gent 2004). Being a non-homology based mechanism, NHEJ does not require DNA pairing for successful ligation and, consequently, unlike NAHR it is not dependent on the presence of SDs. Evidence exists that NHEJ is more prevalent in unstable (or fragile) regions of the genome such as the subtelomeric regions (Nguyen et al. 2006; Kim et al. 2008). Furthermore, many NHEJ events, classified as microhomology-mediated end joining, require end resection and join the ends by base pairing at microhomology sequences (5–25 nucleotides)(McVey and Lee 2008; Pawelczak and Turchi 2008). NHEJ leaves a "molecular scar" since the product of repair often contains additional nucleotides at the DNA end junction (Lieber 2008).

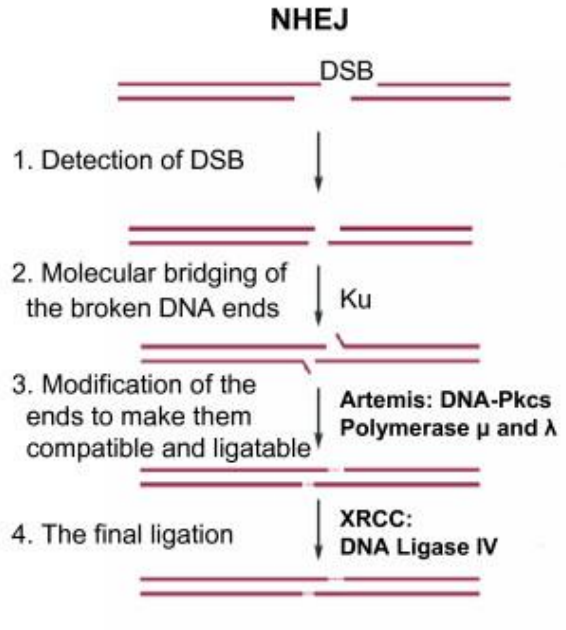


Figure 7. NHEJ mechanism: Once detected, the ends of a Double strand DNA break undergo modifications (nucleotide insertion or removal) to be compatible and ligateble by a DNA ligase. The enzymes involved in each steps are indicated Figure of (Gu et al. 2008)

Fork stalling and template switching (FoSTeS)

To explain the complexity of non-recurrent rearrangements, such as those associated with Pelizaeus-Merzbacher pathology and *MECP2* gene (methyl-CpG-binding protein 2) duplications and triplications associated with mental retardation and disturbance of development in male, Lee et al. proposed the replication Fork Stalling and Template Switching (FoSTeS) Model (Fig. 8). According to this model, during DNA replication, the DNA replication fork stalls at one position, the lagging strand disengages from the original template, transfers and then anneals, by virtue of microhomology at the 3' end, to another replication fork in physical proximity (not necessarily adjacent in primary sequence), 'primes', and restarts the DNA synthesis (Lee et al. 2007a). The invasion and annealing depends on the microhomology between the invaded site and the original site. Upon annealing, the transferred strand primes its own template driven extension at the transferred fork. This priming results in a 'join point' rather than a breakpoint, signified by a transition from one segment of the genome to another – the template-driven juxtaposition of genomic sequences. Switching to another fork located

downstream (forward invasion) would result in a deletion, whereas switching to a fork located upstream (backward invasion) results in a duplication.

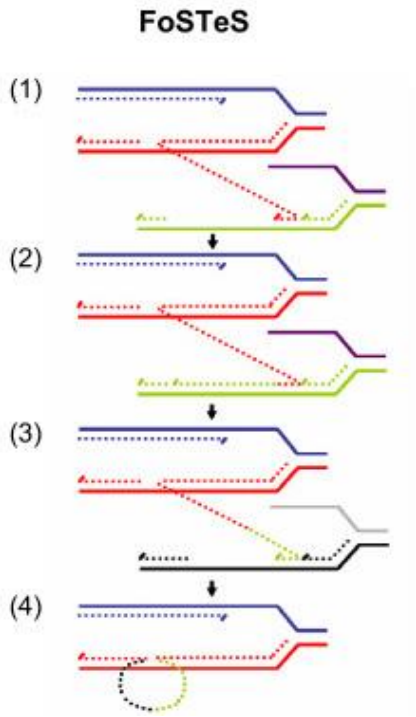


Figure 8. After the original stalling of the replication fork (dark blue and red, solid lines), the lagging strand (red, dotted line) disengages and anneals to a second fork (purple and green, solid lines) via microhomology (1), followed by (2) extension of the now 'primed' second fork and DNA synthesis (green, dotted line). After the fork disengages (3), the tethered original fork (dark blue and red, solid lines) with its lagging strand (red and green, dotted lines) could invade a third fork (gray and black, solid lines). Dotted lines represent newly synthesized DNA. Serial replication fork disengaging and lagging strand invasion could occur several times (e.g. FoSTeS x 2, FoSTeS x 3, ... etc.) before (4) resumption of replication on the original template. Figure of *Gu et al 2008.*

A relation between CNV size and associated mutational mechanism has been hypothesized. It has been shown indeed that larger CNVs are more frequently associated with segmental duplication and thus related to NAHR events, whereas among the smaller known CNVs non-homology-driven mutational mechanisms may be prevalent (Tuzun et al. 2005; Conrad et al. 2006)(Fig. 9)

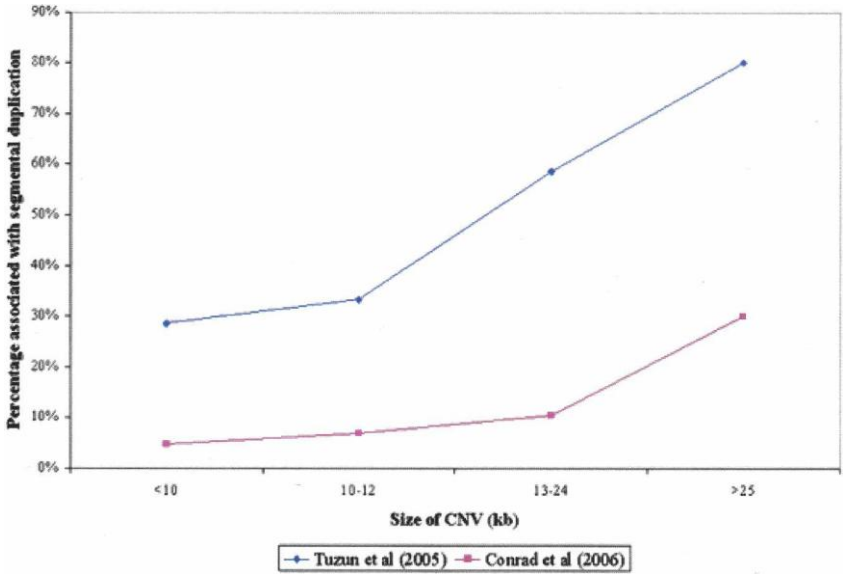


Figure 9. Positive correlation between the size of CNVs and likelihood of association with segmental duplications. This correlation is noted by both the Conrad et al. (2006) and Tuzun et al. (2005) studies. Figure of (Freeman et al. 2006)**.**

When and where CNVs occur

CNVs can occur as inherited polymorpshisms or arise *de novo* both in germ line and in somatic cells. The rate of CNV formation, estimated to be several orders of magnitude higher than point mutations, is especially high during meiosis (in germ cells). However, CNVs can also arise mitotically resulting in mosaic populations of somatic cells. This observation is supported by a number of evidence such as the report of identical monozygotic twins (having identical genomes) bearing different CNVs (Bruder et al. 2008) and the finding of different copy number of repeated sequences in different organs and tissues from the same individual (Piotrowski et al. 2008). The explanation for such interesting findings may be that CNVs must have arisen by spontaneous mutations during the early stages of embryogenesis, either just before or just after the embryo split into two individuals in case of monozygotic twins. A similar scenario may be hypothesized if a mutation occurs in an embryo that does not split leading to a copy-number chimerism in the same individual (Fig. 10).

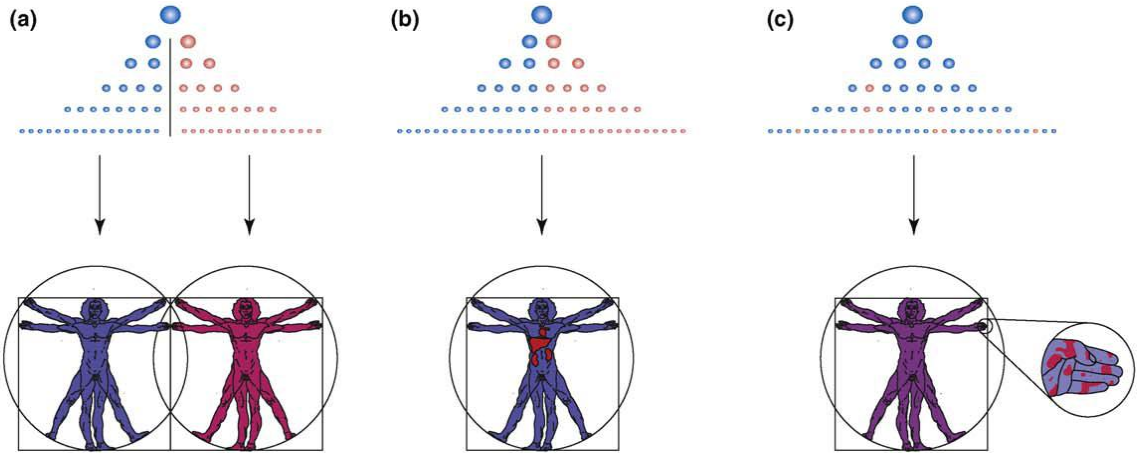


Figure 10. Somatic CNV. The upper parts of each diagram represent the process of cell division from a fertilized zygote; cells shown in red have acquired a spontaneous copy-number change that is then passed on to their descendents. (a) Copy-number differences between monozygotic twins, as reported by Bruder et al., imply that a spontaneous CNV arose at or about the time at which the developing embryo split (black vertical bar) into two individuals. The result is a pair of monozygotic twins differing at the CNV locus (red and blue human figures below). (b) If a spontaneous CNV arose at a similarly early point during the development of a single (non-twinning) embryo, the result would be an individual who is chimeric for copy-number at that locus, with only some tissues (red) showing the new copy-number variant. This phenomenon was reported by Piotrowski et al. (c) However, if CNVs arise spontaneously in later stages of embryogenesis or after birth, they will result in an individual with microchimerism, in which one or more patches of cells, or individual cells, differ in copy-number from their neighbours. Figure of (Dear 2009)**.**

3.3.3 The Y chromosome

The Y chromosome is one of the smallest chromosomes of the human genome having an approximate length of 60 Mb. It is the sole chromosome in our genome that is not essential for survival and it differs markedly from the remaining chromosomes in terms of genomic structure, gene content and evolutionary trajectory (Navarro-Costa 2012).
The human Y chromosome can be divided into two major structural genomic domains: the "Male-Specific region of the Y chromosome (MSY) and the two Pseudoautosomal regions (PAR1 and PAR2).

MSY region
The MSY region, comprising approximately 95% of the chromosome length, is a segment in which there is no X-Y crossing over and this is why it was originally considered as "non-recombining". Relatively recent studies (Rozen et al. 2003) have questioned this view demonstrating that a high level of internal homology promotes homologous recombination events within the MSY region. Our knowledge about MSY structure is based on a single reference sequence, obtained analyzing only one man's Y chromosome (Skaletsky et al. 2003). The region is a mosaic of heterocromatic and eucromatic sequences. The MSY's euchromatic DNA sequences total roughly 23 megabases (Mb), including 8 Mb on the short arm (Yp) and 14.5 Mb on the long arm (Yq). In addition to the large block of centromeric heterochromatin (approximately 1 Mb) found in every nuclear chromosome, the Y chromosome contains a second, much longer heterochromatic block (roughly 40 Mb) that comprises the bulk of the distal long arm.

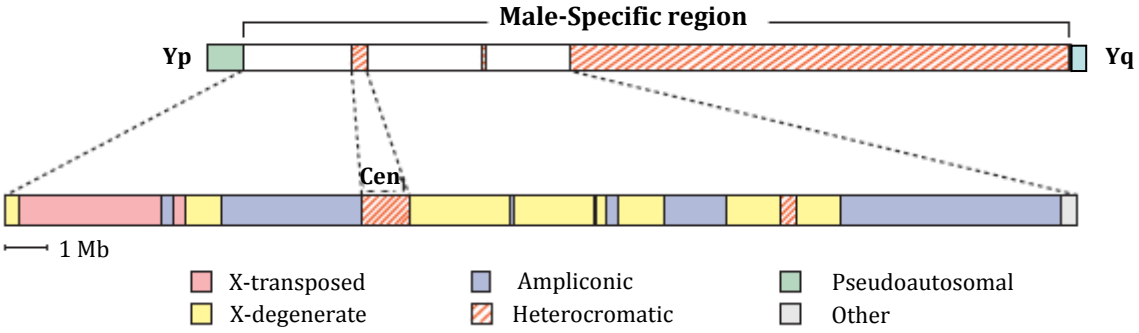


Figure 11. Schematic representation of the whole Y chromosome, including the pseudoautosomal MSY regions. Heterocromatic segments and the three classes of euchromatic sequences (X-transposed, X-degenerate and Ampliconic) are shown.

The MSY region is a patchwork of three different sequence classes (Fig.11):

- **X-transposed sequences** which are so named because their presence in the human MSY is the result of a massive X-to-Y transposition occurred about 3–4 million years ago, after the divergence of the human and chimpanzee lineages (Page et al.; Mumm et al. 1997; Schwartz et al. 1998). These sequences are 99% identical to DNA sequences in Xq21, notwithstanding they do not participate in X–Y crossing over during male meiosis.

- **X-degenerate sequences** seem to be surviving relics of the ancient autosomes from which the X and Y chromosomes co-evolved. These sequences are dotted with single-copy genes or pseudogenes displaying between 60%-96% nucleotide sequence identity to 27 different X-linked genes.

- **Ampliconic segments** are scattered across the euchromatic long arm and proximal short arm of the Y chromosome and display a combined extension of 10.2 Mb. These sequences are composed of repeated DNA blocks of 115-678 Kb, called amplicons, which are organized in six families colour-coded as yellow, blue, green, red, grey and turquoise, characterized by > 99.9% sequence identity among family members. The colour defining each family corresponds to the colour of the fluorescent probe (FISH probe) used for its identification (Kuroda-Kawaguchi et al. 2001). Amplicons, which can be regarded as SDs, are in turn organized in symmetrical arrays of contiguous units named "palindromes". Eight massive palindromes can be identified in the ampliconic region of the Yq, each defined by a symmetry axis separating two largely identical arms (with a sequence identity of 99.94–99.997%) constituted by single or multiple amplicons (Kuroda-Kawaguchi et al. 2001; Skaletsky et al. 2003) (Fig. 12). Amplicons represent approximately 35% of the MSY region and the eight palindromes collectively comprise one quarter of the MSY euchromatin. Therefore, the Y chromosome displays a significantly higher SDs content compared to the rest of chromosomes showing an average content of approximately 5%.

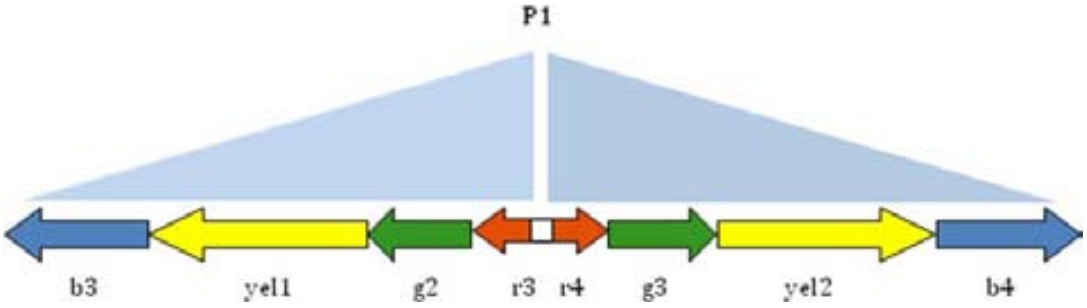


Figure 12. Example of organization of the amplicons (coloured arrows) in a symmetrical array of continuous repeat units (palindrome P1)

The ampliconic sequences evolved from a great variety of genomic sources, and accumulated in the MSY region through two main molecular mechanisms: the amplification of X-degenerate genes and the transposition/retroposition and subsequent amplification of autosomal genes. Such an accumulation of SDs (amplicons) in the MSY region is considered an evolutionarily conserved strategy of the Y chromosome to counteract the accumulation of deleterious mutations, in the absence of conventional recombination with a chromosome partner. The presence of massive near-identical amplicons allows indeed two recombination-based DNA repair mechanisms to occur in the MSY region: gene conversion and NAHR. The first is a non-reciprocal transfer of sequence information from one DNA duplex to another (Szostak et al. 1983), which can occur between duplicated sequences on a single chromosome and in mitosis(Jackson and Fink 1981). Gene conversion (non-reciprocal recombination) is as frequent in the MSY, as crossing over (reciprocal recombination) is in ordinary chromosomes and occurs routinely in 30% of the MSY (Skaletsky et al. 2003). This conversion-based system of gene copy "correction" permits the preservation of Y-linked genes from the gradual accumulation of deleterious mutations ensuring their continuity over time. As stated above, NAHR is a homology-based mechanism of accurate DNA repair, which can also lead to the generation of large-scale *AZF* structural rearrangements such as inversions and CNVs affecting the dosage of a number of Y-linked genes.

The MSY gene content

The early 20th century predominant view that the Y chromosome was a genetic wasteland was denied by the identification of at least 156 transcription units all located in euchromatic sequences of the MSY region and distributed among the three euchromatic sequence classes (Fig. 13). Approximately 78 out of 156 transcription units identified, encode for at least 27 distinct proteins or protein families (Table 1). For the remaining 78 transcription units strong evidence of protein coding is lacking and many of these transcription units are probably non-coding. The ampliconic sequences exhibit by far the highest density of genes, both coding and non-coding, among the three sequence classes in the MSY euchromatin: 135 of the 156 MSY transcription units identified so far are ampliconic.

Nine distinct MSY-specific protein-coding gene families have been identified, with copy numbers ranging from two (*VCY, XKRY, HSFY, PRY*) to three (*BPY2*) to four (*CDY, DAZ*) to six (*RBMY*) to an average of 35 (*TSPY*). Overall, these nine coding multi-copy gene families encompass roughly 60 transcription units and are predominantly or exclusively expressed in the testis. Furthermore, the ampliconic sequences include at least 75 other transcription units for which strong evidence of protein coding is lacking. Of these 75 putative non-coding transcription units, 65 are members of 15 MSY-specific families, and the remaining 10 occur in single copy. The X-transposed segments exhibit the lowest density of genes since only two single copy genes have been identified therein: the TGFB-induced factor homeobox 2-like, Y-

linked (*TGIF2LY*) expressed in testis and Protocadherin 11 Y-linked (*PCDH11Y*) expressed in brain. X-degenerated sequences include 16 single copy genes, among them all the 12 ubiquitously expressed MSY genes and only one gene, *SRY* (Sex Determining region Y), with testis restricted expression belongs to this class of sequences.

PAR1 and PAR2

The MSY is flanked on both sides by pseudoautosomal regions (PAR1 and PAR2), where X–Y crossing over is a normal and frequent event in male meiosis. These regions include 22 transcriptional units coding for 18 proteins.

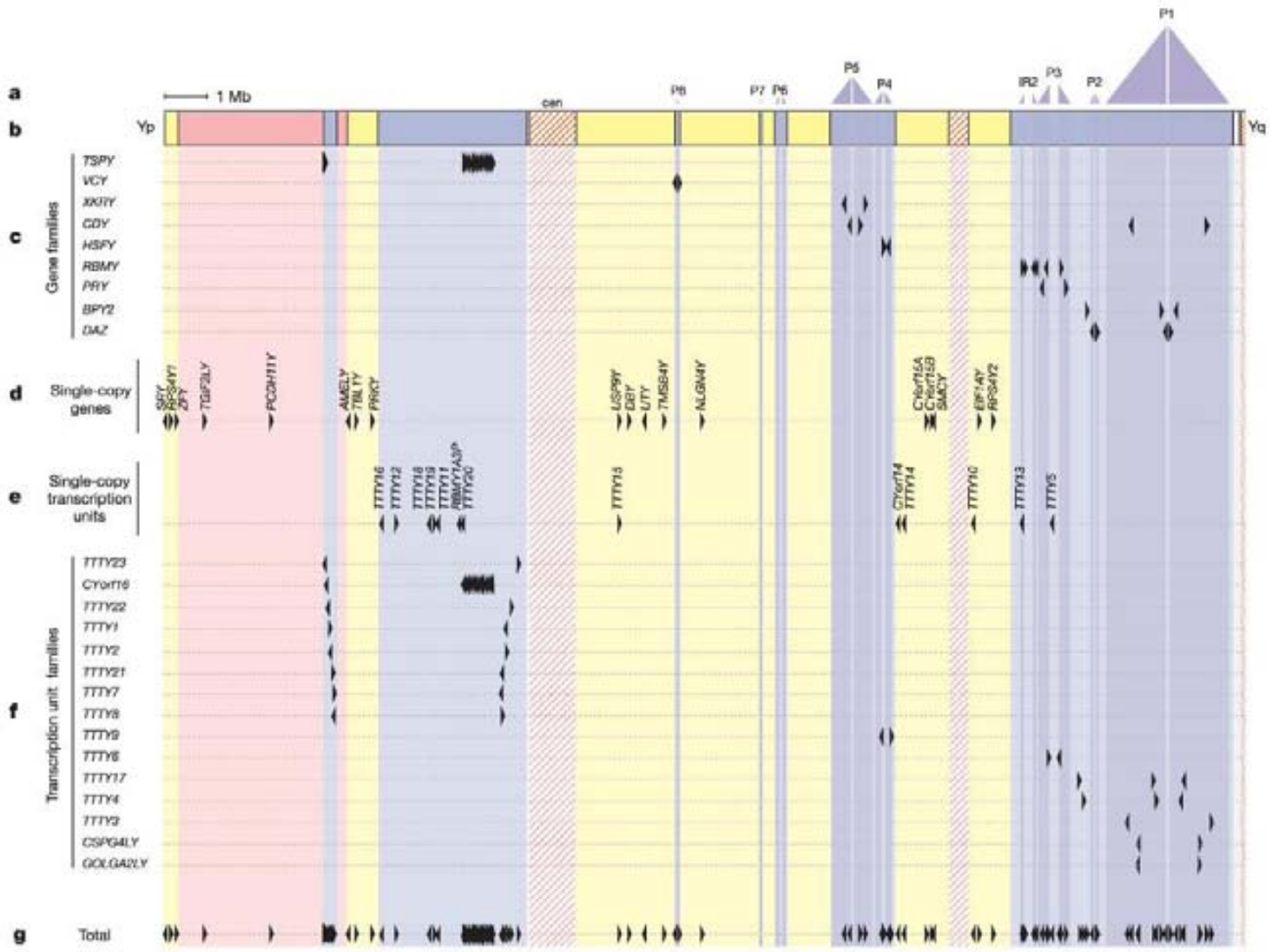


Figure 13. MSY genes, transcription units and palindromes. a) Localization of the 8 palindromes (P1-P8) in the MSY region. b) MSY region as represented in Figure 11. Solid black triangles denote coding genes and transcription units which are classified as follow: c) Nine families of protein-coding genes; d) single-copy protein coding genes and e) single-copy transcription units which give rise to spliced but apparently non-coding transcripts. f) Fifteen families of transcription units. g) Merged map of all genes and transcription units. Figure by *Skaletsky et al 2003.*

Table 1. MSY genes and gene families demonstrated or predicted to encode proteins. From *Skaletsky et al 2003.*

MSY sequence class	Gene Symbol	Gene Name	Number of Copies	Tissue expression	X-linked homologue	Autosomal homologue
X-transposed	*TGIF2LY*	TGF β-induced factor homeobox 2-like Y-linked	1	Testis	*TGIF2LX*	_
	PCDH11Y	Protocadherin 11 Y-linked	1	Fetal brain, brain	*PCDH11X*	_
Total			**2**			
X-degenerate	*SRY*	Sex determining regionY	1	Predominantly Testis	*SOX3*	_
	RPS4Y1	Ribosomal protein S4, Y-linked 1	1	Ubiquitous	*RPS4X*	_
	ZFY	Zinc finger protein Y-linked	1	Ubiquitous	*ZFX*	_
	AMELY	Amelogenin Y	1	Teeth	*AMELY*	_
	TBL1Y	Transducin β-like 1Y-linked	1	Fetal brain, prostate	*TBL1X*	_
	PRKY	Protein kinase Y-linked	1	Ubiquitous	*PRKX*	_
	USP9Y	Ubiquitin-specific peptidase 9 Y-linked	1	Ubiquitous	*USP9X*	_
	DBY	Dead box Y	1	Ubiquitous	*DBX*	_
	UTY	Ubiquitously transcribed tetratricopeptide repeat containing, Y-linked	1	Ubiquitous	*UTX*	_
	TMSB4Y	Thymosin β-4 Y-linked	1	Ubiquitous	*TMSB4X*	_
	NLGN4Y	Neuroligin 4 Y-linked	1	Fetal brain, brain, prostate, testis	*NLGN4X*	_
	CYorf15A	Chromosome Y open reading frame 15A	1	Ubiquitous	*CYorf15*	_
	CYorf15B	Chromosome Y open reading frame 15B	1	Ubiquitous	*CYorf15*	_
	SMCY	SMC (mouse) homologue, Y	1	Ubiquitous	*SMCX*	_
	EIF1AY	Eukaryotic translation initiation factor 1A Y	1	Ubiquitous	*EIF1AX*	_
	RPS4Y2	Ribosomal protein S4 Y-linked 2	1	Ubiquitous	*RPS4X*	_
Total			**16**			

Table 1. Continue

MSY sequence class	Gene Symbol	Gene Name	Number of Copies	Tissue expression	X-linked homologue	Autosomal homologue
Ampliconic	*TSPY*	Testis-specific protein Y-linked	~35	Testis	_	_
	VCY	Variable charge Y-linked	2	Testis	*VCX*	_
	XKRY	XK related Y	2	Testis	_	_
	CDY	Chromodomain Y	4	Testis	_	*CDYL*
	HSFY	Heat shoc transcription factor Y-linked	2	Testis	_	_
	RBMY	RNA-binding motif Y	6	Testis	*RBMX*	_
	PRY	PTP-BL related Y	2	Testis	_	_
	BPY2	Basic protein Y 2	3	Testis	_	_
	DAZ	Deleted in azoospermia	4	Testis	_	*DAZL*
Total			~60			
Grand Total			~78			

Y chromosome haplogroups

Because of its sex-determining role, the Y chromosome is male specific and constitutively haploid. As stated above, it escapes meiotic recombination for up to 95% of its length and thus it is clonally transmitted from father to son. The importance of this feature is that the Y chromosome haplotypes, which are the combinations of allelic states of markers along the chromosome, usually pass intact from generation to generation changing only by spontaneous mutations (Jobling and Tyler-Smith 2003). Therefore, Y chromosome represents an invaluable record of all mutations that have occurred along male lineages throughout evolution. Although all existing Y chromosomes share a single evolutionary history, deriving from the same ancestral Y chromosome, the presence of polymorphisms in non coding regions allowed the definition of monophyletic groups. The Y phylogeny is indeed based on binary markers (mostly SNPs) that have low mutation rates, and therefore can be regarded largely as unique events in human history. Haplotypes constructed using such markers are defined Y haplogroups (hgrs) and are arranged in the phylogenic tree of the Y chromosome. In the last decade many efforts have been focused on the identification of novel binary markers for the construction of a more and more detailed haplogroup tree (Y chromosome consortium 2002) (Jobling and Tyler-Smith 2003). The most recently published version of the Y phylogeny is based on 599 binary markers defining 311 distinct haplogroups (Karafet et al. 2008)(Fig.14).

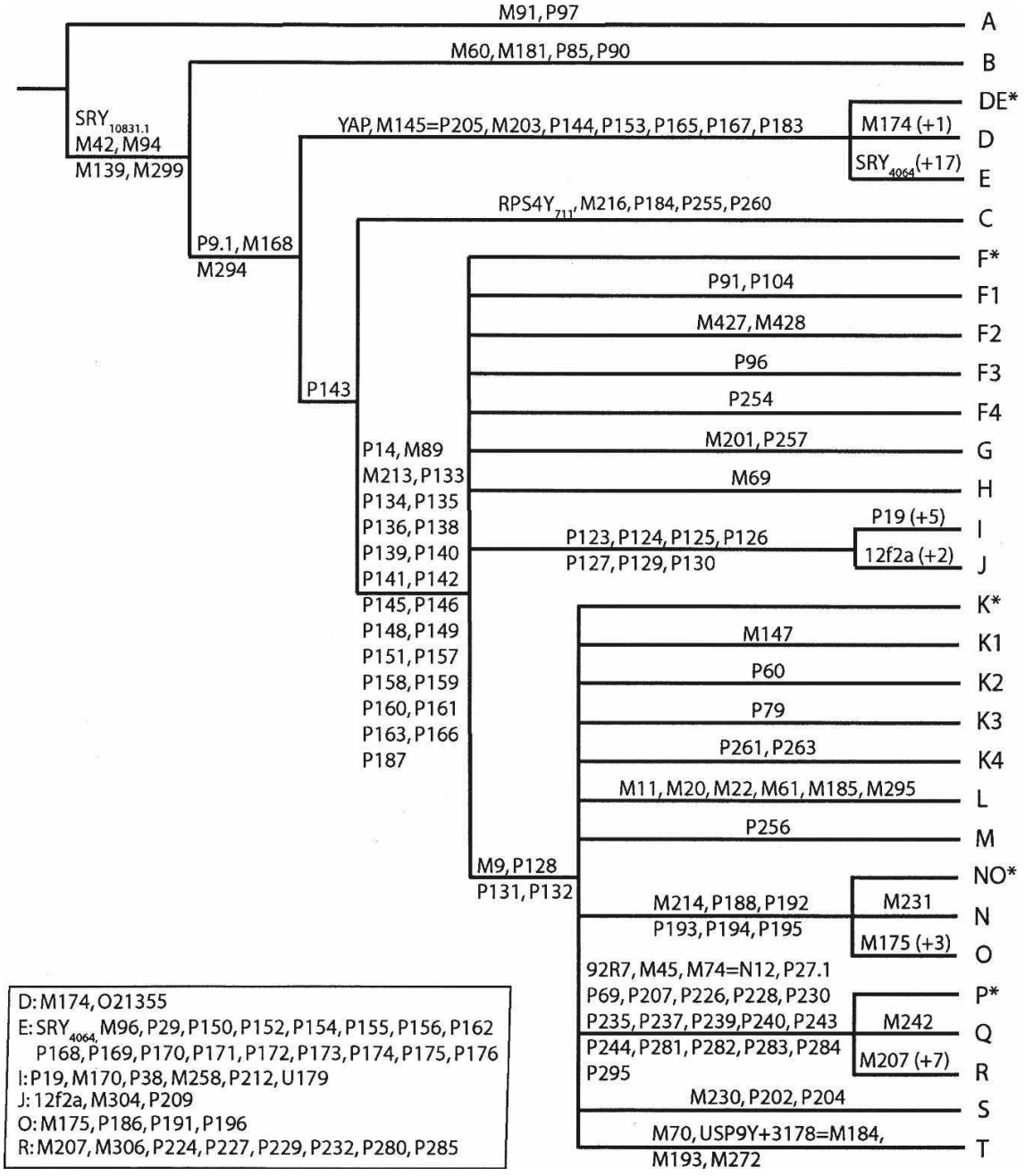


Figure 14. An abbreviated form of the Y chromosome parsimony tree showing the most relevant Y hgrs. Mutation names are indicated on the branches and haplogroups are indicated on the right at the end of each branch.

Survey of the population distributions of Y-chromosomal hgrs have shown that they are highly geographically differentiated (Underhill and Kivisild 2007) with particular populations carrying their own characteristic sets of lineage (Jobling 2008) (Fig.15).

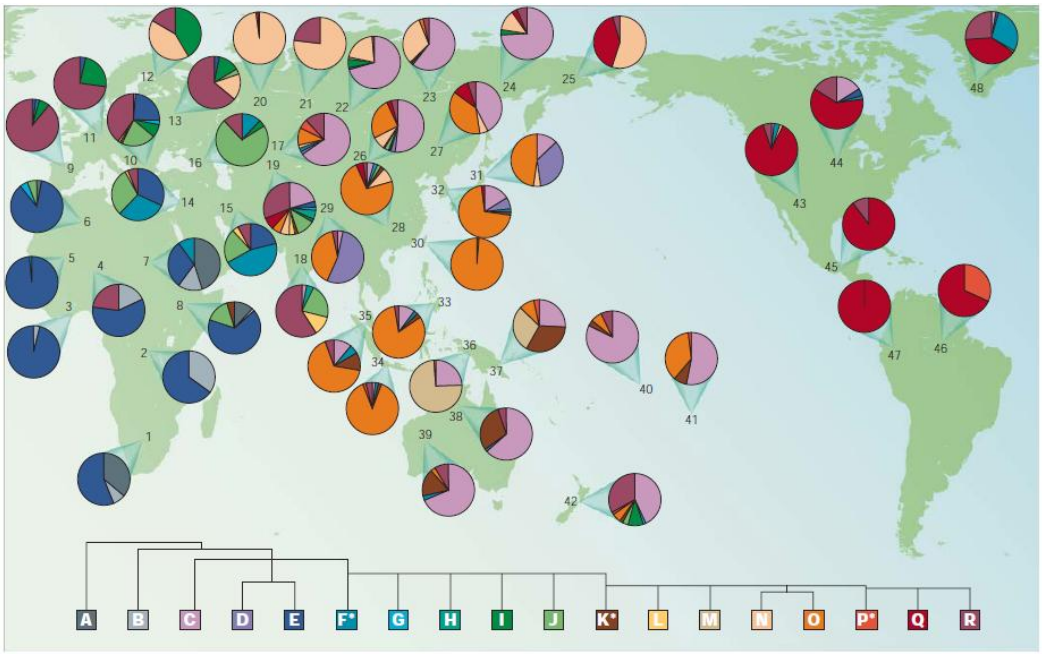


Figure 15. Global distribution of Y hgrs in the world. Each circle represents a population sample with the frequency of the 18 main Y hgrs identified by the Y Chromosome Consortium (YCC) indicated by the colored sectors. It is worth noting the general similarities between neighbouring populations but large differences between different parts of the world. Figure by *Jobling and Tyler-Smith 2003*.

It has been suggested that Y background may modulate the susceptibility to some genetic rearrangements and thus contribute to the inter-population variation of their frequency. Jobling and colleagues (Jobling et al. 1998), provide for the first time findings supporting this hypothesis showing that the recombination between the homologous genes protein kinase X-linked (*PRKX*) on the short arm of the X chromosome (Xp) and protein kinase Y-linked (*PRKY*) on Yp arise predominantly on a particular Y hgr. Furthermore, in case of NAHR-based CNVs, Y background-related variations in DNA sequence or orientation of segmental duplications may increase or decrease the susceptibility to CNV formation via NAHR mechanisms. Contrary to a previous study by Peracchini and collegues (Paracchini et al. 2000) recent surveys on Chinese and Moroccan populations reported an association between some Y hgrs and increased susceptibility to (partial) deletion of *AZFc* region (Imken et al. 2007; Yang et al. 2008a; Yang et al. 2008b; Eloualid et al. 2012). It has also been reported that the Y background may influence the phenotypic expression of certain polymorphisms/mutations due to the linkage between the latter and modulating Y-linked genetic factors. The "*gr/gr* deletion" of *AZFc* region represents a classic example of this phenomenon. In Caucasian Y hgrs, this partial *AZFc* deletion represents a proven genetic risk factor for impaired spermatogenesis (Tüttelmann et al. 2007; Visser et al. 2009; Navarro-Costa et al. 2010; Stouffs et al. 2011; Rozen et al. 2012), whereas *gr/gr* deletion does not have any effect on sperm production in specific Y hgrs commonly represented in Asian populations, such as D2b, Q3 and Q1, where it occurs constitutively (Sin et al. 2010; Yang et al. 2010). However, when non-constitutively *gr/gr* deleted Y hgrs are considered, a significant association with spermatogenic impairment is observed also in Asian populations. Thus, the inclusion of deletion-fixed Y hgrs in the study population can mask the association between such *AZFc* partial deletion and spermatogenic impairment. All these observations indicate that the correct selection of the study groups based on the Y background (ethnic/geographic matching of patients and controls) is pivotal in

association studies involving the Y chromosome for preventing population stratification biases.

Y-linked CNVs

As mentioned above, the accumulation of a high proportion of segmental duplications in the MSY region, provides the structural basis for the generation of CNVs (Skaletsky et al. 2003; Jobling 2008) by promoting the activation of NAHR events .

Given the clonal inheritance of MSY, a phylogenetic approach can be used to provide insights into the dynamic of Y-linked CNV formation. Specifically, by determining the frequency of a given CNV in different Y lineages, it is possible to infer the minimum number of independent mutation events accounting for the CNV distribution. In the case of unique CNVs, present in all the members of a given Y hgr but absent in other lineages (CNV1 in the Fig.16), a unique mutation event have occurred in the ancestral Y chromosome of that specific hgr. Recurrent CNVs (CNV2 and CNV3 in the Fig.16), distributed among different branches may arise through several independent mutation events reflecting the highly mutagenic nature of the involved region. In the case of recurrent CNVs showing high prevalence in Y hgrs (CNV2), belonging to more than one lineage indicates that the mutation is likely to be occurred in the ancestral Y chromosome of more than one lineage but in some members of the same hgr "reversion" of the mutation has occurred.

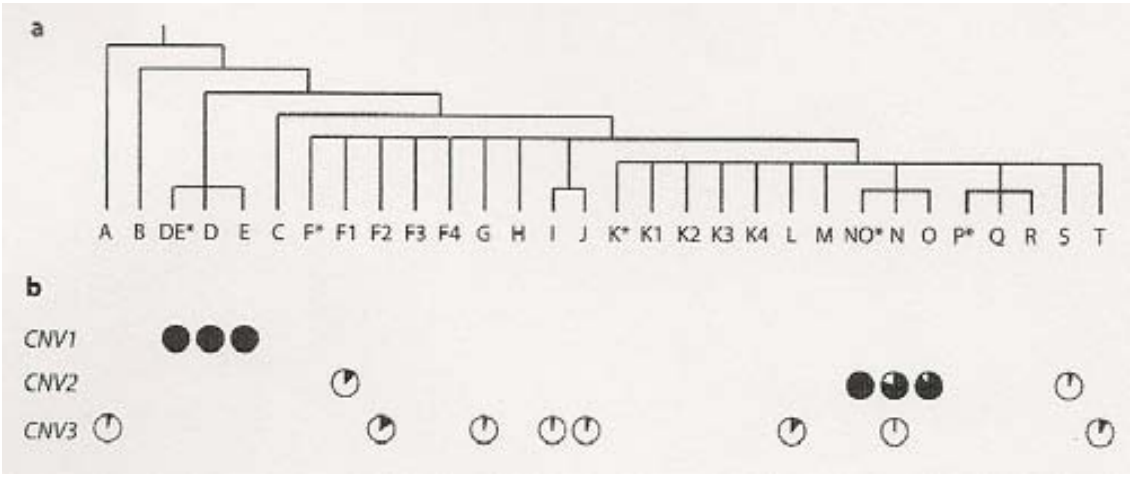


Figure 16. Phylogenic approach used for the study of the dynamic of Y-linked CNV formation. Figure by *Joblin et al 2008*.

Finally, the rate of CNV generation could be deduced if the number of generations encompassed by the sampled chromosomes was known (Hammer MF 2002; Repping et al. 2006; Karafet et al. 2008).

The discovery of Y-CNVs has arisen from several research field such as forensic and population genetics studies and molecular male reproductive genetics. However a more comprehensive picture of Y CNVs derives from systematic genome-wide CNV surveys (Redon et al. 2006; Perry et al. 2008) and whole Y chromosome resequencing data (Levy et al. 2007) has provided a more objective picture of Y-CNVs. The largest scale study (Redon et al. 2006) performed so far explored 104 distinct Y chromosomes from the HapMap sample, revealing that the *AZFc* region corresponds to the most variable euchromatic portion in terms of CNVs (Fig. 17)

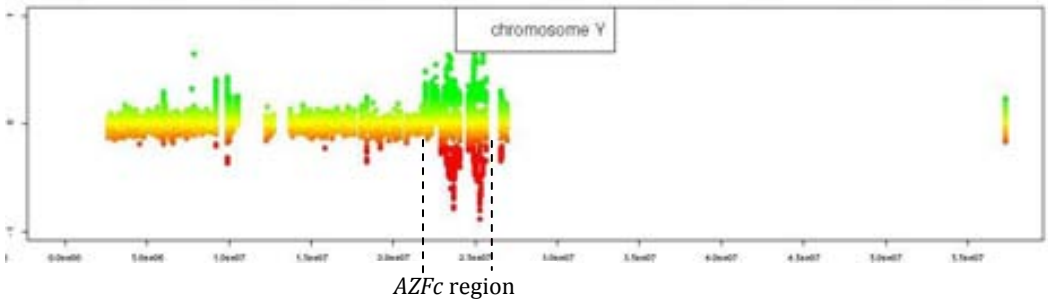


Figure 17. Representation of the log2 ratio from comparative genomic hybridization to BAC clones spanning the Y euchromatin. The most dynamic region correspond to the *AZFc* region

A detailed description of the most common Y-linked CNVs is presented below.

Y chromosome microdeletions: the AZF deletions

The first association between azoospermia and microscopically detectable deletions in the long arm of the Y chromosome (Yq), was reported by Tiepolo and Zuffardi in 1976 (Tiepolo and Zuffardi 1976). The authors proposed the existence of an AZoospermia Factor (*AZF*) on Yq, representing a key genetic determinant for spermatogenesis, since its deletion was associated with the lack of spermatozoa in the ejaculate. Due to the structural complexity of the Y chromosome, the molecular characterization of the *AZF* took about 30 years to be achieved. With the development of molecular genetic tools and the identification of specific markers on the Y chromosome (Sequence Tagged Sites, STSs), it was possible to circumscribe the *AZF* region and to highlight its tripartite organization. Three *AZF* sub-regions were indeed identified in proximal, middle and distal Yq11 and designated as *AZFa*, *AZFb* and *AZFc*, respectively. It was then demonstrated that *AZFb* and *AZFc* overlap, being 1.5 Mb of the distal portion of *AZFb* interval part of the *AZFc* region (Fig. 18).

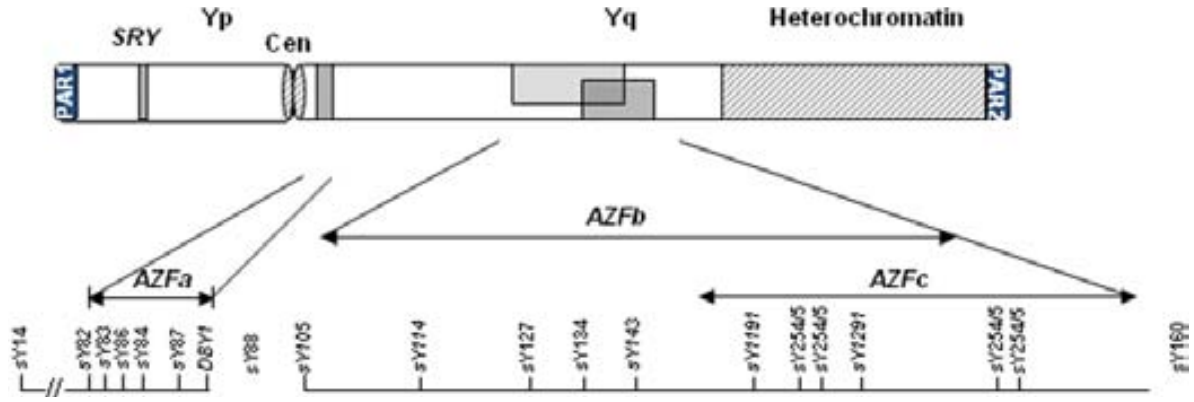


Figure 18. Structure of *AZF* region of the Yq. Schematic representation of the Y chromosome showing the three *AZF* regions (A) with each specific STSs (B).

Sub-microscopic deletions involving the *AZF* regions, regarded to as Y microdeletions, occur in 1/4000 males in the general population. They are now considered a well-established genetic cause of male infertility being exclusively found in men with impaired sperm

production. *AZF* deletion screening is now part of the diagnostic work-up of severe male factor infertility (Simoni et al. 2008). Indications for *AZF* deletions screening are based on sperm count and include azoospermia and severe oligozoospermia (<5 million spermatozoa/ml). Thanks to the European Academy of Andrology (EAA) guidelines and EAA/EMQN (European Molecular Genetics Quality Network) external quality control programme (http://www.emqn.org/emqn/) Yq testing has become more homogeneous and reliable in different routine genetic laboratories. The EAA guidelines provide a set of primers (two markers for each region) that is able to detect virtually all clinically relevant deletions.

Deletion frequency increases with the severity of spermatogenic impairment amounting to 5-10% among non-obstructive azoospermic men, 2-4% among severe oligozoospermic men with less than 5 million spermatozoa/ml and less than 0.5% among men with sperm concentration between 5 and 10 million spermatozoa/ml. Y-microdeletions have never been found among men with normal sperm parameters (normozoospermic men). In addition to the semen phenotype, also ethnic background is likely to influence the occurrence of this genetic anomaly as suggested by the different deletion frequency observed even within similar semen categories amongst infertile men from different populations. In this regard, the lowest deletion frequency (1.8%) was reported in German and Danish idiopathic severely oligozoospermic men,(Cruger et al. 2003; Simoni et al. 2008) whereas the highest in an ethnically admixed population from France (13.7%)(Krausz et al. 1999) and in Romanians (10%) (Raicu et al. 2003). Data on the prevalence of classical AZF deletions in men attending an infertility clinic in Spain derive from two independent surveys, with an overall frequency of 5.4% and 7%, respectively (Oliva et al. 1998; Martínez et al. 2000).
In most of the cases Y microdeletions arise as de novo event and are associated with heterogeneous seminal and testicular phenotypes.

Types of AZF deletions: mechanism of formation, frequency and genotype-phenotype correlation.

Y microdeletions arise through NAHR and can be classified according to the recombination hot spot as *AZFa*, *P5*-proximal *P1* (*AZFb*), *P5*-distal *P1* (*AZFbc*), *P4*-distal *P1* (*AZFbc*) and *b2/b4* (*AZFc*) deletions.

The *AZFa* region spans 792 Kb and unlike either *AZFb* or *AZFc*, is exclusively constituted by single-copy DNA (Figure). The complete deletion of *AZFa* interval results from the NAHR between two flanking HERV elements (human endogenous retroviral elements) spanning 10 Kb each and displaying an overall 94% of sequence identity. Two ubiquitously expressed genes map inside *AZFa* region and are thus involved in deletion:

- The ***USP9Y*** gene (ubiquitin specific peptidase 9, Y-linked) encodes for an ubiquitin specific protease member of the C19 cystein peptidase family. These enzymes promote the intracellular cleavage of ubiquitin molecules from ubiquitinated proteins (Baarends et al. 2000; Lee et al. 2003). USP9Y shares 91% identity with its X homologue (USP9X) suggesting that both target similar molecules and may overlap functionally (Vogt et al. 2008). Studies in murine gametogenesis have shown that while *USP9X* expression starts as early as in the establishment of the primordial germ cell population in both sexes, *USP9Y* only starts to be expressed in the male germline at the spermatid stage (Noma et al. 2002). This markedly distinct expression window hints at a temporal constrain in the regulation of USP9Y function, a probable consequence of its molecular targets only being present at later spermatogenic stages. The available data in human points to *USP9Y* not being essential for male fertility, as observed in other

primate lineages where the gene became inactive (Perry et al. 2007; Tyler-Smith 2008). Although *USP9Y* deletions were initially thought to be exclusively associated to azoospermia and to account for the *AZFa* deletion phenotype, deletions compatible with sperm production and with natural conception have been later identified, the latter corresponding to the complete deletion of the gene (Krausz et al. 2006; Luddi et al. 2009). Based on th above observations, *USP9Y* is currently considered as a fine tuner of male gametogenesis (Krausz et al. 2006; Tyler-Smith and Krausz 2009).

- ***DDX3Y*** (DEAD (Asp-Glu-Ala- Asp) box helicase 3, Y-linked) The DDX3Y protein is an ATP-dependent RNA helicase belonging to the DEAD box protein family characterized by the conserved motif Asp-Glu-Ala-Asp. The exact molecular role of DDX3Y is unknown, although the DEAD box proteins have been implicated in several key processes of RNA metabolism such as secondary structure alteration, splicing, spliceosome assembly and translation initiation (Rosner and Rinkevich 2007). *DDX3Y* and its X homologue (*DDX3X*; 91.7% sequence identity) are ubiquitously expressed, with expression levels peaking in testis (Ditton et al. 2004; Sekiguchi et al. 2004). However, the widespread presence of both transcripts in adult tissues does not directly correlate with actual protein expression since DDX3Y, unlike DDX3X, is testis-specific. Both genes encode for testis-specific transcripts characterized by an overall shorter length and the presence of extended untranslated regions (UTRs) (Foresta et al. 2000; Vogt et al. 2008). This specificity in transcriptional profiles most likely serves to ensure a precise expression window as DDX3Y is detected predominantly in the cytoplasm of spermatogonia whereas DDX3X is mainly detected in spermatids (Ditton et al. 2004). The divergent expression window of the two genes suggests that DDX3Y may represent a specialization of DDX3X functions for premeiotic developmental stages. Yet, despite differences in cell type expression, the molecular functions of DDX3Y and DDX3X are probably analogous, as evidenced by functional rescue studies in murine cell lines (Sekiguchi et al. 2004). Taking into account data from *USP9Y* deletions as well as the regulation of *DDX3Y* function, it is tempting to consider that the absence of the latter is the main causative agent for the complete *AZFa* deletion phenotype. However, this hypothesis still requires the validation warranted by the unambiguous identification of *DDX3Y*-specific deletions.

AZFa deletion is a fairly rare event representing less than 5% of the reported Y microdeletions (Kamp et al. 2001; Krausz and Degl'Innocenti 2006) and it is invariably associated with azoospermia due to the complete absence of germinal cells in seminiferous tubules, a condition known as pure Sertoli Cells Only Syndrome (SCOS) (Kleiman et al. 2012). The low prevalence most likely stems from both limitations of the deletion mechanism (such as the relatively short recombination target), and from the potential negative selection of deletion due to its deleterious effect on fertility. The corresponding NAHR product, the *AZFa* duplication, is detected at a fourfold higher frequency when compared to that of the deletion indicating that increased *AZFa* gene dosage does not affect fertility (Bosch and Jobling 2003).

The *AZFb* region spans a total of 6.23 Mb and contains three single-copy regions, a *DYZ19* satellite repeat array and 14 ampliconic elements organized in palindromes (from P2 to P5 and the proximal part of P1) of which P5/P4 and P1 are the NAHR targets giving rise to the complete and partial *AZFb* deletion, respectively. The *AZFb* gene content reflects the mesh of different sequence types constituting this region, with single copy genes mapping alongside ampliconic gene families (Fig. 19).

More specifically, a total of 5 different single-copy transcription units map to *AZFb*:

- ***KDM5D*** (lysine (K)-specific demethylase 5D) encodes for a histone H3 lysine 4 (H3K4) demethylase that forms a protein complex with the MutS homolog 5 (MSH5) DNA repair factor during spermatogenesis (Lee et al. 2007b; Akimoto et al. 2008). This complex locates to condensed DNA during the leptotene/zygotene stage, suggesting an involvement in male germ cell chromatin remodelling. In accordance, by demethylating di- and tri-methylated H3K4, KDM5D may be involved in chromosome condensation during meiosis. Such possibility fits with the instances of maturation arrest at the spermatocyte stage associated to *AZFb* deletions. Despite the apparently male germline-specific functions, this gene is ubiquitously expressed and is homologous to *KDM5C*, an X-borne gene associated to X-linked mental retardation (Santos et al. 2006; Tahiliani et al. 2007).

- ***EIF1AY*** (eukaryotic translation initiation factor 1A, Y-linked) encodes for a ubiquitously expressed Y-linked member of the EIF-1A family, a sequence family involved in translation initiation (Roll-Mecak et al. 2001). The EIF-1A proteins are required for a high rate of protein biosynthesis since they enhance ribosome dissociation into subunits and stabilize the binding of the 43S complex (a 40S subunit, eIF2/GTP/Met-tRNAi and eIF3) to the 5'_end of capped RNA. *EIF1AY* has an X-homologue (*EIF1AX*) and although evidence at the protein level is available, its functions are largely deduced by similarity to *EIF1AX*. In this regard, the acquisition of male-specific regulatory features by *EIF1AY* and/or the existence of partial functional overlap with *EIF1AX* are valid hypotheses.

- ***RPS4Y2*** (ribosomal protein S4, Y-linked 2), corresponds to a fairly recent duplication of the *RPS4Y* gene, the latter encoding for a ribosomal protein subunit required for mRNA binding to the ribosome (Bergen et al. 1998). Since *RPS4Y2* expression is testis-specific, a putative role in the post-transcriptional regulation of the spermatogenic program can be postulated (Andrés et al. 2008). Indeed, evidence for positive selection in the *RPS4Y2* coding sequence suggests a hypothetical acquisition of germline-specific functions.

- ***CYorf15A*** (chromosome Y open reading frame 15A) and ***CYorf15B*** (chromosome Y open reading frame 15B) sequences have an X homologue (*CXorf15*) that belongs to the taxilin family and has been linked to transcriptional regulation in osteoblasts (Yu et al. 2006). The role of *CYorf15* sequences for either general or reproductive functions is unknown. Although *CYorf15A* and *CYorf15B* apparently encode for proteins homologous to the amino and carboxy-terminal domains of CXorf15, respectively, evidence for their existence is restricted to the identification of ubiquitously espressed transcripts (Skaletsky et al. 2003)

AZFb multicopy genes belong to 7 gene families:

- ***XKRY*** (XK, Kell blood group complex subunit-related, Y-linked) is expressed specifically in testis and maps to the yellow-coded amplicon family. Although protein evidence is still lacking, sequence analysis suggests that the two active copies of *XKRY* (mapping to yel3 and yel4) encode for a multipass transmembrane tran sport protein similar to the XK protein. The latter locates to neuromuscular and hematopoietic cell membranes

- ***HSFY*** (heat shock transcription factor, Y-linked) maps to the blue amplicons of palindrome P4, with two active copies (*HSFY1/2*) located in *b5* and *b6,* respectively. The protein coded by *HSFY* is a member of the heat shock factor family of transcriptional activators. This gene is subjected to alternative splicing, generating three different mRNA transcripts with testis-restricted expression. More specifically, they have been identified in spermatogenic cells up to the spermatid stage and in Sertoli cells (Shinka et al. 2004). The product of mRNA transcript variant 1 contains a heat-shock factor-like DNA-binding domain (Tessari et al. 2004) suggesting that this mRNA is the critical HSFY transcript. Evidence exists suggesting that HSFY may play a critical role in germ cell maturation. For instance, a decreased expression of HSFY protein has been associated with maturation arrest in American and Japanese populations (Sato et al. 2006) and expression analysis of the mouse orthologue demonstrated predominant expression in round spermatids, supporting a role for *HSFY* in the later stages of spermatogenesis (Kinoshita et al. 2006). Interestingly, three oligozoospermic patients with an identical partial *AZFb* deletion (removing the proximal *AZFb* markers from palindrome P4: sY117, a marker with one copy in each *HSFY* and sY114, a marker flanking the central P4 spacer) leading to isolated loss of *HSFY* have recently been reported (Kichine et al. 2012). This deletion was detected in 3/1186 infertile men and 0/1179 control men from France, confirming that *HSFY* deletion is associated with infertility, but contradicting the findings that *HSFY* expression is essential for germ cell development. One plausible explanation for sich inconsistence is that X-linked or autosomal compensatory mechanisms may exist in some populations that can partially compensate for absence of *HSFY*. Indeed, all three of the reported oligozoospermic men with HSFY deletions belonged to the same specific Y-chromosome haplotype (R1b1b1a1b).

- ***RBMY1A1*** (RNA binding motif protein, Y-linked, family 1, member A1) is present in multiple copies along the Y chromosome; nevertheless, the six functional units cluster to the *AZFb* amplicons (Elliott et al. 1997; Chai et al. 1998). This complex arrangement, characterized by an extensive array of *RBMY1A1* pseudogenes and sub-families, had thwarted initial attempts to precisely map this determinant (Schempp et al. 1995; Najmabadi et al. 1996). *RBMY1A1* is part of the RNA binding motif (RBM) gene family that also includes an X homologue (*RBMX*) and a set of autosomal retrogene-derived copies of *RBMX* (of these only *RBMXL1*, *RBMXL2* and *RBMXL9* are expressed, and protein evidence is only available for *RBMXL2*) (Elliott 2004). Unlike its ubiquitously expressed X homologue, *RBMY1A1* is expressed solely in male germ cells, with the protein displaying a nuclear location (Elliott et al. 1997). The main feature of the RBM family is the presence of a N-terminal RNA recognition motif (RRM) responsible for the interaction with target RNA molecules (Shamoo et al. 1995; Chai et al. 1998). In this regard, RBM family members display characteristics of canonical RNA-binding proteins involved in nuclear RNA processing. In fact, this gene has been linked to the storage and transport of mRNA from the nucleus during spermatogenesis. Differently from the other RBM genes, *RBMY1A1* also contains a C-terminal protein interaction repeat domain enriched in serine, arginine, glycine, and tyrosine (SRGY) (Elliott 2004). This serves as a regulatory region for the modulation of RBMY1A1 function. Efforts to identify RBMY1A1-interacting proteins have shown that pre-mRNA splicing regulators, particularly the SR and the SR-related proteins, are potential partners (Elliott et al. 2000; Venables et al. 2000). These ubiquitously expressed factors also contain RRM domains, therefore their functional modulation via RBMY1A1 interaction emerges as a

distinct possibility (Longman et al. 2000). Additionally, RBMY1A1 may modulate cellular processes other than splicing regulation and mRNA metabolism since it has been shown to interact with the steroidogenic acute regulatory protein (STAR) and T-STAR proteins (Venables et al. 1999). These act not only as splicing regulators but also as members of signal transduction pathways involved in cell cycle control. In this regard, RBMY1A1 can be involved in several aspects of meiotic and premeiotic regulation via the establishment of multiple protein complexes. Interestingly, the male germ cell-specific expression of *RBMY1A1* is also mimicked by the autosomal *RBMXL2* gene. In this case, the nuclear localization of the protein during and immediately after meiosis is suggestive of meiotic specialization. In accordance, haploinsufficiency of the murine *RBMXL2* orthologue results in abnormal spermatogenesis in animal models (Ehrmann et al. 2008). The identification of the RNA targets of RBMY1A1 has been partially successful. It is believed that the RRM domain can bind RNA at both high and low affinity, making the characterization of target molecules complex (Elliott 2004). Furthermore, the protein has an unique two-step mechanism for RNA recognition that starts with a sequencespecific interaction with the target molecule before eliciting a conformational modification (Skrisovska et al. 2007). This complex mechanism warrants RBMY1A1 a significant plasticity in terms of RNA partners. Studies in murine models have identified 12 different potential mRNA targets for RBMY1A1, most of them expressed in testis starting from the neonatal period (Zeng et al. 2008). Interestingly, the protein seems to be able to bind to its own alternative transcript, suggesting a complex regulatory network. The existence of alternative *RBMY1A1* transcripts has also been detected in humans (Chai et al. 1997). All the aforementioned properties seem to indicate that the disruption of *RBMY1A1* plays a significant role in the *AZFb* deletion phenotype. In reality, both its expression pattern and putative role in male germ cell development support the notion that RBMY1A1 deletions perturb the meiotic program. Similarly, the disruption of *KDM5D* may also contribute to the deletion phenotype. In fact, RBMY1A1 and KDM5D are located in the germ cell nucleus during prophase I, suggesting involvement in meiotic orchestration. While this regulation may be directly exerted by KDM5D (via changes in chromatin structure), the role of RBMY1A1 might be mediated by effector proteins or by transcriptional regulation of mRNA targets.

- ***PRY*** (Testis-Specific PTP-BL-Related Protein on Y) gene copies map to the blue amplicons, with the two functional units being restricted to b1 and b2. These are designated as *PRY* and *PRY2*, respectively, and encode for a gene product with a low degree of similarity to protein tyrosine phosphatase, nonreceptor type 13. The latter corresponds to a signalling molecule involved in the regulation of a myriad of cellular processes, particularly in programmed cell death (Dromard et al. 2007). *PRY* and *PRY2* display testis-specific expression and additional regulation via alternative splicing (Stouffs et al. 2001). Nevertheless, the alternative transcript seems to correspond to a nonfunctional isoform since it contains a premature stop codon truncating the product at about half. The expression of *PRY* in germ cells is irregular, with the protein being detected only in a few sperm and spermatids (Stouffs et al. 2004). Interestingly, both transcript and protein levels were shown to be higher in the defective germ cell fraction of the ejaculate. Furthermore, PRY levels are increased in ejaculated sperm obtained from men with abnormal semen parameters, suggesting a link between its expression and defective spermatogenesis. Appropriately, a role for PRY in male germ cell apoptosis has been suggested based on the observation that approximately 40% of PRY positive cells show DNA fragmentation.

The remaining 3 multicopy gene families (*CDY*, *BPY2* and *DAZ*) overlap with the *AZFc* region genes and are described in details below.
The *AZFb* deletion occurs at a similar or slightly increased rate than *AZFa* deletion (3 to 10 % of all Yq microdeletions). *AZFb* deletion carriers are azoospermic with testicular histology of maturation arrest at the spermatocyte/spermatid stage. Unlike *AZFa* deletion, no evidence for reciprocal duplications has been reported for *AZFb* deletion, so far.

The *AZFc* deletion spans 3.5 Mb and results from the NAHR between the flanking *b2* and *b4* amplicons. The deletion removes 21 genes and transcriptional units belonging to 8 multicopy gene families (Fig.19). These include 3 protein coding gene families (*BPY2*, *CDY* and *DAZ*) specifically expressed in the testis:

- The ***BPY2*** *(*Basic Protein charge, Y-linked, 2) gene family maps to the green amplicons (one active copy per amplicon) and includes also a set of pseudogene sequences (Kuroda-Kawaguchi et al. 2001). The genomic organization of the *BPY2* gene is quite unique since it is constituted by nine exons only five of which translated into amino-acids (Stuppia et al. 2001). The *BPY2* gene product is a testis-specific highly charged protein displaying a nuclear localization throughout all male germ cell developmental stages, persisting even in ejaculated sperm (Tse et al. 2003). The exact role played by BPY2 in spermatogenesis is unclear, with most of the available knowledge being inferred from its protein partners. Using the yeast two-hybrid assay BPY2 has been shown to interact with ubiquitin protein ligase E3A (UBE3A), a widely-expressed member of the ubiquitin protein degradation system (Wong et al. 2002). Since UBE3A corresponds to a testis-expressed E3 ubiquitin protein ligase (responsible for the transfer of the ubiquitin group to the targeted substrates), BPY2 may modulate its target specificity. Additionally, the two-hybrid assays have also identified microtubule-associated protein 1S (MAP1S) as an interacting protein (Wong et al. 2004). MAP1S is a member of the microtubule-associated proteins (MAPs) family and is involved in microtubule binding, bundling and stabilization, as well as in the crosslinking of microtubules with microfilaments. Since MAP1S is predominantly expressed in testis, a putative role of BPY2/MAP1S in the control of the male germ cell cytoskeletal network has been proposed (Wong et al. 2004).

- The chromodomain protein gene family (***CDY***) consists of two Y-encoded genes (*CDY1* and *CDY2*) and two autosomal copies (*CDYL* in chromosome 6 and *CDYL2* in chromosome 16) (Lahn and Page 1999). The Y family members map to the yellow amplicons, with the *CDY1* copies in *AZFc* (amplicons *yel1* and *yel2*) and the *CDY2* copies in *AZFb* (*yel3* and *yel4*) (Ferlin et al. 2003a). A fairly large number of pseudogene sequences are also scattered throughout *AZFb* and *AZFc*. As expected, the Y-linked copies have testis-specific expression whereas the autosomal units display a more general expression pattern (*CDYL* is even ubiquitously expressed) (Lahn and Page 1997). These genes are involved in postmeiotic nuclear remodelling and transcriptional regulation. The CDY proteins are characterized by two functional motifs: a N-terminal chromatin-binding domain (the chromodomain) and a C-terminal catalytic domain responsible for the Coenzyme A (CoA)-dependent acetyltransferase activity. The chromodomain is a typical signature of proteins involved in chromatin remodelling and gene expression regulation (Tajul-Arifin et al. 2003). Accordingly, in vitro assays have demonstrated that recombinant CDY proteins can acetylate histone H4 (and, to a lesser degree, H2A). Given the nuclear localization of the protein and the

post-meiotic expression window, the CDY family is considered a nuclear remodeling factor promoting histone H4 hyperacetylation (Lahn et al. 2002). The latter, by inducing a more relaxed chromatin configuration, may serve as trigger for the histone-to-protamine transition and subsequent nuclear condensation. The function of the CDY proteins is not restricted to histone acetylation. Studies have associated CDYL and its paralogues to transcriptional corepressor complexes consisting of multiple chromatin modifying proteins (Caron et al. 2003; Shi et al. 2003; Mulligan et al. 2008). Accordingly, the primary function of CDYL may be that of a transcriptional co-repressor, as observed in murine models when histone deacetylases (HDACs) bind to its catalytic domain (Caron et al. 2003). The protein acquires its role in chromatin remodelling only when HDACs are degraded (in the elongating spermatid stage) and the CoA-binding activity of mCdyl is activated. This fits with data obtained from protein structure analysis indicating that the CDY proteins do not show obvious similarities to canonical histone acetyltransferase motifs (Wu et al. 2009). Recently, CDY1 has also been shown to interact with lysine 9-methylated histones (H3K9me2 and H3K9me3), although the exact functional role of this interaction is unknown (Kim et al. 2006). The analysis of such binding properties further suggests that *CDYL2*, not *CDYL*, is the ancestor of the gene family (Fischle et al. 2008). The CDY1 and CDY2 proteins are isoforms with an amino-acid identity of 98% and a similar expression window (Lahn and Page 1999). On the other hand, the global identity score between the Y-linked CDY proteins and CDYL is just 63%. The accelerated protein evolution rate of the Y-borne CDY sequences seems to suggest that these copies have evolved under positive selection for germline specific functions (Dorus et al. 2003). Nevertheless, identity levels are slightly higher when comparing just the functional domains of the Y-derived and autosomal copies. In this regard, functional complementation between *CDY* genes may rescue, to some extent, the loss of the *AZFb* and/or *AZFc* variants. Fittingly, complete *AZFc* deletions do not alter H4 hyperacetylation levels in developing spermatids when compared to those recorded in nondeleted hypospermatogenic men (Kleiman et al. 2008).

- The ***DAZ*** (Deleted in Azoospermia) gene family consists of three different genetic determinants: *BOULE*, *DAZL* and *DAZ* (Xu et al. 2001). Of these, *DAZ* maps to *AZFc* and is consequently organized as a multi-copy gene family, while the remaining two are single-copy autosomal genes. The Y-linked *DAZ* genes are present in one copy per red *AZFc* amplicon, for a total of four copies in the reference *AZFc* sequence (*DAZ1*, *DAZ2*, *DAZ3* and *DAZ4*) (Saxena et al. 2000; Kuroda-Kawaguchi et al. 2001)sharing more than 99% sequence homology. The palindromic organization of the reference sequence results in the clustering of the *DAZ* copies to two red amplicon duplets, with the more proximal cluster in P2 palindrome containing *DAZ1* and *DAZ2*, and the more distal cluster in P1 palindrome containing *DAZ3* and *DAZ4.* These genes are specifically expressed in testis in pre-meiotic germ cells (particularly in spermatogonia). Historically, *DAZ* has been the focus of considerable attention both for its link to Yq microdeletion phenotypes and evolutionary origin. In fact, almost 10 years before the sequencing of the MSY, *DAZ* was considered to be the azoospermia factor (Reijo et al. 1995). This gene also corresponds to the first reported instance of an autosome to Y transposition. It has been proposed that the ancestral *BOULE* gene gave rise to *DAZL* by duplication and transposition in the early vertebrate lineage and that *DAZ* arose during primate evolution by transposition of *DAZL* to the Y chromosome (Saxena et al. 1996). The amplification of *DAZ* to four copies in the Y chromosome possibly results from tandem duplication and inversion events during evolution (Saxena et al. 2000; Kuroda-

Kawaguchi et al. 2001). The *DAZ* gene family (*BOULE*, *DAZL* and *DAZ*) encodes for RNA binding proteins characterized by two functional domains: a N-terminal RRM and a C-terminal DAZ repeat domain involved in protein-protein interactions (Reynolds and Cooke 2005). The number of these domains varies between the DAZ family members and may even be polymorphic in the case of Y-linked DAZ proteins (Lin and Page 2005). The four Y-linked *DAZ* gene copies encode for four RNA binding protein isoforms displaying cytoplasmic localization (Habermann et al. 1998; Reijo et al. 2000; Huang et al. 2008; Kim et al. 2009). These isoforms contain multiple DAZ repeat units (8 -24). The number of RRM domains varies from one, in the case of *DAZ2* and *DAZ3*, to 2 and 3 in the case of *DAZ1* and *DAZ4,* respectively. These proteins are thought be involved in the germ-cell progression to meiosis and the formation of haploid cells. They may act inducing the translational activation of developmentally regulated transcripts by binding to their 3'-UTR.

The *AZFc* region is also enriched in spliced but apparently non-coding transcripts of the testis-specific transcript Y-linked (TTTY) family (*TTTY3.1, TTTY4.1* and *TTY17.1*). Additionally, it contains an extensive array of pseudogenes including copies of *AZFb* and *AZFc* genes (*RBMY1A1*, *PRY*, *CDY* and *BPY2*), as well as the *AZFc*-exclusive sequence families *GOLGA2LY1* (Golgi autoantigen, golgin subfamily a, 2-like, Y-linked 1) and *CSPG4LYP1* (Chondroitin sulfate proteoglycan 4 pseudogene 1, Y-linked) (Fig.19).
Although the predicted function of the majority of the *AZFc* gene products is known their exact role in spermatogenesis remain unclear.
The *AZFc* deletion, accounting for approximately 60% of all recorded *AZF* deletions (Navarro-Costa et al. 2010), is associated with severe spermatogenic impairment phenotype (azoospermia or severe oligozoospermia) related to variable testicular pictures ranging from pure and mixed SCOS to hypospermatogenesis and maturation arrest. A deterioration of semen quality over time has been suggested for *AZFc* deleted oligozoospermic men based on indirect observations such as the difference in age between microdeleted patients with azoospermia and oligozoospermia or the increase of FSH concentrations over time in some subjects. However, this is nowadays a still debated issue.

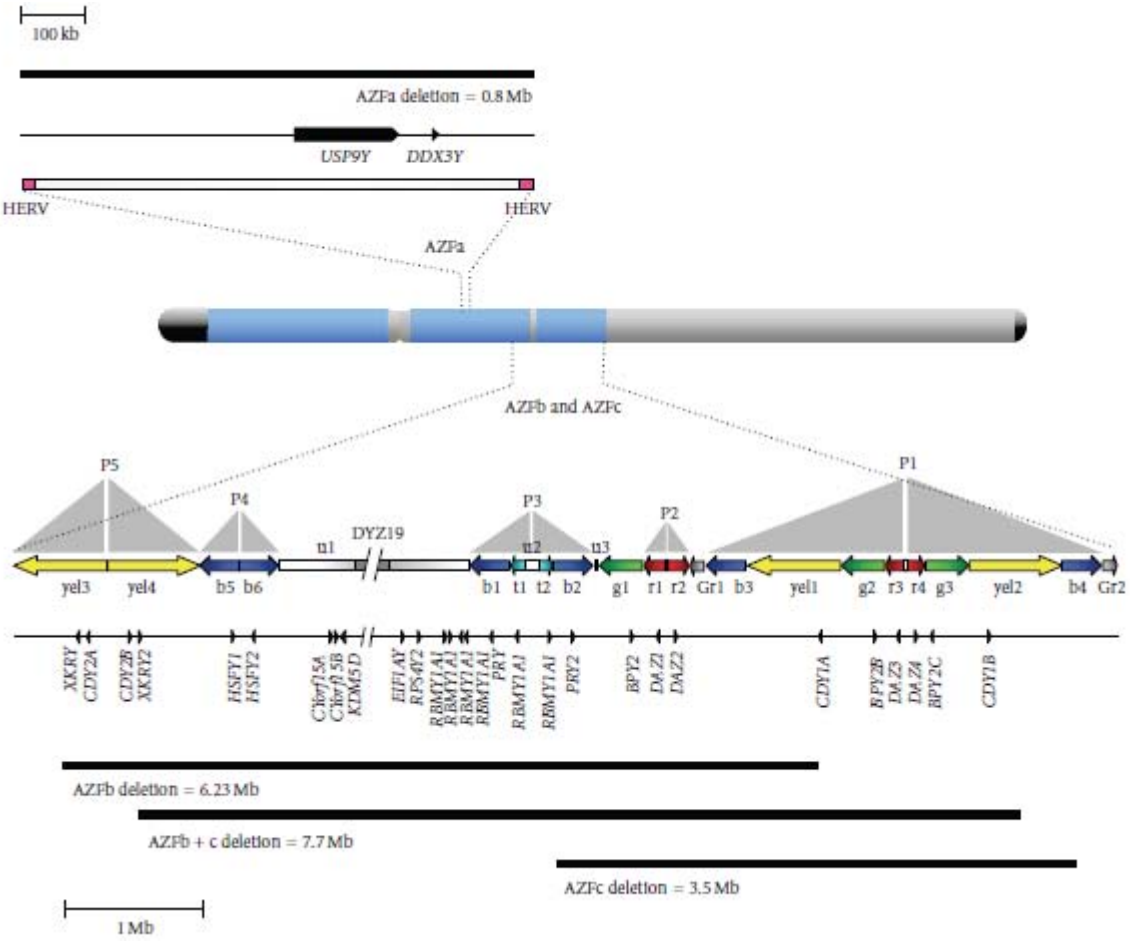


Figure 19. *AZF* deletion patterns. Recombining amplicons /palindroms responsible for each *AZF* deletions and genes involved are shown. *AZFa* is flanked by two HERV elements that mediate the occurrence of *AZFa* deletions via non-allelic homologous recombination. *AZFb* and *AZFb+c* deletions are caused by *P5*/proximal *P1* yel3/yel1) and P5-distal P1 (yel3/yel2) recombination, respectively. NAHR between *b2* and *b4* amplicons lead to *AZFc* deletion. Figure by *Navarro – Costa 2010*

The identification of Yq microdeletions is not only relevant for the diagnosis, allowing the etiology of the impaired spermatogenesis to be defined, but it may have prognostic value prior testicular biopsy (TESE) (Brandell et al. 1998; Krausz et al. 2000b). In this regard, in case of complete *AZFa* and *AZFb* deletions of the Y chromosome testicular biopsy is not advised because the chance of finding spermatozoa is virtually zero. *AZFc* deletion is compatible with the presence of spermatozoa in the testis or in the ejaculate, and is obligatorily transmitted to the male offspring. Therefore, genetic counseling is mandatory. The severity of spermatogenic failure of the son may vary substantially, however given the strict cause–effect relationship between *AZF* deletions and impaired spermatogenesis, normal spermatogenesis cannot be expected.

AZFc-linked CNVs

The reference *AZFc* sequence consists almost entirely (approximately 95%) of ampliconic units arranged in direct and/or inverted repeats (Kuroda-Kawaguchi et al. 2001; Repping et al. 2003). This peculiar architecture makes the *AZFc* region particularly prone to inter and intrachromatid homology-based recombination events and hence to structural variations such as CNVs and inversions. Highlighting the extreme plasticity of the *AZFc* region, a

multitude of possible *AZFc* architectures can be generated from the reference sequence in three or fewer homologous recombination events (Repping et al. 2006)
In addition to the complete *AZFc* (*b2/b4*) deletion, a number of recurrent internal rearrangements have been reported. They are overall regarded as partial *AZFc* rearrangements and include deletions and duplications both affecting *AZFc* gene dosage.

AZFc partial deletions
Three main partial *AZFc* deletion patterns have been identified: the *gr/gr*, *b2/b3* and *b1/b3*deletion (Repping et al. 2006) (Fig. 20). Their names derive from the colour-code of the ampliconic units involved in the underlying recombination event(r=red; b=blu; g=green) (Kuroda-Kawaguchi et al. 2001; Repping et al. 2003). Partial deletions remove between 1.6 and 1.8 Mb of the region and reduce approximately by half the overall dosage of the *AZFc* gene families (Table 2). They are described in details below.

- ***gr/gr deletion***: this partial *AZFc* deletion results from the recombination between green (g) or red (r) amplicons. Three recombination patterns can be identified in the reference *AZFc* sequence: *g1/g2, r1/r3, r2/r4* (Fig 20F). Additional recombination hot spots (*g1/g3, r1/r4* and *r2/r3*) are exclusively present in Y chromosomes carrying an inversion polymorphism in the *AZFc* region, i.e *b3/b4* inversion in *P1* palindrome (Repping et al. 2006) (Fig 20G). *gr/gr* deletion removes, in any case, 1.6 Mb of the *AZFc* region and, according to the sequence tagged site-polymerase chain reaction (STS/PCR) results as provided by the original report by Repping and colluegues (2006), it is defined by the absence of STS *sY1291* and the presence of all flanking markers, including the single copy marker *sY1191* and the multi-copy markers *sY1206* and *sY1161.* Eight testis-specific gene and transcription unit families (Table 2) are affected by this deletion pattern. Specifically, *gr/gr* deletion removes 2 copies of *DAZ* and 1 copy of *CDY1* which are the two most important *AZFc* candidate infertility genes. It is worth-noting that based on the *AZFc* structure and the underlying recombination event, a different types of *DAZ* (*DAZ1/2* or *DAZ3/4*) and *CDY1* copies *(CDY1a or CDY1b)* can be removed. Therefore, based on the missing copies of *DAZ* and *CDY1*, the following 4 *gr/gr* deletion subtypes can be defined: *DAZ1/2+CDY1a*; *DAZ1/2+CDY1b; DAZ3/4+CDY1a*; *DAZ3/4+CDY1b. gr/gr* deletion is the most commonly detected of the three partial *AZFc* deletion patterns. Its prevalence varies, indeed, from 2.1 to 12.5% in infertile men and from 0 to 10.2% among normozoospermic-fertile controls including Y hgrs with constitutive deletions (Stouffs et al. 2011)

- ***b2/b3 or g1/g3*** deletion removes 1.8 Mb of the *AZFc* interval and affects 12 genes and transcription units. This deletion pattern can occur only on two inverted variants of the *AZFc* region, the *gr/rg* and the *b2/b3* inverted organizations (Fig 20H), but not on the reference *AZFc* architecture (Repping et al. 2004; Repping et al. 2006). The *b2/b3* deletion arising on either of the two inverted *AZFc* variants resulting in the same organization of amplicons and in the loss of only *sY1192* and *sY1191.*

- ***b1/b3 deletion is caused by the recombination between b1 and b3 amplicons and*** removes 1.6 Mb of the *AZFc* region (Fig 20E). It is characterizaed by the absence of both *sY1291* and *sY1191* as well *sY1197* and *sY1161. b1/b3* deletion involves 12 *AZFc* gene copies and transcription units. *b1/b3* deletion is rare and its frequency varies amongst populations with only 18 deletions published to date (Eloualid et al. 2012)

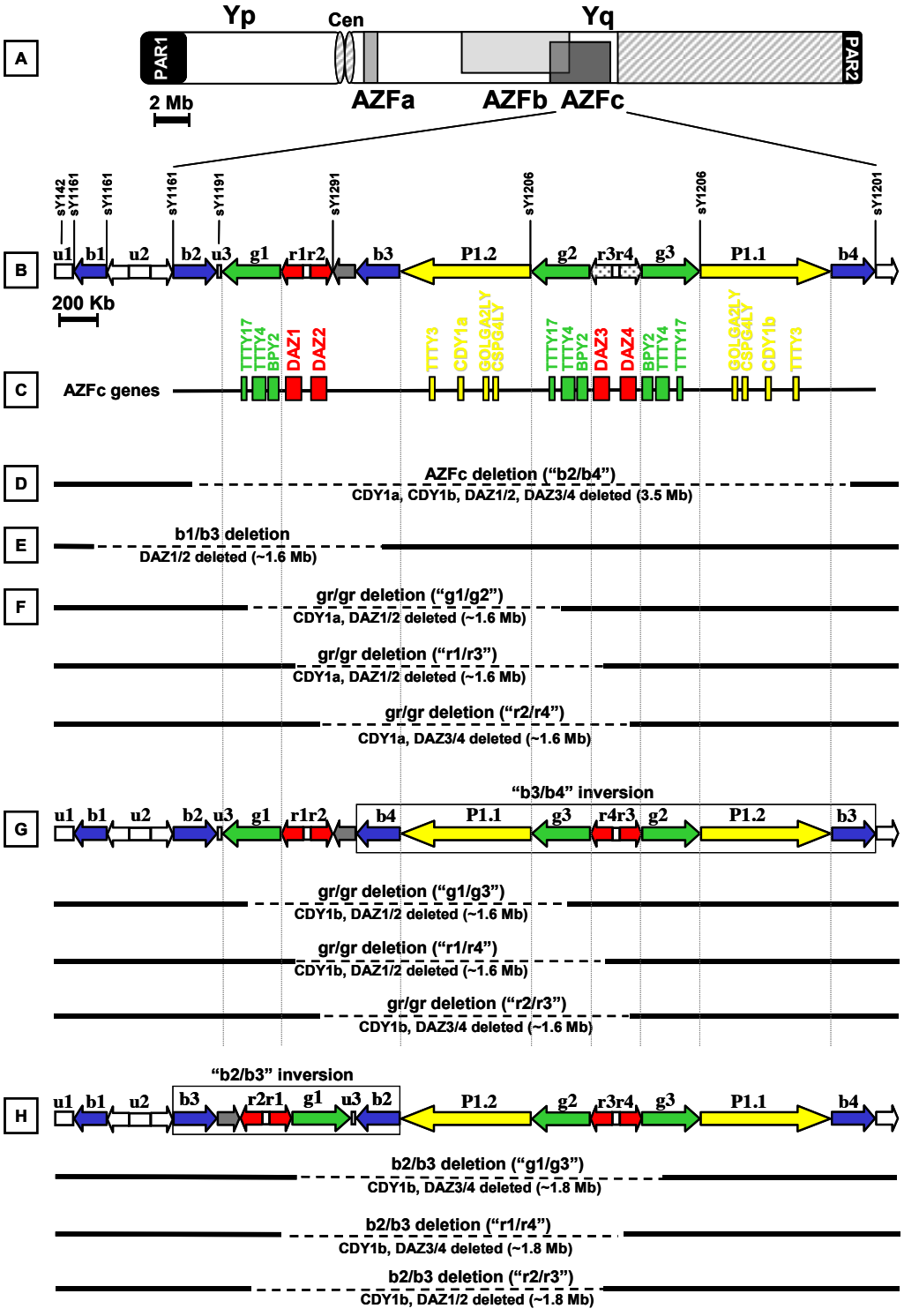


Figure 20. (Partial) *AZFc* deletions. Schematic representation of Y chromosome. The three *AZF* regions are shown (A). Reference sequence of *AZFc* regions, almost entirely composed of amplicons (indicated as colored arrows), and STSs localization (B). Location of the *AZFc* multicopy genes (C). Complete (*b2/b4*) *AZFc* deletion (D). *b1/b3* deletion (E); *gr/gr* deletion, possible recombination mechanisms and *DAZ* gene set and *CDY1* copy type removed are indicated (F); *b3/b4* inversion of the *AZFc* reference sequence and recombination pattern underlying *gr/gr* deletion with gene set and *CDY1* copy type removed (G); *b2/b3* inversion allowing a *b2/b3* deletion to occur, three possible deletion breakpoints are shown (H).

Table 2. Gene and transcription unit copy number in the reference *AZFc* sequence and in presence of partial deletions. Both "*gr/gr* " and "*b2/b3*" deletions remove 2 *DAZ* and 1 *CDY1* gene copies.

GENE	COPY NUMBER				
	No deletion	*"b2/b4"* Del.	*"gr/gr"* Del.	*"b2/b3"* Del.	*"b1/b3* Del."
RBMY	6	6	6	6	4
BPY2	3	0	2	1	2
DAZ	4	0	2	2	2
CDY1	2	0	1	1	2
PRY	2	2	2	2	0
CSPG4LY	2	0	1	1	2
GOLGA2LY	2	0	1	1	2
TTTY3	2	0	1	1	2
TTTY4	3	0	2	1	2
TTTY5	1	1	1	1	0
TTTY6	2	2	2	2	0
TTTY17	3	0	2	1	2
TOTAL	**32**	**11**	**23**	**20**	**20**

Genotype/phenotype correlation

Since the pilot study by Repping and colleagues (Repping et al. 2003), reporting a significant association between gr/gr deletion and spermatogenic impairment, the clinical significance of this recurrent deletion has been extensively explored in different populations. A number of studies corroborated the initial conclusions whereas others did not. Such inconsistencies can be related to a number of issues. First, methodological differences exist between studies, which are mainly related to the lack of confirmatory steps in some of them. As stated above, *gr/gr* deletion is defined by the absence of *sY1291* and the presence of all flanking STSs. Since *sY1291* PCR-based analysis can produce false-positive results (5.3% false gr/gr deletion rate has been reported by Krausz and colleagues (2009) (Krausz et al. 2009), validation steps such as the *CDY1* and *DAZ* gene dosage analysis are essential in order to confirm the effective loss of genetic material and discard false deletions. Therefore, results from studies performing only the STS-PCR analysis are likely to be affected by methodological biases.

Discrepancies between studies can also be related to a number of selection biases such as:

The inclusion of infertile men with known causes of impaired sperm production (non-idiopathic patients); this can "dilute" the frequency of *gr/gr* deletion in the patient group thus weakening or completely masking the association with spermatogenic impairment;

The presence of a high proportion of idiopathic azoospermic compared to oligozoospermic men. *gr/gr* deletion is more frequent among oligozoospermic patients, therefore the inclusion of a high proportion of azoospermic men may lead to miss those subjects which are the more likely carriers of *gr/gr* deletion;

The use of "fertile men" with unknown sperm count as controls. Since fertile subjects may also include men with abnormal sperm parameters (fertility is not synonymous of normozoospermia), normozooospermic men should be used as controls when exploring the association of a genetic variant (such as *gr/gr* deletion) with spermatogenic impairment;

The lack of ethic/geographic matching of patients and controls. In this case, the finding of a different *gr/gr* deletion frequency between the two study groups is more likely related to the existence of systematic differences in patients' and controls' Y background. Studies based on Asian populations are especially susceptible to this recruitment bias due to the high prevalence of the deletion-fixed Y hgrs D2, Q1 and Q3. As stated above gr/gr deletion constitutively present in those lineages does not affect sperm production, probably due to the compensatory effect of other Y-related factors. Therefore, the lack of association between *gr/gr* deletion and spermatogenic impairment in some studies based on Asian populations may be related to the over-representation of constitutive neutral *gr/gr* deletions. Accordingly, when occurring outside the deletion-fixed Y hgrs *gr/gr* deletion represents a significant risk factor for impaired sperm production also in these populations. At this regard, a recent study based on the Korean population, reported a significant association between *gr/gr* deletion and male infertility only in Y lineage not including the deletion-fixed D2b hgr (Y chromosome Alu Polymorphism, YAP-), suggesting that *gr/gr* deletion is associated with impaired spermatogenesis in Koreans belonging to the YAP- lineage (Choi et al. 2012). On the contrary, the phenotypic heterogeneity of *gr/gr* deletion is largely independent from the Y background in Caucasians as demonstrated by a multicentre survey based on the detailed characterization of *gr/gr* deletion subtypes and Y hgrs (Krausz et al. 2009).

With all the abovementioned limitations, 4 meta-analyses have been published on *gr/gr* deletion so far, all achieving significant Odds Ratios (ORs) with an average 2-2.5 fold increase of associated risk of spermatogenic failure. The available data, therefore, indicate that *gr/gr* deletion represents a significant risk factor for impaired sperm production (Tüttelmann et al. 2007a; Visser et al. 2009; Navarro-Costa et al. 2010; Stouffs et al. 2011). The clinical relevance of *gr/gr* deletion has been confirmed further by a recent multiethnic population-based survey

of over 20.000 Y chromosomes, reporting a 1.9-fold increased risk of severe spermatogenic failure associated to this partial deletion (Rozen et al. 2012). Due to the extreme variability of deletion frequencies in different countries, which clearly limits the *gr/gr* deletion screening in admixed populations, the calculation of the effect size should be done separately for distinct populations.

Initial studies on *b2/b3* deletion failed to identify any association between this partial rearrangement and spermatogenic failure (Fernandes et al. 2004; Repping et al. 2004) mainly because of the high prevalence of hgr N1 chromosomes. *b2/b3* deletion is indeed fixed on this Y hgrwhich is highly prevalent in Northern Eurasian populations (especially in Baltic sea area). On the contrary, a strong association was found between *b2/b3* deletion and spermatogenic failure in populations with a different Y background, as Han Chinese and Moroccan populations (Wu et al. 2007; Eloualid et al. 2012). Similarly to that observed in relation to the phenotypic expression of constitutive *gr/gr* deletion, the *b2/b3* deletion represents a risk factor for spermatogenic impairment when occurring outside N1 hgr. A propensity to evolve in a complete *AZFc* deletion in the next generation has been suggested for both *b2/b3* and *gr/gr* partial deletions (Zhang et al. 2007; Lu et al. 2009).

The *b1/b3* deletion is a rare event, with an estimated 0.1% (1 every 994 males) prevalence in the general population and of 0.25% among men with impaired sperm production. Due to its rarity, the role of *b1/b3* deletion in spermatogenic impairment remains elusive. Nevertheless, according to estimates of population-genetics parameters for this partial *AZFc* deletion, *b1/b3* deletion would increase the risk of severely impaired sperm production by a factor of 2.5. Even so, only 1.8% of men with the *b1/b3* deletions show severely impaired sperm production. For the remaining 98.2% the phenotype could range from oligozoospermia to normozoospermia (Rozen et al. 2012).

The prevalence and clinical significance of partial *AZFc* rearrangements in the Spanish population has been little explored since only one pilot study on *gr/gr* deletion has been performed so far suggesting a significant association between this partial deletion and spermatogenic impairment (de Llanos et al. 2005).

AZFc partial duplications

The inter-chromatid NHAR, possible underlying mechanism of (partial) *AZFc* deletions, always generates reciprocal deletion and duplication. This fits with the existence of the *gr/gr* and *b2/b3* duplication which are reciprocal of the homonym partial deletions (Lin et al. 2007). These duplications increase the dosage of 9 and 12 *AZFc* genes, respectively.

A third duplication pattern, the *b2/b4* duplication has been frequently reported as secondary to a *gr/gr* or *b2/b3* deletion. This type of rearrangement named "deletion-duplication" requires at least two round of NAHR and can restore or even increase, in case of multiple NAHR events, the *AZFc* gene dosage (Fig. 21).

A detailed characterization of Y chromosome belonging to different lineages found limited variation in the Y-linked genes copy number (Repping et al. 2006). In this regard, about 90% of men carriers four *DAZ* and two *CDY1* copies implying that a normal spermatogenesis requires an optimal gene dosage. Supporting this observation, Noordam and colleagues (Noordam et al. 2011) reported a significant reduced total motile sperm count (TMSC) in men with lower *DAZ* dosage, in respect to the optimal one, as well as a lower TMSC in a single case of increased *DAZ* copy number due to a *gr/gr* deletion followed by *b2/b4* duplication. Therefore, not only the reduction but also the increase of *AZFc* gene dosage may have a negative effect on sperm production. This hypothesis led three different research groups to explore the clinical relevance of *AZFc* partial duplications in male infertility. The first study addressing this topic (Lin et al. 2007) reported a significant association between *AZFc* partial duplications and spermatogenic impairment in the Han Chinese population of Taiwan. These results were then replicated in the Han Chinese population of the East of China (Lu et al. 2011) but not in the Italian population (Giachini et al 2008) where *AZFc* partial duplications have been reported not to be associated with spermatogenic impairment. Since this discordance may reflect genuine ethnic differences, if increased *AZFc* gene content is to play a role in the spermatogenic impairment, the effect will probably be modulated by population-specific factors (Krausz et al. 2011).

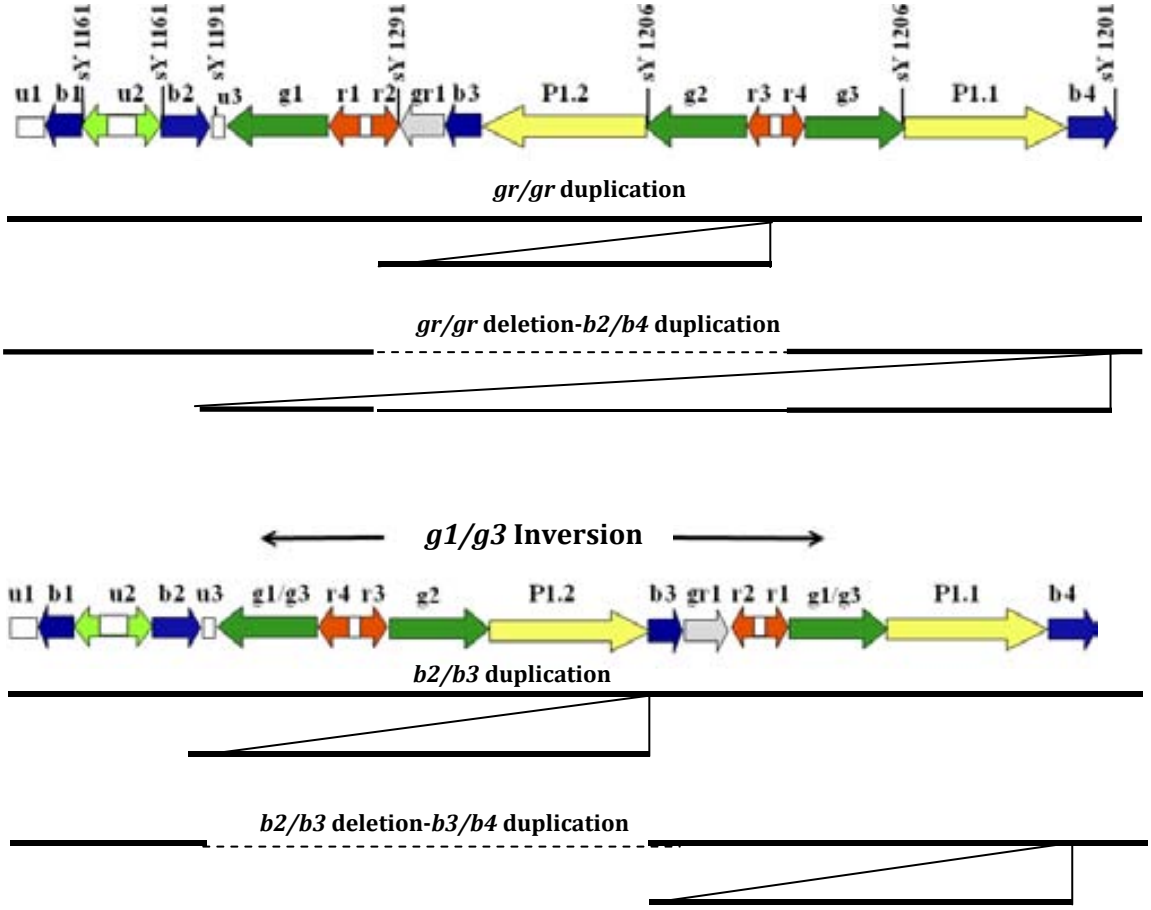


Figure 21. Partial *AZFc* duplications and deletion-duplications. At the top are shown the reference *AZFc* architecture and the "*gr/gr*" CNVs: *gr/gr* duplication and *gr/gr* deletion followed by *b2/b4* duplication. In the middle is shown a structure derived from the reference architecture through *g1/g3* inversion and the "*b2/b3*" CNVs: *b2/b3* duplication and *b2/b3* deletion followed by *b3/b4* duplication.

3.3.4 The X chromosome

The X chromosome is a sub-metacentric chromosome showing many features that are unique in the human genome. According to the novel reassembly of the human chromosome reference, the total X-chromosome size has been estimated to be about 155.3 Mb. It displays a low (G+C) content (39%) compared with the genome average (41%) and it is highly enriched in repetitive sequences. These regions account for 56% of the euchromatic X-chromosome sequence and are represented by:

- Short Interspersed Nuclear Elements (SINEs) belonging to the *Alu* family, the content of which in the X chromosome is below the genome average.
- Long Terminal Repeats (LTRs) the coverage of which is above average.

- Long Interspersed Nuclear Elements (LINEs) of the L1 family, which are the most represented class of repetitive elements of the X chromosome, accounting for 29% of the chromosome sequence compared to a genome average of only 17% (Ross et al.2005).
- Ampliconic sequences (segmental duplications of >10 Kb sharing > 99% nucleotide identity) represent approximately 2% (3.15 Mb) of the chromosome length (Mueller et al. 2013).

The cross-species alignment of orthologous X-linked genes highlighted the presence of two evolutionary domains of the X chromosome:

- The X-conserved region (XCR), an ancestral region including all the long arm and PAR1, which would descend from the proto-X chromosome, one of two 'proto' sex chromosomes evolved from the ancestral autosome pair according to the Ohno's theory (Ohno 1967). This evolutionary domain is share by all mammals (placental and not).

- X-added region (XAR) including the short arm and PAR2 region, which established on X chromosome by translocation from a second autosome. This region is exclusively present in placental mammals.

The X chromosome has a low gene density, with half as many genes per megabase on the X chromosome as on the average human autosomes (Ross et al. 2005). The unusually low gene density of the X chromosome is probably due to the massive expansion of non-coding intergenic sequences during evolution, which spread the genes further apart.
Bellot and colleagues (Bellott et al. 2010) estimated that intergenic regions were expanded by about 80 Mb in the X chromosome. No single class of non-coding sequence can account for this change, but the significant enrichment for LINEs on X chromosome suggests that the expansion of the intergenic sequence may have been driven by recurrent insertions and divergence of transposable elements. Accordingly in the human genome low gene density is often associated with increased interspersed repeat content, specifically LINEs (Lander et al. 2001) and this association is also observed in the X chromosome.
The new assembly 800 protein coding genes map to the human X chromosome. Among them 144, corresponding to 18%, violate the so-called Ohno's law stating that the gene content of X chromosomes is conserved among placental mammals. These genes in fact do not show orthologs in mouse and in other species. The majority of them (76/144) would have been acquired independently by the human X chromosome, through transposition or retroposition from autosomes or through having arisen *de novo*, since humans and mice began to diverge from a common ancestor 80 million years ago. Among the independently acquired X-linked genes, roughly two-thirds (48/76) are ampliconic (i.e embedded in duplicated segments of >10 kb in length and exhibiting >99% nucleotide identity), whereas the remaining one-third are multicopy (only the gene structure is duplicated) or single copy. Overall only 31% of the human X-ampliconic genes had orthologs in the other species. Ampliconic genes are predicted to have a function in male fitness (see paragraph below).

X-linked genes and spermatogenesis
It has been recently reported that most independently acquired human and mouse X-linked genes exhibit high expression in the testis and little or no expression in other tissues. In mice, this prevalent testis expression is related to the male germ cell-restricted expression of these genes regardless of whether they are single, multi-copy or ampliconic (Mueller et al 2013).

These novel findings are in line with previous genomic studies reporting an enrichment on the mammalian X chromosomes, compared to the autosomes, for male-specific single and multi-copy genes showing testis-restricted or predominant expression (Wang et al. 2001; Lercher et al. 2003; Mueller et al. 2009; Zheng et al. 2010). Given that the the independently acquired genes are expressed predominantly in spermatogenic cells, one might anticipate that loss-of-function mutations affecting these genes or gene families would perturb male gametogenesis

Based on the reported X chromosome enrichment for single copy genes expressed during the early stages of murine spermatogenesis, it was originally suggested that mainly pre-meiotic genes were located on the X chromosome (Wang et al 2001). Accordingly, X chromosome is transcriptionally active only in mitotically dividing spermatogonia and in the early meiotic (pre-pachytene) spermatocytes. During meiosis X-linked genes undergo the so-called meiotic sex chromosome inactivation (MSCI) and are thus transcriptionally silenced (Zeng et al. 2010). Notwithstanding, evidences have been recently reported of transcriptional activity also during and after meiosis. Indeed, many microRNA have been found to be expressed at the pachytene stage, when MSCI occurs (Song et al. 2009). The escape from MSCI silencing by X-linked microRNA suggests that they may contribute to MSCI or be involved in post-transcriptional regulation of autosomal mRNA during meiotic and post-meiotic stages of spermatogensis. In addition, a post-meiotic transcription reactivation has been reported for several multi-copy mouse X-linked gene families (Wang et al. 2005; Mueller et al. 2008) showing higher expression levels compared to single copy genes (Fig. 22). It was therefore hypothesized that increasing copy number may be a mechanism to counteract transcriptional repression of the X chromosome in post-meiotic germ cells.

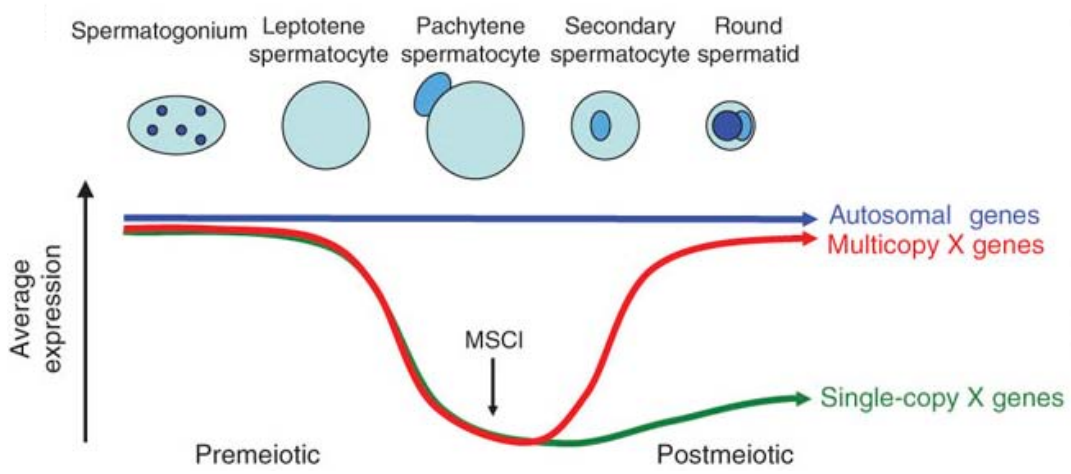


Figure 22. Multi-copy genes evade the effects of X chromosome postmeiotic repression in the mouse. Single-copy and multi-copy X-linked genes exhibit similar average levels of expression during pre-meiotic spermatogenesis. All X-linked genes are subsequently silenced during MSCI. Following MSCI, single-copy X-linked genes exhibit low reactivation levels whereas multi-copy X-linked genes exhibit expression levels similar to autosomal genes, thus evading the effects of post-meiotic repression.

The most represented X-linked testis specific gene families are the CTA genes, which have been suggested to account for 10% of human X-chromosome gene content (Ross et al 2005).

CTA genes are defined by a unique expression pattern: amongst normal tissues, they are expressed exclusively or predominantly in male germ cells and in embryonic trophoblasts, but their gene products are also found in a significant number of human tumors of different histological origin. At least 70 families of CTA genes with over 140 members have been identified so far and recently listed in a database established by the Ludwing Institute for Cancer Research (http:// www.cta.Incc.br/) (Almeida et al. 2009). The X-linked CTA genes (X-CTA) represent more than half of all CTA genes and often constitute multi-copy gene families organized in well-defined clusters along the X chromosome, where the different members are arranged into complex direct and inverted repeats (segmental duplications) (Fratta et al. 2011). This feature account for the susceptibility of CTA genes to CNVs even though their multi-copy gene status may be a strategy to increase the chance to escape MSCI during meiosis, as observed for mouse X-linked multi-copy genes.

The MAGE (Melanoma antingen) and GAGE (G antigen) are the largest and best known X-CTA gene families containing at least 24 and 16 members, respectively (Stouffs et al. 2009).
The biological function of most X-CTA genes is still largely unknown. However, evidence is emerging that the best studied of these, the MAGE genes, can act as signal transducing transcriptional modulators. Moreover, MAGE genes appear to be able to mediate proliferative signals (Park et al. 2002; Duan et al. 2003; Glynn et al. 2004). In normal testis, X-CTA genes are expressed primarily in spermatogonia. According to the so-called Rice's theory, such enrichment of male-specific genes on the X chromosome would be related to the accumulation of recessive alleles/genes with beneficial effect for men ("masculinization" of the X chromosome). Indeed, recessive alleles that are beneficial to males are expected to become fixed more rapidly on the X chromosome than on an autosome (Hurst 2001) and if these alleles were detrimental to females, their expression could become restricted to male tissues.

X chromosome and male infertility

Two major characteristics of this chromosome make it an interesting object of study in the field of male infertility. As stated above, it is enriched in genes potentially involved in spermatogenesis. Furthermore, because of the hemizygous condition of males for genes located on this chromosome, any *de novo* mutation may have an immediate impact.
Low-frequency *de novo* mutations are expected to be involved in the etiology of male infertility since mutations in genes crucial for spermatogenesis will obviously not be transmitted through generations and thus will not be spread through the population.
Furthermore, considering the low prevalence of single gene mutations in candidate spermatogenesis genes, it is currently postulated that infertility should be regarded as a polygenic disease (Cram et al. 2004). In this view, the classical candidate gene approach, focusing on single genes of interest, is a limited strategy as highlighted by the paucity of X-linked candidate gene mutations (*AR, SOX3, USP26, NXF2, TAF7L, FATE* and *AKAP*), since potentially causative mutations have been reported only in the *AR*, to date.
Conversely, high-throughput technologies such as the Comparative Genomic Hybridization (aCGH) and SNP arrays, allowing the scanning of entire chromosomes or the whole genome, provide a global view of the individual genetic background. These technologies are being demonstrated a useful tool in exploring the genetic of male factor infertility and have allowed the identification of novel genetic factors, among them a number of X-linked CNVs of potential clinical relevance in the etiology of male infertility (Tüttelmann et al. 2011; Krausz et al. 2012; Stouffs et al. 2012; Lopes et al. 2013).

4. AIMS OF THE THESIS

The global aim of the present thesis is to enhance our understanding on genetic factors involved in non-syndromic cryptorchidism and spermatogenic impairment.

The first part of the thesis was focused on the study of two SNPs in relation to non-syndromic cryptorchidism. Specific aims of this part are:

- To clarify the clinical significance of the T222P variant of *RXFP2* (a gene coding for the receptor of the major hormone responsible for the abdominal phase of testis descent) in non-syndromic cryptorchidism, in two Mediterranean populations (Spanish and Italian).

- To test whether the $(TA)_n$ polymorphism in the promoter of the *ESR1* gene is associated with isolated cryptorchidism, in order to evaluate whether this polymorphism may predispose to the disease in the context of gene-environmental interaction.

The second part of the thesis was aimed at investigating the role of sex-chromosomes linked CNVs in the etiology of spermatogenic impairment, specifically:

- To provide new insights into a number of still debated issues concerning the routine genetic testing for Y microdeletions through a comprehensive phenotypic and genotypic description of consecutive infertile men screened for Y microdeletions.
- To corroborate the role of *gr/gr* deletion as significant risk factor for spermatogenic impairment in Spanish population.
- To provide for the first time data on the prevalence and clinical significance of *AZFc* partial duplications in the Spanish population**.**

- To identify X-linked CNVs potentially involved in male infertility.
- To investigate the role of 3 recurrent X-linked deletions in the etiology of spermatogenic impairment.

5. RESULTS

The results of the present thesis can be divided into two parts. The first one concerns the study of two genetic variants of *ESR1* and *RXFP2* genes in relation to non-syndromic cryptorchidism. This part includes the following two papers:

1) Further insights into the role of T222P variant of *RXFP2* in non-syndromic cryptorchidism in two Mediterranean populations.
Ars et al Int J Androl. 2011; 34:333-8. PMID: 20636340

2) *ESR1* promoter polymorphism is not associated with non-syndromic cryptorchidism.
Lo Giacco et al Fertil Steril. 2011; 95:369-71. PMID: 20887985

The second part of results refers to the investigation of the role of Y and X-linked CNVs in the etiology of spermatogenic impairment. This section includes the following papers:

3) Clinical relevance of Y-linked CNV screening in male infertility: new insights based on the 8-year experience of a diagnostic genetic laboratory.
Lo Giacco et al 2013 (under review at European Journal of Human Genetics)

4) High Resolution X Chromosome-Specific Array-CGH detects New CNVs in infertile males.
Krausz et al PLoS One. 2012; 7:e44887. PMID: 23056185.

5) Recurrent X chromosome-linked deletions: discovery of new genetic factors in male infertility.
Lo Giacco et al 2013 (under review at Journal of Medical Genetics)

Papers 1, 2 and 4 form part of the *compendium* whereas papers 3 and 5, not yet accepted for the publication, do not. Nevertheless, the results of these two works will be presented and discussed in the following sections, being an important part of the present thesis.

PAPER 1

Further insights into the role of T222P variant of *RXFP2* in non-syndromic cryptorchidism in two Mediterranean populations.

Ars E., Lo Giacco D., Bassas L., Nuti F., Rajmil O., Ruíz P., Garat J.M., Ruiz-Castané E., Krausz C. *International Journal of Andrology*. 2011 34:333-8. (PMID: 20636340)

SUMMARY 1

Background:
The etiology of non-syndromic cryptorchidism is still largely unknown. Given the key role of INSL3/RXFP2 system in controlling the transabdominal phase of testis descent, genetic variants in the respective genes represent the strongest candidate genetic factors in respect to monogenic causes of non-syndromic cryptorchidism. In particular, the T222P genetic variant represents the so far most extensively studied and, at the same time, the most controversial genetic variant of *RXFP2*.

Aim:
Given the controversies in the literature and the reported sharp contrast between Italian and Spanish populations, we sought to investigate further into the clinical significance of T222P variant in these two Mediterranean populations. Moreover, we also aimed to verify the hypothesis of a "founder effect" as underlying factor of the high prevalence of T222P in the Mediterranean area.

Materials & Methods:
A total of 577 subjects from Spain (187 cryptorchid and 390 controls) and 550 from Italy (199 cryptorchid and 351 controls were genotyped for the T222P variant by direct sequencing.
For *RXFP2* haplotype analysis, a total of 5 exonic and intronic polymorphisms were analyzed in 9 T222P carriers and in relatives of three of them.

Results:

In the entire study population the T222P was found, always in heterozygosity, in a total of 12 cryptorchid patients (3 Spanish and 9 Italian) and 12 controls (7 Spanish and 5 Italian). This finding excludes a causative relationship between the heterozygous variant and cryptorchidism. The T222P variant was found at a similar frequency in both cases (1.6%) and controls (1.8%) in the Spanish population, whereas in Italy, the frequency of T222P resulted significantly higher in cryptorchid patients (4.5%) than controls (1.4%) (p=0.031; OR= 3.17 [95% confidence interval (CI): 1.07–9.34). The observed difference between the two countries and the highly variable phenotypic expression of the T222P variant may depend on the genetic background or on environmental conditions. The haplotype analysis of the *RXFP2* in 3 Spanish T222P carriers and their parents showed that this variant is linked to the previously inferred C-C-G-A-13 haplotype and consequently provides further support to the 'founder effect' hypothesis.

Conclusion:
Our data indicate that T222P is a frequent variant in the Spanish population with no pathogenic effect. Although in Italy it seems to confer a mild risk to cryptorchidism, the screening for this variant for diagnostic purposes is not advised because of the relatively high frequency of control carriers.

International Journal of Andrology
international journal of andrology ISSN 0105-6263

ORIGINAL ARTICLE

Further insights into the role of T222P variant of *RXFP2* in non-syndromic cryptorchidism in two Mediterranean populations

E. Ars,*†[1] D. Lo Giacco,*†[1] L. Bassas,† F. Nuti,‡ O. Rajmil,† P. Ruíz,* J. M. Garat,§ E. Ruiz-Castané† and C. Krausz†‡

*Molecular Biology Laboratory, Fundació Puigvert, Universitat Autònoma de Barcelona, Barcelona, Spain, †Andrology Service, Fundació Puigvert, Universitat Autònoma de Barcelona, Barcelona, Spain, ‡Andrology Unit, Department of Clinical Physiopathology, Center of Research, Transfer and Higher Education on Chronic, Inflammatory, Degenerative and Neoplastic Disorders, University of Florence, Florence, Italy, and §Paediatric Urology Division, Urology Service, Fundació Puigvert, Barcelona, Universitat Autònoma de Barcelona, Barcelona, Spain

Keywords:
cryptorchidism, gene polymorphisms, genetics, RXFP2, testis

Correspondence:
Csilla Krausz, Andrology Unit, Viale Pieraccini, 6 50139 Firenze, Italy.
E-mail: c.krausz@dfc.unifi.it
Elisabet Ars, Molecular Biology Laboratory, Fundació Puigvert, Cartagena, 340-350 08025 Barcelona, Spain.
E-mail: ears@fundacio-puigvert.es

[1]These authors contributed equally to the study.

Received 3 December 2009; revised 1 April, 21 April and 19 May 2010; accepted 22 May 2010

doi:10.1111/j.1365-2605.2010.01088.x

Summary

The aetiopathogenesis of isolated cryptorchidism remains largely unknown. Mutation screenings in the most relevant candidate genes for testicular maldescent lead to controversial data in the literature. In particular, the role of the T222P genetic variant of the *RXFP2* gene is still debated. Given the controversies, the aim of this study was to provide further data on this genetic variant in two Mediterranean populations. A total of 577 subjects from Spain and 550 from Italy (with and without a history of cryptorchidism) were analysed. The T222P substitution was found in both unilateral and bilateral cases and in a total of 12 controls. These data exclude a clear-cut cause–effect relationship between T222P variant and testicular maldescent. The T222P variant was found at a similar frequency in both cases and controls in the Spanish population, whereas in Italy, the frequency of T222P resulted significantly higher in the cryptorchid group ($p = 0.031$). The observed difference between the two countries and the highly variable phenotypic expression of the T222P variant may depend on the genetic background or on environmental conditions. The haplotype analysis of the *RXFP2* gene in T222P carriers and their parents showed that this variant is linked to the previously inferred C-C-G-A-13 haplotype and consequently provides further support to the 'founder effect' hypothesis. In conclusion, our data indicate that T222P is a frequent variant in the Spanish population with no pathogenic effect. Although in Italy it seems to confer a mild risk (odds ratio = 3.17, 95% confidence interval: 1.07–9.34) to cryptorchidism, the screening for this variant for diagnostic purposes is not advised because of the relatively high frequency of control carriers (1.4% of Italian men without a history of cryptorchidism).

Introduction

Cryptorchidism is the most frequent congenital birth defect in male children and can occur as an isolated disorder or in association with other congenital anomalies (syndromic cryptorchidism) (Toppari *et al.*, 2006). Our understanding of the physiology of testis descent has greatly advanced in the last few years (Ivell & Hartung, 2003; Virtanen & Toppari, 2008). It has become evident that a major role in testis descent is played by two major hormone/receptor systems: testosterone and its receptor AR (androgen receptor) and the INSL3/RXFP2 (former

LGR8) system, insulin-like factor 3 (INSL3) and its G-protein-coupled receptor, RXFP2 (relaxin family peptide 2). INSL3/RXFP2 are essential for the first phase of testicular descent, whereas androgens are involved in the inguinal phase (Foresta *et al.*, 2008). It was predicted that mutations in genes coding for the two hormone/receptor systems should cause cryptorchidism. Since the first mutation screening of the *INSL3* gene in cryptorchid men (Krausz *et al.*, 2000), a large number of polymorphisms and putative mutations have been described (Adham & Agoulnik, 2004; Ferlin *et al.*, 2006). Of all reported mutations, only two missense mutations, the P49S and V18M, have a deleterious effect on the ability of INSL3 to activate its receptor (Bogatcheva *et al.*, 2003; El Houate *et al.*, 2007). However, the role of these mutations (all heterozygous) in the pathogenesis of cryptorchidism remains to be established. Concerning the *RXFP2* gene, only one missense mutation at codon 222 (T222P) has been reported. According to the first two studies, this mutation was present in heterozygosity exclusively in men with history of testicular maldescent (Gorlov *et al.*, 2002; Ferlin *et al.*, 2003). In addition, functional analysis showed a severely reduced receptor surface expression of the mutant protein further supporting its causative role in the pathogenesis of cryptorchidism (Bogatcheva *et al.*, 2007). However, the proposed clear-cut cause–effect relationship between heterozygous mutation and cryptorchidism has been recently challenged by a multicentre study, which has clearly demonstrated that this genetic variant may not be considered further as a pathogenic mutation as it can be also found in men with no history of testicular maldescent (Nuti *et al.*, 2008). In this study, it was also demonstrated that the T222P variant is not restricted to the Mediterranean region, as it was observed also in a Central European country. Another recent study from Morocco has also reported a high frequency for this variant in the range of polymorphism in both cryptorchid men and non-cryptorchid men (El Houate *et al.*, 2008). Despite these controversies, the T222P variant has been reported as a genetic alteration associated with persistent bilateral cryptorchidism in a recent study on newborn children from Italy (Ferlin *et al.*, 2008).

The role of T222P variant in Spain has not been fully investigated and only a small study population (77 patients and 48 controls) was screened in a multicentre study (Nuti *et al.*, 2008). In the aforementioned study, the frequency of T222P mutation was 0% in the patient and 4% in the control groups, showing an inverse situation to that observed in the Italian population in which 5% of patients had the mutation vs. 1.8% of the controls.

Given the controversies in the literature and the apparent sharp contrast between Italy and Spain, we sought to investigate further into the clinical significance of this genetic variant in the two Mediterranean populations by screening a total of 386 cases and 741 controls. Moreover, we also aimed to clarify the founding mutation issue by performing the *RXFP2* gene haplotype analysis in the mutation carriers and in their parents.

Materials and methods

Study populations

A total of 577 subjects of Spanish ancestry were analysed for the T222P variant of *RXFP2*. Among them, 187 patients had a history of cryptorchidism, whereas the remaining 390 were controls with no history of testicular maldescent at birth. The cryptorchid group included 113 with unilateral (1 of them had 'retractile' testis), 74 with bilateral cryptorchidism (6 of them had 'retractile testis'). Part of this study population has been already screened in the multicentre study by Nuti *et al.* (2008) (77 patients and 44 controls). The Italian study population included 550 subjects from Central Italy (Italian ancestry): among them, 199 had a history of testicular maldescent at birth, whereas 351 were controls. The cryptorchid group included 132 with unilateral and 67 bilateral cryptorchidism. Also in this case, part of the study population has been analysed in the previous multicentre study (Nuti *et al.*, 2008): 159 patients and 275 controls.

Molecular analysis of *RXFP2* gene

DNA was extracted from peripheral blood in all the participants in this study with the exception of the Italian controls for whom, depending on sample availability, the genomic DNA was isolated either from the peripheral blood or from the frozen semen.

Exon 8 of the *RXFP2* gene (RefSeq: NM_130806.3) containing the T222P variant, was amplified using the following primers: forward: 5′-GGGGAGGCAGGTTT TATTTC-3′; reverse: 5′-AAGCTAGTGCTAGATGTCATT GC-3′. The resulting DNA fragments were analysed by direct bidirectional sequencing using automated sequencer (ABI PRISM 3100; Applied Biosystems, Foster City, CA, USA). For *RXFP2* haplotype analysis, a total of five *RXFP2* polymorphisms were analysed in nine carriers of the T222P mutation. For the analysis of exon 12–intron 12 polymorphisms (exon 12: 957G>A, 993A>G; intron 12: INV12(-2)A9 > A13), the primers were described by Ferlin *et al.* (2003). Primers for the intron 7 [INV7(-351)A>C] and intron 9 polymorphisms [INV9(+380) A>C] and polymerase chain reaction conditions were described by Nuti *et al.* (2008). The resulting DNA fragments were sequenced using automated sequencer (ABI PRISM 3100; Applied Biosystems).

Statistical analyses

For statistical analysis of genotype distribution, test for deviation of Hardy–Weinberg Equilibrium (HWE) or two-point association studies, we employed tests adapted from Sasieni (1997). Odds ratio (OR) estimates were computed from 2 × 2 tables. These calculations were performed on the online resource facility at the Institute for Human Genetics, Munich, Germany (http://ihg.gsf.de) and using the Statistical Package for the Social Sciences software (SPSS, Evanston, IL, USA).

Results

Sequencing analysis of exon 8 of the *RXFP2* gene

Spanish study population

The T222P genetic variant was found in both cryptorchid and control subjects. The variant frequency observed in the two groups was almost the same; in the cryptorchid group 3/187 (1.6%), whereas in the controls 7/390 (1.8%).

Italian study population

The variant was significantly more frequent in the patient group (9/199, 4.5%) than in the control (5/351, 1.4%), with an OR = 3.17 [95% confidence interval (CI): 1.07–9.34, $p = 0.031$]. Data for both study populations are reported in Table 1.

Table 1 T222P *RXFP2* variant in patients with a history of testicular maldescent and in controls with no history of cryptorchidism

	No. of mutated/ total men	T222P variant frequency (%)	Statistical analysis, *p*-value [odds ratio (95% CI)]
Spanish			
Cryptorchid	3/187[a]	1.6	0.8 [0.89 (0.2–3.4)]
Unilateral	2/113	1.7	
Bilateral	1/74	1.3	
Control	7/390[a]	1.8	
Italian			
Cryptorchid	9/199	4.5	0.031
Unilateral	5/132	3.7	[3.17 (1.07–9.34)]
Bilateral	4/67	5.9	
Control	5/351	1.4	

[a]A total of 77 patients and 48 controls from Spain and 159 patients and 275 controls of Italian origin, from a previous multicentre study are included (Nuti *et al.*, 2008).

RXFP2 gene haplotyping

The analysis of five exonic and intronic polymorphisms allowed the discrimination of different possible haplotypes on the carriers (Table S1). The analysis of Spanish triplets (mother, father and son) available for two patients and of one control/mother pair also indicated that the T222P is in linkage with the C-C-G-A-13 haplotype (Fig. 1). In all three carriers, the variant had been transmitted from the father. Patient 08-202 shared with his father the C-C-G-A-13 haplotype, whereas patient sw09-9 shared the C-C-G-A-12 allele. In the case of the control carrier (07-276), only the mother was available for the analysis and she did not carry the variant. The haplotype in linkage with the variant in this subject is C-(C/A)-G-(A/G)-13, which potentially includes the C-C-G-A-13 haplotype.

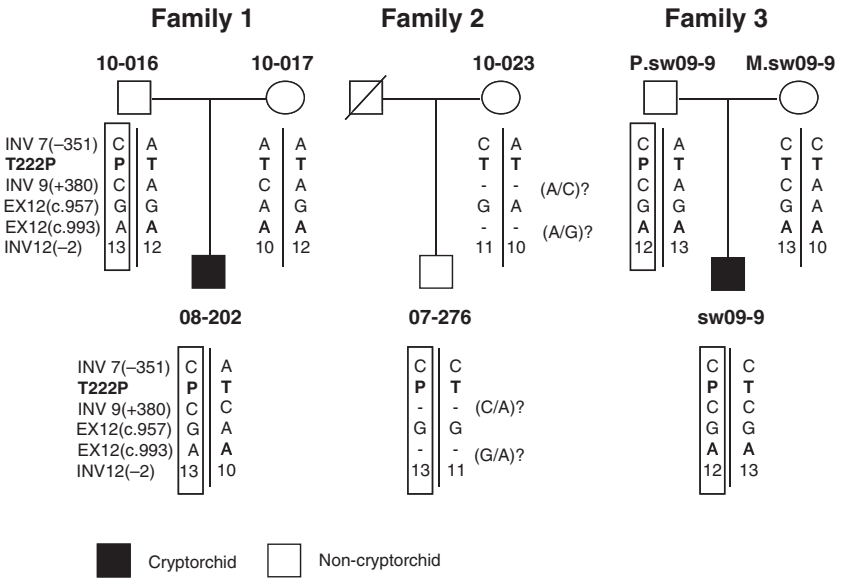


Figure 1 Haplotype analysis of the three Spanish families carrying the T222P variant of the *RXFP2* gene. The haplotype linked to the T222P variant is boxed. In all three carriers, the variant is inherited from the father.

Genotype/phenotype correlation

In all 12 cases and 12 controls, the T222P variant was found in heterozygosity. The phenotypic features of T222P variant carriers were heterogeneous in both geographical groups. In fact, the phenotypic expression of this variant also includes normal testicular descent in 12 cases (7 from Spain and 5 from Italy). This finding further excludes a causative relationship between the heterozygous variant and cryptorchidism. The testicular phenotype of the 12 cryptorchid patients was unilateral cryptorchidism in seven patients and bilateral cryptorchidism in five. In contrast to a previous study (Ferlin *et al.*, 2008), the T222P variant was not predominantly present in persistent bilateral cryptorchid patients. In all cases, orchidopexy was performed during infancy. Concerning the three Spanish T222P carriers, whose parents' DNA was available, two of them were cryptorchidic (08-202, sw09-9) and the third one had normal testicular descent (07-276). In all three subjects, the variant was transmitted from the father and all of them had normal testes descent. In the Italian group, DNA from parents was available only for one carrier, code MMP122 (already published in Nuti *et al.*, 2008), and in this case, the variant was transmitted by the mother.

Discussion

The search for clinically relevant genetic factors involved in non-syndromic cryptorchidism has been largely unsuccessful (Giachini *et al.*, 2007; Foresta *et al.*, 2008). Paradoxically, not even the strongest candidate genes such as *INSL3/RXFP2* appear to be involved in the aetiopathogenesis of testicular maldescent. The *INSL3* gene has been the subject of a number of studies in which the entire coding region has been sequenced (for review, see Foresta *et al.*, 2008). In contrast, the *RXFP2* gene has been sequenced only in a small number of subjects (approx. 160 patients; Gorlov *et al.*, 2002). In fact, based on the study by Gorlov *et al.* (2002) which reported for the first time the T222P mutation in exon 8, all subsequent studies have focused on the screening of this specific variant. More than 4000 subjects from different geographical/ethnic origin have been analysed so far (Table 2). After the publication of the first studies, it became evident that this genetic variant is prevalent in the Mediterranean area (only two carriers have been described from Hungary in a later study) and is completely absent from other geographical areas such as Northern Europe and Asia. Previous studies suggested that all (except three) T222P carriers share a common inferred haplotype: C-C-G-A-13 (Ferlin *et al.*, 2003; El Houate *et al.*, 2008; Nuti *et al.*, 2008). This finding leads to the hypothesis about a common ancestor with the C-C-G-A-13 haplotype in the Mediterranean area. The haplotype analysis in our study population confirmed the presence of the common C-C-G-A-13 inferred haplotype in all (except one) Italian and all Spanish carriers (Table S1). The analysis of two Spanish triplets (mother, father and cryptorchid son) and a mother/son pair also supported the hypothesis of a linkage between the T222P variant and the C-C-G-A-13 haplotype in these cases: in one case, the variant is on the C-C-G-A-13 haplotype, whereas in the other, it is on the C-C-G-A-12 haplotype. The latter may also derive from the common C-C-G-A-13 through the contraction of the polyadenine stretch. Contraction and expansion of repeated sequences is a common phenomenon. In the control carrier, the markers are less informative (C-A/C-G-A/G-13) because of the non-availability of the DNA from the father. However, in this subject also, the T222P variant can be potentially linked to the C-C-G-A-13 haplotype.

Table 2 The results of T222P analyses published in the literature in cryptorchid patients and controls from different countries

Geographical origin	Total no.	No. of T222P variants (%)	References
Cryptorchid			
France	20	1 (5)	1
United States	13[a]	0 (0)	3
Mixed origin	41		1
Finland	23	0 (0)	2
Italy (Northern Italy)	1198	29 (2.4)	4, 5, 9
Italy (Central Italy)	**199**[b]	**9 (4.5)**	7
Hungary	70	2 (2.8)	7, cs
Japan	62	0 (0)	6
Morocco	109	3 (2.7)	8
Egypt	53	3 (5.6)	7
Spain	**187**[b]	**3 (1.6)**	7, cs
Control			
France	62	0 (0)	1
United States	31	0 (0)	1, 3
Finland	33	0 (0)	2
Italy (Northern Italy)	850	0 (0)	4, 5, 9
Italy (Central Italy)	**351**[b]	**5 (1.4)**	7, cs
Hungary	140	1 (0.7)	7
Germany	100	0 (0)	1
Japan	60	0 (0)	6
Morocco	250	4 (1.6)	8
Spain	**390**[b]	**7 (1.8)**	7, cs

[a]Patients with familial cryptorchidism.

[b]Includes 77 patients and 48 controls from Spain and 159 patients and 275 controls of Italian origin, from a previous multicentre study (Nuti *et al.*, 2008).

Current study population is shown in bold; cs, current study.

References: 1, Gorlov *et al.*, (2002); 2, Roh *et al.*, (2003); 3, Feng *et al.*, (2004); 4, Ferlin *et al.* (2003); 5, Bogatcheva *et al.*, (2007); 6, Yamazawa *et al.*, (2007); 7, Nuti *et al.*, (2008); ; 8, El Houate *et al.*, (2008); 9, Ferlin *et al.*, (2008).

These data clearly indicate that a founder effect seems to be the most plausible explanation for the geographical restriction of this variant and both Italian and Spanish carriers probably descend from a common ancestor.

Our results further support the relatively high incidence of T222P variant in the two Mediterranean countries. However, although in Spain the frequencies are similar in cases and controls, in Italy, the T222P variant appears to be a mild risk factor for crytorchidism. This later finding contradicts our previous conclusion on the Italian population (Nuti *et al.*, 2008) in which a higher but statistically non-significant ($p = 0.054$) difference in mutation frequencies was observed between cases and controls. The enlarged sample size allowed us to ascertain that this variant, although not causative, can be considered as a mild genetic susceptibility factor to cryptorchidism in Italy (OR = 3.17; 95% CI: 1.07–9.34). Similar to the Spanish population, a lack of association between the T222P variant and cryptorchidism has also been reported in another Mediterranean country, Morocco, where the frequency of T222P was 2.7% in patients and 1.6% in controls (El Houate *et al.*, 2007).

The T222P variant has been transmitted by the father in the three Spanish subjects and this is in contrast to data from Italy. In fact, both in Central and Northern Italy, the variant has been transmitted by the mother in all 10 triplets tested (Bogatcheva *et al.*, 2007; Nuti *et al.*, 2008). Interestingly, also in Morocco, in two of three families, a paternal transmission of the variant was reported (El Houate *et al.*, 2007). All carrier fathers had normal testis descent.

The observed marked geographical differences in allelic frequencies make it very difficult to combine data from different geographical areas. Even a meta-analysis restricted to the Italian studies is biased by the fact that in the Northern Italian studies, the T222P variant was absent in 850 controls, whereas in Central Italy, the frequency reached 1.4%. This controversy cannot be explained by differences in the methodology (both laboratories used direct sequencing) or in the selection criteria of controls (with the exception of 300 newborn subjects included as controls in the study of Ferlin *et al.* (2008); therefore, it remains an unsolved paradox. According to Ferlin *et al.* (2008), no cases with T222P variant (all heterozygous) showed spontaneous descent of the testis and were reported exclusively in association with bilateral cryptorchidism. The observed 7 unilateral carriers and 12 controls in our study population are in sharp contrast to this finding and strongly argue against considering the T222P variant as a pathogenic mutation with clear-cut cause–effect relationship with cryptorchidism.

The highly variable phenotypic expression in heterozygous carriers may reflect the presence of other genetic or environmental factors acting in combination with the T222P variant. Animal models indicate that the phenotypic manifestation of *Insl3/Rxfp2* knockout alleles is susceptible to genetic background with noted delay in testis descent in heterozygotes described in one report of an *Insl3* mutant (Nef & Parada, 1999), but not in another *Insl3* mutant (Zimmermann *et al.*, 1999) and in the *Rxfp2* (Gorlov *et al.*, 2002). Different environmental conditions, especially exposure to xeno-oestrogens, may also influence the phenotypic expression of heterozygote carriers. In this regard, it has been shown that exposure to mono-*N*-butyl phthalate, an endocrine disruptor with weak oestrogenic activities, impairs *Insl3* gene expression and affects testicular descent in rats (Wilson *et al.*, 2004; Shono *et al.*, 2005). Only heterozygous carriers have been identified so far; it remains therefore an open question whether homozygosity is strictly associated with cryptorchidism also in humans.

The different scenario in the two countries may be related to the accumulation of different genetic or environmental modifiers leading to a more mild or wild-type phenotype in heterozygotes of Spanish descent and to a more frequently observed pathological condition in Italy. This hypothesis is supported by a recent study which provides evidence about relevant genetic differences between populations belonging to different geographical areas within Europe (Novembre *et al.*, 2008).

As cryptorchidism is a significant risk factor for impaired sperm production and testis cancer, the identification of transmissible genetic factors is of considerable interest. We provide evidence that the T222P variant does not have any pathogenic effect on cryptorchidism in the Spanish population. Although in Italy it seems to confer a mild risk to cryptorchidism, the screening for this variant for diagnostic purposes is not indicated because of the relatively high frequency of control carriers (1.4% of Italian men without a history of cryptorchidism).

Acknowledgements

This work was supported by a grant from the Spanish Health Ministry (FIS-08/1138). The authors thank the patients for participating in this study, and the president of the Fundacio Puigvert (Esperança Martí i Salís) for her continuous support to this project. The authors are grateful to Dr J. Galan for his precious advice regarding data analysis and interpretation.

References

Adham IM & Agoulnik AI. (2004) Insulin-like 3 signalling in testicular descent. *Int J Androl* 27, 257–265.

Bogatcheva NV, Truong A, Feng S, Engel W, Adham IM & Agoulnik AI. (2003) GREAT/LGR8 is the only receptor for insulin-like 3 peptide. *Mol Endocrinol* 17, 2639–2646.

Bogatcheva NV, Ferlin A, Feng S, Truong A, Gianesello L, Foresta C & Agoulnik AI. (2007) T222P mutation of the insulin-like 3 hormone receptor LGR8 is associated with testicular maldescent and hinders receptor expression on the cell surface membrane. *Am J Physiol Endocrinol Metab* 292, E138–E144.

El Houate B, Rouba H, Sibai H, Barakat A, Chafik A, Chadli el B *et al.* (2007) Novel mutations involving the INSL3 gene associated with cryptorchidism. *J Urol* 177, 1947–1951.

El Houate B, Rouba H, Imken L, Sibai H, Chafik A, Boulouiz R, Chadli E, Hassar M, McElreavey K & Barakat A. (2008) No association between T222P/LGR8 mutation and cryptorchidism in the Moroccan population. *Horm Res* 70, 236–239.

Feng S, Cortessis VK, Hwang A, Hardy B, Koh CJ, Bogatcheva NV & Agoulnik AI. (2004) Mutation analysis of *INSL3* and *GREAT/LGR8* genes in familial cryptorchidism. *Urology* 64, 1032–1036.

Ferlin A, Simonato M, Bartoloni L, Rizzo G, Bettella A, Dottorini T, Dallapiccola B & Foresta C. (2003) The INSL3-LGR8/GREAT ligand–receptor pair in human cryptorchidism. *J Clin Endocrinol Metab* 88, 4273–4279.

Ferlin A, Arredi B, Zuccarello D, Garolla A, Selice R & Foresta C. (2006) Paracrine and endocrine roles of insulin-like factor 3. *J Endocrinol Invest* 29, 657–664.

Ferlin A, Zuccarello D, Zuccarello B, Chirico MR, Zanon GF & Foresta C. (2008) Genetic alterations associated with cryptorchidism. *JAMA* 300, 2271–2276.

Foresta C, Zuccarello D, Garolla A & Ferlin A. (2008) Role of hormones, genes, and environment in human cryptorchidism. *Endocr Rev* 29, 560–580.

Giachini C, Nuti F, Marinari E, Forti G & Krausz C. (2007) Partial AZFc deletions in infertile men with cryptorchidism. *Hum Reprod* 22, 2398–2403.

Gorlov IP, Kamat A, Bogatcheva NV, Jones E, Lamb DJ, Truong A, Bishop CE, McElreavey K & Agoulnik AI. (2002) Mutations of the GREAT gene cause cryptorchidism. *Hum Mol Genet* 11, 2309–2318.

Ivell R & Hartung S. (2003) The molecular basis of cryptorchidism. *Mol Hum Reprod* 9, 175–181.

Krausz C, Quintana-Murci L, Fellous M, Siffroi JP & McElreavey K. (2000) Absence of mutations involving the INSL3 gene in human idiopathic cryptorchidism. *Mol Hum Reprod* 6, 298–302.

Nef S & Parada LF. (1999) Cryptorchidism in mice mutant for Insl3. *Nat Genet* 22, 295–299.

Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A *et al.* (2008) Genes mirror geography within Europe. *Nature* 456, 98–101.

Nuti F, Marinari E, Erdei E, El-Hamshari M, Echavarria MG, Ars E *et al.* (2008) The leucine-rich repeat-containing G protein-coupled receptor 8 gene T222P mutation does not cause cryptorchidism. *J Clin Endocrinol Metab* 93, 1072–1076.

Roh J, Virtanen H, Kumagai J, Sudo S, Kaleva M, Toppari J & Hsueh AJ. (2003) Lack of *LGR8* gene mutation in Finnish patients with a family history of cryptorchidism. *Reprod Biomed Online* 7, 400–406.

Sasieni PD. (1997) From genotypes to genes: doubling the sample size. *Biometrics* 53, 1253–1261.

Shono T, Shima Y, Kondo T & Suita S. (2005) *In utero* exposure to mono-*n*-butyl phthalate impairs insulin-like factor 3 gene expression and the transabdominal phase of testicular descent in fetal rats. *J Pediatr Surg* 40, 1861–1864.

Toppari J, Virtanen H, Skakkebaek NE & Main KM. (2006) Environmental effects on hormonal regulation of testicular descent. *J Steroid Biochem Mol Biol* 102, 184–186.

Virtanen HE & Toppari J. (2008) Epidemiology and pathogenesis of cryptorchidism. *Hum Reprod Update* 14, 49–58.

Wilson VS, Lambright C, Furr J, Ostby J, Wood C, Held G & Gray LE Jr. (2004) Phthalate ester-induced gubernacular lesions are associated with reduced insl3 testis gene expression in the fetal rat testis. *Toxicol Lett* 146, 207–215.

Yamazawa K, Wada Y, Sasagawa I, Aoki K, Ueoka K & Ogata T. (2007) Mutation and polymorphism analyses of *INSL3* and *GREAT/LGR8* in 62 Japanese patients with cryptorchidism. *Horm Res* 67, 73–76.

Zimmermann S, Steding G, Emmen JM, Brinkmann AO, Nayernia K, Holstein AF, Engel W & Adham IM. (1999) Targeted disruption of the Insl3 gene causes bilateral cryptorchidism. *Mol Endocrinol* 13, 681–691.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Haplotype analysis of *RXFP2* gene based on five intronic (introns 7 and 9) and exonic (exon 12) polymorphisms with all possible inferred haplotypes. The common inferred haplotype C-C-G-A-13 is indicated in bold.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

SUPPLEMENTARY TABLE 1. Haplotype analysis of *RXFP2* gene based on five intronic (intron 7 and 9) and exonic (exon 12) polymorphisms with all possible inferred haplotypes. The common inferred haplotype C-C-G-A-13 is indicated in bold.

Code	Polymorphisms				Possible haplotypes
	INV7 (-351) A>C	INV9 (+380) A>C	Exon 12 SNPs 957G>A 993A>G	Intron 12 INV12(-2)A_9 >A_{13}	
Spanish					
Cryptorchid					
07-413	CC	AC	GG GA	10/13	CCGG10 / CAGG10 / CCGA10 / CAGA10 CCGG13 / CAGG13 / **CCGA13** / CAGA13
08-202	AC	CC	GA AA	10/13	ACAA10 / CCAA10 / ACGA10 / CCGA10 ACAA13 / CCAA13 / ACGA13 / **CCGA13**
sw 09-9	CC	CC	GG AA	12/13	CCGA12 / **CCGA13**
Control					
05-143	AC	CC	GA AA	10/13	ACAA10 / CCAA10 / ACGA10 / CCGA10 ACAA13 / CCAA13 / ACGA13 / **CCGA13**
07-22	AC	AC	GG AA	12/13	AAGA12 / ACGA12 / CAGA12 / CCGA12 AAGA13 / ACGA13 / CAGA13 / **CCGA13**
FQ 01-388	AC	CC	GA AA	10/13	ACAA10 / CCAA10 / ACGA10 / CCGA10 ACAA13 / CCAA13 / ACGA13 / **CCGA13**
FQ 01-397	AC	AC	GG AA	12/13	AAGA12 / ACGA12 / CAGA12 / CCGA12 AAGA13 / ACGA13 / CAGA13 / **CCGA13**
07-276	CC	AC	GG GA	11/13	CCGG11 / CAGG11 / CCGA11/ CAGA11 CCGG13 / CAGG13 / **CCGA13** / CAGA13
Italian					
Cryptorchid					
MMP122*	AC	AC	GG GA	12/13	AAGA12 / ACGA12 / CAGA12 / CCGA12 AAGA13 / ACGA13 / CAGA13 / **CCGA13**
MMP299*	AC	CC	GG AA	10/13	ACGA10 / CCGA10 / ACGA13 **/ CCGA13**
MMP475*	AC	AC	GG AA	12/13	AAGA12 / ACGA12 / CAGA12 / CCGA12 AAGA13 / ACGA13 / CAGA13 / **CCGA13**
A191*	AC	CC	GG AA	10/13	ACGA10 **/** CCGA10 / ACGA13 / **CCGA13**
A251*	CC	CC	GA AA	12/13	CCGA12 / CCAA12 / **CCGA13 /** CCAA13
A639*	AC	CC	GG AA	10/13	ACGA10 / CCGA10 / ACGA13 / **CCGA13**
CH4*	AC	CC	GA AA	10/13	ACGA10 / CCGA10 / ACAA10 / CCAA10 ACGA13 / **CCGA13** / ACAA13 / CCAA13
CH16*	AC	AC	GG AA	12/13	AAGA12 / ACGA12 / CAGA12 / CCGA12 AAGA13 / ACGA13 / CAGA13 / **CCGA13**
CH25	AC	AC	GG AA	12/13	AAGA12 / AAGA13 / CAGA12 / CAGA13 ACGA12 / ACGA13 / CCGA12 /**CCGA13**
Control					
A268*	AC	CC	GG AA	11/13	ACGA11 / CCGA11 / ACGA13 / **CCGA13**
A337*	AC	CC	GG AA	10/13	ACGA10 / CCGA10 / ACGA13 / **CCGA13**
A448*	CC	AA	GG AA	10/13	CAGA10 / CAGA13
CS32*	CC	AC	GG GA	11/13	CAGG11 / CCGG11 / CAGA11 / CCGA11 CAGG13 / CCGG13 / CAGA13 / **CCGA13**
C68*	AC	AC	GA AA	10/13	AAGA10 / CAGA10 / ACGA10 / CCGA10 AAAA10 / CAAA10 / ACAA10 / CCAA10 AAGA13 / CAGA13 / ACGA13 / **CCGA13** AAAA13 / CAAA13 / ACAA13 / CCAA13

* carriers already haplotyped in the previous multicenter study.

PAPER 2

ESR1 promoter polymorphism is not associated with non-syndromic cryptorchidism.

Lo Giacco D., Ars E, Bassas L., Galan J.J., Rajmil O., Ruíz P., Caffaratti J., Guarducci E., Ruiz-Castané E., Krausz C.
Fertility & Sterility. 2011 95:369-71, (PMID: 20887985)

SUMMARY 2

Background:
A proportion of cases of non-syndromic cryptorchidism may derive from the action of environmental factors (such as the Endocrine Disruptors) on a predisposing genetic background. The *ESR1* polymorphisms are supposed to contribute to the individual genetic susceptibility to the environment. Among them the *ESR1* promoter microsatellite (TA)n is the most interesting. It was previously reported as a potential functional polymorphism in relation to bone mineralization and spermatogenesis, whereas its role in non-syndromic cryptorchidism has been never investigated so far.

Aim:
Given the relationship between cryptorchidism and low sperm count in the context of testicular dysgenesis syndrome, our aim was to establish whether the (TA)n polymorphism may also influence testis descent.

Materials & Methods:
A total of 569 subjects from Spain (180 cryptorchid patients 389 controls) and 544 subjects from Italy (193 cryptorchid patients and 351 controls) were analyzed for the $(TA)_n$ polymorphism of *ESR1*. The fragment containing the microsatellite $(TA)_n$ was amplified by PCR using fluorescent primers and the size of PCR products was determined by capillary electrophoresis using GeneScan software. Selected PCR products with different sizes were sequenced for the definition of the exact TA repeat length and used as reference samples. The number of TA repeats was calculated comparing the detected PCR fragment by GeneScan software with the reference fragments. We established three allele groups: group H, including alleles with a high number of TA repeats (TA ≥20); group M, including alleles with a medium number of TA repeats (15< TA <20); and group L, including low TA alleles (TA ≤15). Since genotypes HH and HM (referred to as genotype A) showed the lowest sperm count compared to the other allelic combinations (referred to as genotype B) we considered them as the functionally most active genetic variants.

Results:
In both study populations, the distribution of genotypes A and B was similar in patients and controls, showing a higher frequency of genotype B in both phenotypic groups. Therefore no association between specific (TA)n genotypes and non-syndromic cryptorchidism was found neither in the Spanish (*p* value=0.91; OR=0.984 95% CI=0.729-1,329) nor in the Italian (*p* value=0.41; OR=1.146; 95% CI=0.828-1.585) study populations.

Conclusion:
Our findings excluded an association between the TA repeat genotype A and cryptorchidism in both Mediterranean populations. Further investigations are needed to establish whether under specific conditions, such as a higher environmental exposure, this polymorphism may contribute to the disease.

ESR1 promoter polymorphism is not associated with nonsyndromic cryptorchidism

The *ESR1* promoter microsatellite $(TA)_n$ was reported as a potential functional polymorphism. In a case–control study, we were unable to demonstrate any association between $(TA)_n$ and nonsyndromic cryptorchidism in Italian and Spanish study populations. (Fertil Steril® 2011;95:369–71. ©2011 by American Society for Reproductive Medicine.)

Key Words: Cryptorchidism, genetics, endocrine disrupters, polymorphism, *ESR1*, infertility

Cryptorchidism is the most frequent congenital birth defect in male children, and its etiology, in most cases, remains unknown (1). Two hormone/receptor systems play a major role: [1] testosterone and its receptor, and [2] insulin-like factor 3 (INSL3) and its G-protein–coupled receptor, relaxin family peptide 2 (RXFP2). Mutations in the androgen receptor (*AR*) gene are associated prevalently with syndromic forms of cryptorchidism (2), whereas polymorphisms in the first exon (particular combinations of CAG/GGC repeats) have been reported as potential risk factors for nonsyndromic cryptorchidism (3–5). The identification of environmental factors able to interfere with steroid receptors (endocrine disrupters) and to affect the development of the male urogenital tract increased the interest not only in polymorphisms of the *AR* but also of the estrogen receptor (*ESR1*) genes (6). Taking into consideration that xenoestrogens down-regulate *INSL3* gene expression in Leydig cells through an *ESR1*-dependent mechanism (7–9), it is plausible that *ESR1* polymorphisms may also be involved in the interaction with endocrine disrupters and thus in the etiology of cryptorchidism.

Genetic screening of the human *ESR1* gene locus has revealed the existence of several polymorphic sites (10–14). Recently, a haplotype called AGATA and its tag single nucleotide polymorphism (SNP), SNP12, was proposed as being of potential clinical importance, and it was first reported as a significant risk factor for cryptorchidism in the Japanese population (15). However, data on the same polymorphism in the Italian population (16) gave contradictory results, showing a higher incidence of SNP12 in controls and thus defining it as a "protective" factor against cryptorchidism.

Although the literature on *ESR1* polymorphisms and cryptorchidism is limited to the AGATA haplotype (SNP12), *ESR1* polymorphisms have been extensively analyzed in relationship to male infertility/spermatogenic impairment or hypospadias (15–20). Among them the $(TA)_n$ variable number of tandem repeat polymorphism within the promoter region of *ESR1* gene seems to be the most promising because its position in the promoter implies a potential regulatory function. It has been reported that the number of TA repeats is significantly associated with bone mineral density in postmenopausal women (10, 21) and with spermatogenic efficiency in normozoospermic men (17). The role of this polymorphism in testis descent is unknown. Given the relationship between cryptorchidism and low sperm count in the context of testicular dysgenesis syndrome (22), our aim was to establish whether the same genotype may also influence testis descent.

A total of 569 subjects from Spain (ascertained Spanish origin) and 544 subjects from Italy (ascertained Italian origin) were analysed for the $(TA)_n$ polymorphism of *ESR1*. The Spanish group included 180 patients with a history of nonsyndromic cryptorchidism, whereas the remaining 389 were controls with no history of testicular maldescent at birth. In the cryptorchid group 112 men had unilateral and 68 bilateral cryptorchidism.

Deborah Lo Giacco, M.Sc.[a,b]
Elisabet Ars, Ph.D.[b]
Lluis Bassas, M.D., Ph.D.[a]
José Jorge Galan, Ph.D.[c]
Osvaldo Rajmil, M.D., Ph.D.[a]
Patricia Ruíz[b]
Jorge Caffaratti, M.D.[d]
Elena Guarducci, M.Sc.[e]
Eduard Ruiz-Castané, M.D.[a]
Csilla Krausz, M.D., Ph.D.[a,e]

[a] *Andrology Service, Fundació Puigvert, Universitat Autònoma de Barcelona, Barcelona, Spain*
[b] *Molecular Biology Laboratory, Fundació Puigvert, Universitat Autònoma de Barcelona, Barcelona, Spain*
[c] *Department of Structural Genomics, Neocodex S.L., Seville, Spain*
[d] *Pediatric Urology Division, Urology Service, Fundació Puigvert, Universitat Autònoma de Barcelona, Barcelona, Spain*
[e] *Andrology Unit, Department of Clinical Physiopathology, University of Florence, Florence, Italy*

Received June 5, 2010; revised August 2, 2010; accepted August 20, 2010; published online October 2, 2010.
D.L.G. has nothing to disclose. E.A. has nothing to disclose. L.B. has nothing to disclose. J.J.G. has nothing to disclose. O.R. has nothing to disclose. P.R. has nothing to disclose. J.C. has nothing to disclose. E.G. has nothing to disclose. E.R-C. has nothing to disclose. C.K. has nothing to disclose.
Supported by a grant from the Spanish Health Ministry (grant FIS-08/1138).
Reprint requests: Csilla Krausz, M.D., Ph.D., Andrology Unit, Viale Pieraccini 6, 50139 Florence, Italy (E-mail: c.krausz_@dfc.unifi.it).

TABLE 1

Frequencies of genotype A and genotype B and statistical analyses in patients with a history of testicular maldescent and in controls with no history of cryptorchidism from the Spanish and Italian study populations and from the combined group.

Phenotype	Genotype A, n (%)	Genotype B, n (%)	*P* value [odds ratio (95% confidence interval)]
Spanish (n = 569)	147 (25.8)	422 (74.1)	.91 [0.984 (0.729–1.329)]
Cryptorchid patients (n = 180)	46 (25.5)	134 (74.4)	
Controls (n = 389)	101 (25.9)	288 (74)	
Italian (n = 544)	119 (21.8)	425 (78.1)	.41 [1.146 (0.828–1.585)]
Cryptorchid patients (n = 193)	46 (23.8)	147 (76.2)	
Controls (n = 351)	73 (20.7)	278 (79.2)	
Combined data (n = 1,113)	266 (23.9)	847 (76.1)	.67 [1.049 (0.842–1.307)]
Cryptorchid patients (n = 373)	92 (24.7)	281 (75.3)	
Controls (n = 740)	174 (23.5)	566 (76.5)	

Lo Giacco. Correspondence. Fertil Steril 2011.

The Italian study population (n = 544) showed a similar distribution of cases (193 patients with history of cryptorchidism) and controls (n = 351). In the cryptorchid group 129 subjects had a history of unilateral and 64 bilateral cryptorchidism.

Written informed consent was obtained from all of the patients and controls included in the present study. Institutional board approval was obtained from the local ethics committees of both institutions (Fundació Puigvert and University of Florence).

Deoxyribonucleic acid was extracted from peripheral blood in all the participants in this study with the exception of the Italian controls, for whom, depending on sample availability, the genomic DNA was isolated either from peripheral blood or from frozen semen. The TA repeat located in the promoter element of the *ESR1* gene (RefSeq NC_000006.11) was amplified by polymerase chain reaction (PCR). The sequence of primers was as follows: forward 5′-GACGCATGATATACTTCACC-3′ and reverse 5′-GCAGAAT CAAATATCCAGATG-3′. The TA repeat length analysis was performed with an automated sequencer (ABI PRISM 3100-Avant; Applied Biosystems, Foster City, CA). The size of the PCR products was determined by GeneScan software (Applied Biosystems). Selected PCR products with different sizes were subjected to direct sequencing on the autosequencer for the definition of the correct TA repeat length. The number of TA repeats was calculated comparing the detected PCR fragment by GeneScan software with the sequenced fragments.

For the statistical analysis of genotype distribution, two-point association studies, and joint analyses we used Statistical Package for the Social Sciences software (SPSS, Chicago, IL). Statistical power calculations were performed by using Episheet software (spreadsheets for the analysis of epidemiologic data; available at www.epidemiolog.net/studymat/). Assuming the exposure prevalence for genotype A (0.237) previously reported by our group (17), we estimated having enough statistical power to detect an odds ratio above 1.73.

First we analysed a total of 1,138 alleles from a study population of Spanish origin. The allelic frequency distribution in cases and controls (Supplemental Fig. 1) shows a similar pattern to that observed in other studies dealing with populations from Europe (10, 17, 20). The median number (range) of the TA repeat was 17 (9–26) and 16 (9–28) for the cryptorchid and control groups, respectively. Although bilateral cases had a higher median value of $(TA)_n$ (19, with a range of 13–26) with respect to the unilateral group (16, with a range of 9–26), the difference was not statistically significant.

The distribution pattern of TA alleles in this study population shows two major peaks at 13–15 and 20–24 repeats and a lower distribution of intermediate 16–19 repeat alleles. According to previous studies (17, 21), we established three allele groups: [1] group H, including alleles with a high number of TA repeats (TA $\geq$20); [2] group M, including alleles with a medium number of TA repeats (15< TA <20); and [3] group L, including low TA alleles (TA $\leq$15).

The combination of these alleles gave rise to six different genotypes (HH, HM, HL, MM, LM, and LL). Because genotypes HH and HM showed the highest bone mineral density with respect to the other allelic combinations (21) and the lowest sperm count (17), we considered them as the functionally more active genetic variants. Consequently, we grouped the six possible allelic combinations into two major groups: "genotype A" (including HH and HM genotypes) and "genotype B" (including genotypes HL, MM, LM, and LL). To evaluate whether genotype A of the *ESR1* is associated with cryptorchidism, we compared their frequencies between the two study groups. The distribution of genotypes A and B was similar in the cryptorchid and in the control groups, showing a higher frequency of genotype B in both groups (74.4% vs. 74%, respectively) (Table 1).

We analyzed an independent series of cases and controls of Italian origin (a total of 1,088 alleles). The allelic distribution was similar to that observed in the Spanish population (Supplemental Fig. 1). The median number (range) of the TA repeat was 15 (11–26) and 15 (7–26) for the cryptorchid and control groups, respectively. The allelic distribution of patients with unilateral and bilateral cryptorchidism was similar, with median values of 16 (11–26) and 15 (12–26), respectively. Comparison of the frequencies of genotypes A and B revealed no significant differences between patients and controls; additionally, in this study population there was a higher frequency of genotype B in both groups (76.2% vs. 79.2%, respectively) (Table 1).

Given the similar allelic distribution in the two study populations and their common geographic origin (both Mediterranean countries), we performed a joint analysis on the combined group on a total of 2,226 $(TA)_n$ alleles (373 patients and 740 controls), taking into account the geographic origin in each sample as covariable. No association between specific $(TA)_n$ genotypes and nonsyndromic cryptorchidism was found (Table 1 and Supplemental Table 1). Because data on the position of testes at the time of orchidopexy were not available for the majority of cases, we were unable to evaluate the effect of $(TA)_n$, if any, on the severity of cryptorchidism expressed in terms of position of the gonad(s).

Case–control association studies are subjected to false-positive results, especially when the sample size and the selection of cases and controls (including ethnicity) are not appropriate (23). The lack of consistency between different studies might therefore represent spurious results or chance findings or a real difference between populations (i.e., different genes contributing to disease risk). Because most of the associations are not replicated by subsequent studies, the best way to validate the findings from a genetic association study is to obtain the same results by analyzing other independent series of cases and controls at the same time (24). Given that the $(TA)_n$ polymorphism has not been previously studied in relation to cryptorchidism, we sought to provide highly reliable data by screening two independent study populations from the Mediterranean area. To minimize all potential biases we studied a large number (569 and 544 from Spain and Italy, respectively) of carefully selected subjects. Much care was taken for the clinical selection and for geographic/ethnic matching of cases and controls. The analysis of a total of 2,226 alleles showed a similar allelic distribution of $(TA)_n$ in cases and controls in both study populations. The frequency of genotype A, with a predicted stronger estrogen effect, was also similar in cases and controls both in the Italian and Spanish study populations separately and jointly.

In conclusion, our finding excludes a clinically relevant role for the TA repeat genotype A in the etiopathogenesis of cryptorchidism. However, it remains to be established whether under specific environmental conditions, such as exposure to high doses of endocrine disrupters, this polymorphism may exert a pathogenic effect.

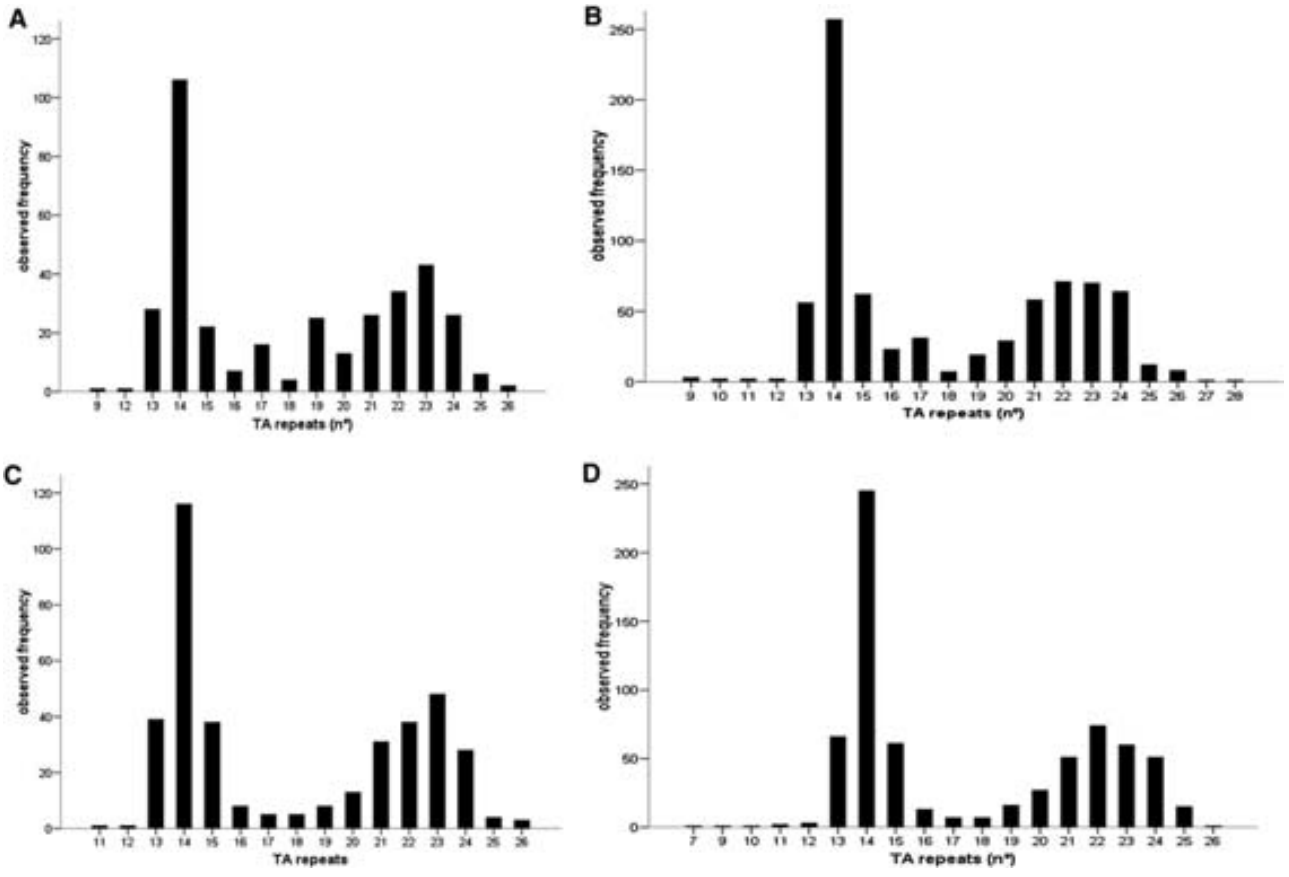
Acknowledgments: The authors thank the clinical staff of the Andrology Service of Fundacio Puigvert (Drs. A. Vives, J. Sarquella, C. Rojas, S. Curvelo, and M.G. Echavarria); Dr A. Bujons (Pediatric Urology Service); and the president of the Fundació Puigvert (Esperança Martí i Salís) for her continuous support of this project.

REFERENCES

1. Toppari J, Virtanen H, Skakkebaek NE, Main KM. Environmental effects on hormonal regulation of testicular descent. Steroid Biochem Mol Biol 2006;102:184–6.
2. Gottlieb B, Beitel L, Wu JH, Trifiro M. The androgen receptor gene mutations database (ARDB). Hum Mutat 2004;23:527–33.
3. Ferlin A, Garolla A, Bettella A, Bartoloni L, Vinanzi C, Roverato A, et al. Androgen receptor gene CAG and GGC repeat lengths in cryptorchidism. Eur J Endocrinol 2005;152:419–25.
4. Radpour R, Rezaee M, Tavasoly A, Solati S, Saleki A. Association of long polyglycine tracts (GGN repeats) in exon 1 of the androgen receptor gene with cryptorchidism and penile hypospadias in Iranian patients. J Androl 2007;28:164–9.
5. Silva-Ramos M, Oliveira JM, Cabeda JM, Reis A, Soares J, et al. The CAG repeat within the androgen receptor gene and its relationship to cryptorchidism. Int Braz J Urol 2006;32:330–4.
6. Damgaard IN, Skakkebaek NE, Toppari J, Virtanen HE, Shen H, et al. Persistent pesticides in human breast milk and cryptorchidism. Environ Health Perspect 2006;114:1133–8.
7. Wilson VS, Lambright C, Furr J, Ostby J, Wood C, Held G, et al. Phthalate ester-induced gubernacular lesions are associated with reduced insl3 testis gene expression in the fetal rat testis. Toxicol Lett 2004;146:207–15.
8. Shono T, Shima Y, Kondo T, Suita S. In utero exposure to mono-n-butyl phthalate impairs insulin-like factor 3 gene expression and the transabdominal phase of testicular descent in fetal rats. J Pediatr Surg 2005;40:1861–4.
9. Cederroth CR, Schaad O, Descombes P, Chambon P, Vassalli JD, Nef S. Estrogen receptor alpha is a major contributor to estrogen-mediated fetal dysgenesis and cryptorchidism. Endocrinology 2007;148:5507–19.
10. Gennari L, Merlotti D, De Paola V, Calabro A, Becherini L, Martini G, et al. Estrogen receptor gene polymorphisms and the genetics of osteoporosis: a HuGE review. Am J Epidemiol 2005;161:307–20.
11. Castagnoli A, Maestri I, Bernardi F, Del Senno L. *Pvu*II RFLP incide the human estrogen receptor gene. Nucleic Acids Res 1987;15:866.
12. Coleman RT, Taylor JE, Shine JJ, Frossard PM. Human estrogen receptor (*ESR*) gene locus: *Pss*I dimorphism. Nucleic Acids Res 1988;16:7208.
13. Zuppan PJ, Hall JM, Ponglikitmongkol M, et al. Polymorphisms at the estrogen receptor (*ESR*) locus and linkage relationships on chromosome 6q. Cytogenet Cell Genet 1989;51:1116.
14. Del Senno L, Aguiari GL, Piva R. Dinucleotide repeat polymorphism in the human estrogen receptor (*ESR*) gene. Hum Mol Genet 1992;1:354.
15. Yoshida R, Fukami M, Sasagawa I, Hasegawa T, Kamatani N, Ogata T. Association of cryptorchidism with a specific haplotype of the estrogen receptor alpha gene: implication for the susceptibility to estrogenic environmental endocrine disruptors. J Clin Endocrinol Metab 2005;90:4716–21.
16. Galan JJ, Guarducci E, Nuti F, Gonzalez A, Ruiz M, Ruiz A, et al. Molecular analysis of estrogen receptor alpha gene AGATA haplotype and SNP12 in European populations: potential protective effect for cryptorchidism and lack of association with male infertility. Hum Reprod 2007;22:444–9.
17. Guarducci E, Nuti F, Becherini L, Rotondi M, Balercia G, Forti G, et al. Estrogen receptor alpha promoter polymorphism: stronger estrogen action is coupled with lower sperm count. Hum Reprod 2006;21:994–1001.
18. Galan JJ, Buch B, Pedrinaci S, Jimenez-Gamiz P, Gonzalez A, Serrano-Rios M, et al. Identification of a 2244 base pair interstitial deletion within the human *ESR1* gene in the Spanish population. J Med Genet 2008;45:420–4.
19. Watanabe M, Yoshida R, Ueoka K, Aoki K, Sasagawa I, Hasegawa T, et al. Haplotype analysis of the estrogen receptor 1 gene in male genital and reproductive abnormalities. Hum Reprod 2007;22:1279–84.
20. Galan JJ, Buch B, Cruz N, Segura A, Moron FJ, Bassas L, et al. Multilocus analyses of estrogen-related genes reveal involvement of the *ESR1* gene in male infertility and the polygenic nature of the pathology. Fertil Steril 2005;84:910–8.
21. Becherini L, Gennari L, Masi L, Mansani R, Massart F, Morelli A, et al. Evidence of a linkage disequilibrium between polymorphisms in the human estrogen receptor alpha gene and their relationship to bone mass variation in postmenopausal Italian women. Hum Mol Genet 2000;9:2043–50.
22. Skakkebaek NE, Rajpert-De Meyts E, Main KM. Testicular dysgenesis syndrome: an increasingly common developmental disorder with environmental aspects. Hum Reprod 2001;16:972–8.
23. Nuti F, Krausz C. Gene polymorphisms/mutations relevant to abnormal spermatogenesis. Reprod Biomed Online 2008;16:504–13.
24. Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. Genet Med 2002;4:45–61.

SUPPLEMENTAL FIGURE 1

Frequency distribution of the *ESR1* gene dinucleotide repeat polymorphism in the Spanish and Italian study populations. (**A**) Spanish cryptorchid patients; (**B**) Spanish controls; (**C**) Italian cryptorchid patients; (**D**) Italian controls.



Lo Giacco. Correspondence. Fertil Steril 2011.

SUPPLEMENTAL TABLE 1

Joint analysis of genotype A in the combined Spanish and Italian populations.

Variable	*P* value [odds ratio (95% confidence interval)]
Geographic origin	.168 [0.839 (0.654–1.077)]
Genotype A	.625 [0.930 (0.695–1.244)]

Lo Giacco. Correspondence. Fertil Steril 2011.

PAPER 3

Clinical relevance of Y-linked CNV screening in male infertility: new insights based on the 8-year experience of a diagnostic genetic laboratory.

Lo Giacco D., Chianese C., Sánchez-Curbelo J., Bassas L., Ruiz P., Rajmil O., Sarquella J., Vives A., Ruiz-Castané E., Oliva R., Ars E. and Krausz C.
European Journal of Human Genetics.2013 (Under review).

SUMMARY 3

Background:
Y-chromosome microdeletions are a well-established genetic cause of severe spermatogenic failure and their molecular diagnosis is part of the diagnostic work-up of severe male factor infertility; however, some few issues remain debated. Several recurrent partial deletions (named *gr/gr*, *b2/b3* and *b1/b3*) and duplications (*b2/b4* duplication) have been reported in the *AZFc* region, in addition to the classical *AZFc* deletion. Among them only the *gr/gr* deletion resulted to be clinically relevant. In the Spanish population, partial *AZFc* rearrangements have been poorly explored and no data exist on partial duplications.

Aims:
We aimed to provide new insights on a number of still debated issues concerning the routine genetic testing for Y microdeletions through a comprehensive phenotypic and genotypic description of consecutive infertile men screened for Y microdeletions.
As for partial *AZFc* rearrangements, the aims of the study were to: i) corroborate the role of *gr/gr* deletion as significant risk factor for spermatogenic impairment; ii) study the relationship of Y hgrs with the phenotypic expression of *gr/gr* deletions; iii) provide information about the role of *b2/b3* deletions and iv) to provide for the first time data on the prevalence and clinical significance of *AZFc* partial duplications the in Spanish population.

Methods:
A total of 806 mainly Spanish consecutive infertile men referring to the genetic laboratory of the clinic were screened for Y microdeletions according to the EAA/ EMQN guidelines.
The analysis of *AZFc* partial rearrangements included 330 idiopathic infertile patients and 385 controls of Spanish origin. Partial *AZFc* deletions and duplications were detected by STS+/- PCR and confirmed by gene dosage. Restriction Fragment Length Polymorphism analysis was used for molecular characterization of partial deletions.

Results:
In our study, 27/806 (3.3%) patients carried complete *AZF* deletions. All were azoo/cryptozoospermic, except for one whose sperm concentration was 1-2x10^6/ml. In *AZFc* deleted men, we observed a lower sperm recovery rate upon TESE (9.1%) compared to the literature (60%-80% with microTESE and multiple TESE). Hgr E was the most represented among non-Spanish whereas hgr P among Spanish *AZF* deletion carriers.
Gr/gr deletion was significantly associated with spermatogenic impairment ($P<0.05$; OR=2.8; 95% CI=1.017-8.007). The majority of *gr/gr* deletion carriers belonged to the P,Q,R branches in both patients and controls. Our data together with a previous pilot study further confirms the *gr/gr* deletion as a significant risk factor in the Spanish population (OR=4.8; 95% CI=1.863-12.623; $P<0.001$). *b2/b3* deletion was found exclusively in patients mostly on hgrs C,F,G,H,I. No statistically significance difference was found in frequency of partial *AZFc* duplication between patients and controls. Accordingly with the results of case-control study, *DAZ* gene dosage analysis revealed that a reduced *DAZ* copy number (2 copies) is associated with a significant reduction in semen quality in terms of sperm concentration and total motile sperm count, whereas an increase (6 or 8 copies) in respect to the most common *DAZ* gene dosage (4 copies) does not affect significantly semen quality.

Conclusions:
Our data integrated with the literature suggest that: i) routine *AZF* microdeletion testing could

eventually include only men with ≤2x10^6/ml; ii) classical TESE is associated with low sperm recovery rate in azoospermic *AZFc* deleted men, therefore microTESE should be regarded as the best option for sperm retrieval in *AZF* deletion carriers; iii) Y background could partially explain the differences in deletion frequencies among populations. Our data confirm *gr/gr* as significant risk factor for spermatogenic impairment in the Spanish population. The clustering of *gr/gr* deletion carriers in P,Q,R hgrs further suggests that the phenotypic variability of *gr/gr* deletion is independent of Y-chromosomal background in Europeans. Conversely, Y background seems to modulate a potential deleterious effect of *b2/b3* on spermatogenesis.
Partial *AZFc* duplications do not represent a risk for spermatogenic failure in the Spanish population.

Clinical relevance of Y-linked CNV screening in male infertility: new insights based on the 8-year experience of a diagnostic genetic laboratory.

Running title**:** Y-linked CNVs in male infertility

Authors:
Deborah Lo Giacco[1, 2], Chiara Chianese[3], Josvany Sánchez-Curbelo[2], Lluis Bassas[2], Patricia Ruiz[1], Osvaldo Rajmil[2], Joaquim Sarquella[2], Alvaro Vives[2]**,** Eduard Ruiz-Castañé[2]**,** Rafael Oliva[4]**,** Elisabet Ars[* 1]and Csilla Krausz[* 2, 3]

* These authors contributed equally to the study

Affiliations:
1. Molecular Biology Laboratory, Fundació Puigvert Universitat Autònoma de Barcelona, Barcelona, Instituto de Investigaciones Biomédicas Sant Pau (IIB-Sant Pau), Barcelona, Spain
2. Andrology Service, Fundació Puigvert, Universitat Autònoma de Barcelona, Instituto de Investigaciones Biomédicas Sant Pau (IIB-Sant Pau), Barcelona, Spain
3. Department of Experimental and Clinical Biomedical Sciences, University of Florence, Italy
4. Human Genetics Laboratory, Genetics Unit, Faculty of Medicine, Molecular Genetics Section, Biochemistry and Molecular Genetics Service, Hospital Clínic, Barcelona, Spain

Corresponding authors:

Csilla Krausz MD, PhD
Department of Experimental and Clinical Biomedical Sciences, University of Florence; Viale Pieraccini 6- 50139 Florence, Italy. Phone: +39 055 4271415 – Fax: +39 055 4271413 ; email: c.krausz@dfc.unifi.it

Elisabet Ars PhD
Molecular Biology Laboratory-Fundació Puigvert; Cartagena 340-350- 08025 Barcelona, Spain
Phone: +34 93 4169700 – Fax +34 93 4169730
email: ears@fundacio-puigvert.es

Abstract

AZF microdeletion screening is routinely performed in the diagnostic work-up for male infertility; however, some issues remain debated. In this study, we provide insights into the sperm concentration cut-off value for routine testing, the predictive value of *AZFc* deletion for testicular sperm retrieval and the Y-background contribution to the inter-population variability of deletion frequencies. In the Spanish population, partial *AZFc* rearrangements have been poorly explored and no data exist on partial duplications. In our study, 27/806 (3.3%) patients carried complete *AZF* deletions. All were azoo/cryptozoospermic, except for one whose sperm concentration was 1-2x10^6/ml. In *AZFc* deleted men, we observed a lower sperm recovery rate upon conventional TESE (9.1%) compared to the literature (60%-80% with microTESE). Haplogroup E was the most represented among non-Spanish whereas hgr P among Spanish *AZF* deletion carriers. The analysis of *AZFc* partial rearrangements included 330 idiopathic infertile patients and 385 controls of Spanish origin. *Gr/gr* deletion, but not *AZFc* partial duplications, was significantly associated with spermatogenic impairment. Our data integrated with the literature suggest that: i) routine *AZF* microdeletion testing could eventually include only men with ≤2x10^6/ml; ii) classical TESE is associated with low sperm recovery rate in azoospermic *AZFc* deleted men, therefore microTESE should be preferred; iii) Y background could partially explain the differences in deletion frequencies among populations. Finally, our data on *gr/gr* deletion further support the inclusion of this genetic test in the work-up of infertile men, while partial *AZFc* duplications do not represent a risk for spermatogenic failure in the Spanish population.

Keywords: male infertility, Y-microdeletions, *gr/gr* deletion, *AZFc* duplications, Y-linked CNVs.

INTRODUCTION

Y-chromosome microdeletions are a well-established genetic cause of severe spermatogenic failure and their molecular diagnosis is part of the diagnostic work-up of severe male factor infertility.[1] These sub-microscopic deletions, involving the *AZF* region of the Yq, can be classified according to the recombination hot-spot and have been designated as *AZFa*, P5-proximal P1 (*AZFb*), P5-distal P1 (*AZFbc*), P4-distal P1 (*AZFbc*) and *b2/b4 (AZFc*).[2,3] The deletion frequency clearly varies according to the semen phenotype; indeed, severely oligozoospermic and azoospermic men have the highest risk of carrying Y microdeletions. The different deletion frequency observed even within similar semen categories amongst infertile men from different populations suggests that also ethnic background could influence the occurrence of this genetic

anomaly. The lowest deletion frequency (1.8%) was reported in German and Danish idiopathic severely oligozoospermic men,[1,4] whereas the highest in an ethnically admixed population from France (13.7%)[5] and in Romanians (10%).[6] Data on the prevalence of classical *AZF* deletions in men attending an infertility clinic in Spain derive from two independent surveys, with an overall frequency of 5.4% and 7%, respectively.[7,8]

Due to its complex structure, rich in massive near-identical amplicons, the *AZFc* region is particularly susceptible to homology-based intra-chromosomal recombination events and hence to structural variations as Copy Number Variations (CNVs).[9,10] In addition to the classical *AZFc* deletion, several recurrent partial deletions (named *gr/gr, b2/b3* and *b1/b3*) and duplications (*b2/b4* duplication) have been reported.[11,12] Even though all partial rearrangements produce either a decrease or an increase in *AZFc* gene dosage, only the "*gr/gr* deletion" resulted to be clinically relevant. The clinical significance of this recurrent deletion has been object of a long-lasting debate. Controversies are mainly due to selection biases and the lack of ethnic matching between cases and controls.[13] Notwithstanding, the 4 meta-analysis published so far on this topic indicate that *gr/gr* deletion represents a significant risk factor for impaired sperm production.[14-17] The clinical relevance of *gr/gr* deletion has been confirmed further by a recent multi-ethnic population-based survey of over 20.000 Y chromosomes, reporting a 1.9-fold increased risk of severe spermatogenic failure.[18] The prevalence and clinical significance of partial *AZFc* rearrangements in the Spanish population has been little explored and only one pilot study was performed.[19] Similarly, partial *AZFc* duplications in male infertility have been poorly explored. To date, only two groups have addressed this topic in the Taiwanese and Italian populations, reaching contradictory conclusions.[12,20] In addition,

by using a consecutive cohort study model Noordam et al suggested that both lower and higher *DAZ* gene dosage could be deleterious for spermatogenesis.[21]

This study presents the 8-year experience of our clinic in testing infertile men for Y-linked CNVs. Our first aim was to thoroughly describe the genetic makeup (karyotype and classical *AZF* deletions) of consecutive infertile men referring to our genetic laboratory and, thus, to provide further data on the clinical indications for routine genetic testing. Our second aim was to corroborate the clinical relevance of *gr/gr* deletion in Spain by performing a detailed molecular characterization of the *AZFc* region in a carefully matched case/control study setting.

For the first time, we provide data on the prevalence and clinical significance of *AZFc* partial duplications in the Spanish population, contributing to outline the effect of an increased *DAZ* gene dosage on sperm production in a Caucasian Y background.

PATIENTS AND METHODS

Subjects

We retrospectively analyzed a set of 806 consecutive infertile men, screened for Y-chromosome microdeletions between November 2004 and December 2012. Most of them (72.95%) were Spanish while the remaining (27.05%) was of different geographic origin. The majority of non-Spanish patients (53.7 %) came from North-Western Africa, mostly from Morocco (44.0 %). The second most represented countries were Southern and Central America (22%) followed by Middle and Far East (9.2%), Eastern, North-Western and Southern Europe accounting for 7.3%, 6.0% and 1.8% of non-Spanish patients, respectively. All patients underwent a comprehensive andrological examination (including physical examination, scrotal ultrasound and hormone analysis) and karyotype analysis was performed for 747 men. Based on clinical and karyotype

126 data, patients were classified into “idiopathic” and “non-idiopathic” **(Table 1)** except
127 for 27 (3.3%), whose medical history resulted insufficient for an etiologic classification.
128 Semen analysis was performed according to WHO guidelines[22] except for morphology,
129 for which strict criteria were used. Two hundred ninety one patients were azoospermic
130 (AZ); 392 and 88 presented severe (SOZ; 0<Sperm Concentration (SC) $\leq 5x10^6$/ml) and
131 moderate oligozoospermia (MOZ; 5<SC<$20x10^6$/ml), respectively, 31 had normal SC
132 ($\geq 20x10^6$/ml) but low motility (asthenozoospermia) or <4% of normal morphology
133 (teratozoospermia) or a combination of both (asthenoteratozoospermia). For four
134 patients semen parameters were not available. Bilateral testicular biopsy was performed
135 in 213 patients. A single biopsy was retrieved after scrotal incision from each testis for
136 both diagnostic (to define the type of tubular damage) and therapeutic purposes (to
137 recover spermatozoa for assisted reproductive techniques-ART). Overall, mature sperm
138 could be retrieved in 45.1% of cases (for further details see **Supp**. **Table 1**).
139 *Study population for the screening of partial AZFc rearrangements:* A total of 715
140 Spanish subjects, 330 strictly selected “idiopathic” infertile patients and 385 controls,
141 were analyzed for partial *AZFc* rearrangements. This group included: 94 AZ, 190 SOZ
142 and 46 MOZ men. Controls were recruited on the basis of normal sperm parameters[22]
143 among sperm donors and men with proven fertility (pre-vasectomy). The Total Motile
144 sperm Count (TMC) was calculated for all subjects, by multiplying semen volume by
145 sperm concentration and the percentage of progressively motile spermatozoa.
146 To prevent recruitment bias, much care was taken for the ethnic and geographic
147 matching of patients and controls. All of them were explicitly asked for their paternal
148 and maternal origin and only subjects with proven Spanish ancestry were included. The
149 Y chromosome haplogroup (hgr) analysis further confirmed the similar Y chromosome

background in cases and controls (see **Supp. Figure 3**). This study was approved by the local Ethics Committees and all participants signed an informed consent.

Methods

Molecular Analysis: Genomic DNA was extracted from peripheral blood samples using a standard method.[23] The screening for Yq microdeletions was performed according to the European Academy of Andrology (EAA)/European Molecular Genetics Quality Network (EMQN) guidelines[24] **(Figure 1B)**. Detection and molecular characterization of partial *AZFc* rearrangements was performed according to a previously reported method.[20] Briefly, we analyzed STSs sY1291, sY1191, sY1189, sY1197 and sY1192 (see GeneBank accessions G72340, G73809, GF102061, G67168 and G67166 for PCR primers and conditions) and identified the *gr/gr* deletion by the absence of sY1291 and sY1189 and *b2/b3* deletion by the absence of sY1192 and sY1191. .

Y hgr definition: Y hgr was defined in all individuals with partial *AZFc* rearrangements and in 21 *AZF* classical deletion carriers. In addition, about 60% of subjects recruited for the case-control study were analyzed in order to exclude population stratification bias. All individuals were genotyped for 6 binary markers (M145, M96, M9, M45, M168 and LLy22g), using a multiplexed primer set previously described[25] and adapted for SNaPshot single base extension (Applied Biosystems, CA, USA). This allowed the definition of 8 hgr branches: A,B; DE,D; E; C,F,G,H,I; K,L,M,NO,O,S,T; J; N; P,Q,R.[26] Marker 12f12 was tested only to discriminate between hrgs C,F,G,H,I and hgr J.

Statistical analysis: SPSS software (version 17.0) was used. We tested the significance of the observed difference in the incidence of partial *AZFc* deletions between patients and controls using Fisher's exact test. As SC and TMC were not normally distributed, we used the non-parametric Median Test to compare median values of SC and TMC

between individuals grouped according to *DAZ* copy number (CN) (P<0.05 as statistically significant). Potential confounding factor, for partial *AZFc* deletions, were avoided by screening only individuals with no partial duplications, and vice-versa.

RESULTS

I. Routine diagnostic screening: *AZF* deletions and karyotype anomalies

Among 806 patients, a total of 27 were found with a complete *AZF* deletion (3.3%). Karyotype anomalies were reported in a total of 58 patients with the most frequent anomaly represented by Klinefelter syndrome (for details see **Supp. Table 2**). Six out of 27 deletion carriers showed abnormal karyotype: i) 4 cases with the entire Yq missing (three 46,XX male and one with 46,X,i(Yp)), ii) 2 terminal *AZFbc* deletions (ChrY. hg19:g.(19357589_22570359)_(58912042_?)del) (LOVD3 data base Variant ID: 0000021249)[27]with breakpoint at P5 palindrome described at the karyotype analysis as idic(Yp). Among the 689 infertile patients with normal karyotype, 21 "genuine" Y-chromosomal microdeletions were identified (3.0%): i) 1 complete *AZFa* (ChrY. hg19:g.(14328345_14607475)_(15132293_15603923)del) (Variant ID: 0000021250); ii) 20 *AZFc*: two terminal (ChrY:hg19:g.(24524070_24872541)_(58912042_?)del) (Variant ID: 0000021251) and 18 *b2/b4* (ChrY:hg19:g. (24524070_24872541)_(25316578_28457316)del) (LOVD3 ID:chrY_000070) *AZFc* deletions (**Figure 1C**).

AZF deletion frequency: estimating deletion frequencies according to the aetiology showed a relatively higher frequency of deletions in the "idiopathic" (4.2%, 19/454) compared to the "non-idiopathic" group (2.4%, 8/325) (**Table 1**). In order to evaluate the impact of semen phenotype and aetiology on the deletion frequency, we calculated the frequency for distinct semen categories belonging to different etiologic/sperm

concentration groups **(Table 2)**. The large majority of *AZF* deletion carriers (21/27) were AZ men, most of whom were "idiopathic" (13/152; 8.5%). Also in this case, the aetiology seems to play an important role, since the deletion frequency in the "non-idiopathic" group was significantly lower (2.2%; *P*=0.037). The deletion frequency in Spanish "idiopathic" infertile men was significantly lower compared to non-Spanish (2.7% versus 7.9%; *P*=0.018).

Genotype-phenotype correlation: only 6 subjects (all *AZFc* deletion carriers) presented spermatozoa in their ejaculate, 5 with $<1x10^6$/ml and 1 with $1.2x10^6$/ml. At least 3 semen analyses were performed for each individual over 1-2 years. Among patients with $<1x10^6$/ml, two (07-026, 06-192) displayed a nearly stable SC over time (both around $0.01x10^6$/ml), whereas more evident although not significant oscillations in the range of cryptozoospermia (CR) were observed in 06-012 (SC=$0.095\text{-}0.044x10^6$/ml), 08-039 (SC=$0.04\text{-}0.250x10^6$/ml) and 07-313 (SC=$0.15\text{-}0.250\ x10^6$/ml). A temporal trend for sperm number reduction was observed in the *b2/b4 AZFc* deletion carrier with $>1x10^6$/ml (09-067), who displayed a SC decrease from SOZ ($1.6x10^6$/ml) to CR ($0.260x10^6$/ml), respectively. In two SOZ patients ICSI was performed with success using ejaculated spermatozoa and resulted in the birth of a healthy girl in both cases. No pregnancy was achieved in the other 4 cases.

AZ men with *AZFc* deletion had variable testicular phenotypes ranging from "pure" or "mixed" Sertoli Cell-Only Syndrome (SCOS) to "pure" bilateral hypospermatogenesis. TESE was performed in 11 patients and only one had spermatozoa (9.1% sperm recovery rate upon TESE)

The patient with complete *AZFa* deletion (sample 10-452) had pure bilateral SCO histology and no spermatozoa could be recovered upon TESE. Detailed

genotype/phenotype description and ART results for *AZF* deletion carriers are reported in **Table 3**.

II. Partial *AZF*c rearrangements: case/control association study

***gr/gr* deletion:** a conventional *gr/gr* deletion (ChrY.hg19:g.(24876071_25505070)_(25505734_25316178)del) (LOVD3 DB-ID:chrY_000067) was found in 17 subjects (12 infertile patients and 5 normozoospermic controls). The deletion frequency between the two groups was statistically significant (12/302, 3.9% vs 5/359, 1.4%; OR=2.853; 95% CI=1.017-8.007; *P*=0.032) (**Table 4A).**

***b2/b3* deletion:** this type of deletion (ChrY.hg19:g.(24524070_24872541)_(24876071_25505070)del) (LOVD3 DB-ID: chrY_000068) was found only in the patient group (4/296) with a significantly different frequency compared to controls (0/354) (1.3% vs 0.0%, *P*=0.043).

A detailed description about *gr/gr* and *b2/b3* deletion carriers is given in **Supp. Table 3.**

Atypical deletions: two patients (11-513 and 05-236) presented an atypical deletion pattern (both sY1291 and sY1191 positive), associated with the removal of *DAZ3/4* and *CDY1B* (g.[26909216_27053187del;26191377_26194161del]) (Variant ID: 0000021252) in the first subject and of *DAZ3/4* and both *CDY1* copies (g.[26909216_27053187del; 26191377_26194161del; 27768264_27771049del]) (Variant ID: 0000021253) in the second one.

All partial deletions: we calculated the combined frequency of all the deletions that decrease the *AZFc* gene content of at least 50%, observing a significantly higher frequency in patients compared to controls (5.8% vs 1.8%, respectively; OR=4.196; 95% CI=1.576-11.170; *P*=0.001) (**Table 4A).**

***AZFc* duplications:** an increased *DAZ* gene dosage (6 or 8 *DAZ* copies) associated with a simultaneously increased *CDY1* gene dosage (3 or 4 copies) was observed in 28 Y chromosomes. These CNVs likely derive either from *gr/gr* (ChrY. hg19: g.(24876071_25505070)_(25505734_25316178)dup) (Variant ID: 0000021256) or *b2/b4* duplication (ChrY. hg19: g.(24524070_24872541)_(25316578_28457316)dup) (Variant ID: 0000021258) events and did not show significant differences between patients (4.9%) and controls (3.5%) (**Table 4A**). We also found one control displaying a *gr/gr* deletion-duplication (ChrY. hg19: g.[(24876071_25505070)_(25505734_25316178)del;(24524070_24872541)_(25316578_28457316)dup] (Variant ID:0000021254) characterized by 4 *CDY1B* and 8 *DAZ3/4* gene copies. Moreover, another patient carried a *b2/b3* deletion followed by a *b2/b4* duplication (ChrY. hg19: g.[(24524070_24872541)_(24876071_25505070)del;(24524070_24872541)_(25316578_28457316)dup] (Variant ID: 0000021255) that restored the reference gene dosage, and thus presented 4 *DAZ* copies (*DAZ1/2* and *DAZ3/4*) and 2 copies of *CDY1B*.

Isolated *CDY1* and *DAZ* CNVs: two controls showed an isolated increase of *CDY1* CN with 3 and 4 *CDY1,* respectively. Finally, isolated amplification of *DAZ* was found in 13 subjects: 10 (4 patients and 6 controls) with 6 *DAZ* copies while 3 (all controls) presented 8 *DAZ* copies (**Table 4B**).

III. Impact of the *DAZ* CNVs on semen quality

In order to evaluate the effect of *DAZ* gene CNVs on semen quality, we grouped all subjects into 5 different categories: 0, 2, 4, 6 and 8 *DAZ* gene copies (for details see **Table 4B**). Men with 0 and 2 *DAZ* gene copies showed a significantly lower SC (Median with 25th/75th percentiles: $0.0x10^6$; 0.0-0.07 and $3.0x10^6$/ml; 0.16-15.0,

respectively)) and TMC (0.0 x10^6; 0.0-0.0 and 2.4 x 10^6; 0.13-30.0, respectively) compared to those bearing 4 *DAZ* gene copies (Median SC 35.0 x 10^6/ml; 0.16-15.0 and Median TMC of 31.3 x 10^6; 0.13-125.9). Increased *DAZ* gene CN (both 6 and 8 copies), although showing lower sperm count, was not significantly associated to reduced sperm quality (**Figure 2**).

IV. Effect of Y chromosome background

Y hgr analysis was performed in order to: i) search for a putative association between Y background and formation of complete *AZF* deletions and partial *AZFc* rearrangements; ii) further evaluate the contribution of Y hgr to the phenotypic expression of the *gr/gr* deletion. Branches P,Q,R were the most represented in the whole study population (63.1% of all subjects analyzed) (**Supp. Figure 1**). All Spanish *AZF* deletion carriers belonged to these hgrs (**Supp. Figure 2A**), which conversely made up only 15.4% of non-Spanish carriers. In fact, the Y hgr mostly observed in this cohort was hgr E (23.1%) found in 3 African patients (2 from Morocco and 1 from Cameroon), followed by the branches C,F,G,H,I; J (15.4% each) and K,L,M,NO,S,T/N (7.7%) **(Supp. Figure 2B).** Concerning the *gr/gr* deletion study (exclusively based on the Spanish population), we observed a similar Y hgr distribution in patients and controls: the majority belonged to the hgrs P,Q,R (63.5% and 62.7% respectively) whereas the rest showed matching frequencies between the two groups (**Supp. Figure 3**). Similarly, branches P,Q,R were the most represented among *gr/gr* deletion carriers both in patients (9/12; 75%) and controls (3/5; 60%) and none of the other Y hgrs observed (E; J; K,L,M,NO,S,T/N; C,F,G,H,I) showed a significant enrichment in the two phenotypic groups.

DISCUSSION

Many aspects of Y-chromosome microdeletions have been clarified (mechanism of formation, identification of the genes involved, semen phenotype-dependent variation of the deletion frequency), but a few debated issues merit further discussion. First of all, it is unclear whether Y background might predispose to the formation of deletions, thus contributing to the observed "inter-population" variation in the deletion frequency. In our study, the significantly higher deletion frequency observed in non-Spanish compared to Spanish idiopathic infertile men is plausibly due to the different proportion of AZ men (44.4% in the non-Spanish versus 29.3% in the Spanish group); however, the Y background might also represent a contributory factor influencing deletion frequencies. Y hgr analysis showed, as expected, that Spanish *AZF* deletion carriers all belonged to the P,Q,R lineages. In the non-Spanish cohort, consistently with the high proportion of North African patients (53.2%) included, *AZF* deletions were mostly found on hgr E, which is seemingly more prone to Y microdeletions.[28,29] Moreover, the deletion frequency reported in idiopathic AZ and SOZ men (9.09% and 5.5%) in Moroccan population[28] is consistent with our findings in non-Spanish idiopathic AZ and SOZ men (10.7% and 6.9%, respectively). The prevalence of Y microdeletions in our study population (3.3%) is in line with the overall data presented in the literature (3.5%, according to a recent meta-analysis[30]). When comparing our results with the German population (the lowest ever frequency in the literature[1]) for similar semen categories, we found slightly higher frequencies in our Spanish population. On the other hand, the deletion frequency in AZ Spanish and Italian[31] (displaying a more similar Y background) was almost identical (7.3% versus 7.2%), further supporting a possible Y-background effect on deletion frequencies.

316 Another debated issue concerns the sperm concentration cut-off value for routine
317 diagnostic testing and more precisely whether Y microdeletion screening should be
318 indicated for all oligozoospermic men with less than $5x10^6$/ml. By an in-depth literature
319 review, we could observe that only a 2.0% of all the *AZF* deletion carriers with an
320 explicitly indicated SC presented more than $2x10^6$ /ml (**Supp. Table 4**). Accordingly, in
321 our study only one carrier had >$1x10^6$/ml but the SC did not exceed the $2x10^6$/ml. These
322 findings suggest that Y microdeletion screening could be eventually restricted to
323 infertile men with SC ≤$2x10^6$ /ml. We found an *AZF*c deletion in two apparently non-
324 idiopathic AZ patients: one (13-124) presenting with unilateral absence of vas deference
325 and one (07-339) with bilateral cryptorchidism. In both cases, the presence of the
326 microdeletion, rather than the mere clinical condition, explains the AZ phenotype. This
327 implies that the Yq screening in azoospermic men should be performed independently
328 from the presence or absence of other abnormal andrological findings.
329 The predictive value of *AZF* deletions for sperm retrieval at TESE is also still debated.
330 The majority of complete *AZFa* deletion carriers show SCOS in their testes; however,
331 data are extremely scarce and the largest published review reported the presence of
332 spermatids in the testes in 2/26 men.[32] Our patient with the complete *AZFa* deletion
333 showed a complete bilateral SCOS further supporting that TESE should not be
334 attempted in *AZFa* carriers. Our survey reports a sperm retrieval of 9.1% (1/11) in AZ
335 men with *AZFc* deletion. This value lies below the lower limit of the range of sperm
336 recovery rates reported in *AZFc* deleted patients (14.3%-80.0%).[1,31,33–37] This is most
337 likely related to technical issues such as low amount of testicular sample retrieved
338 (single biopsy from each testis) and the procedure used (classical TESE); indeed,
339 laboratories which performed microTESE reported higher sperm recovery (**Supp. Table**

340 **5**). The high proportion of pure SCOS cases among our *AZFc* carriers represents
341 another possible explanation for such a low retrieval rate.
342 Concerning the OZ subjects, there are studies both in favor and against the need to
343 cryopreserve spermatozoa to counteract the progressive deterioration of sperm quality
344 (from SOZ to AZ). We observed a single *AZFc* deletion carrier with SOZ that
345 developed into CR, indicating that a progressive decline in spermatogenic activity in
346 patients bearing *AZFc* deletions may occur. However, further longitudinal studies are
347 needed to distinguish between physiological oscillations and real impairment of sperm
348 parameters over time. Several authors proposed a higher risk for Turner syndrome and
349 ambiguous genitalia in ICSI babies born from *AZFc* deletion carriers (for review see ref
350 24). Our survey reveals two successful pregnancies with healthy female offspring,
351 providing additional data to the presently scarce literature about this issue (44 babies
352 described so far). As for partial *AZFc* rearrangements, one of the strengths of this study
353 lies in the careful selection of patients and controls considering both phenotype (only
354 strictly idiopathic infertile and normozoospermic controls were included) and
355 geographic origin (all individuals were rigorously of Spanish ancestry). The Y-
356 chromosome hgr analysis in patients and controls further demonstrated the lack of
357 population stratification bias in the study. The selective recruitment strategy, together
358 with the detailed molecular characterization of the *AZFc* region in the whole study
359 population, allowed us to provide highly reliable data both on partial *AZFc* deletions
360 and duplications.
361 Concerning the *gr/gr* deletion, we found that Spanish *gr/gr* deletion carriers have an
362 increased probability (OR=2.8) of impaired spermatogenesis compared to non-carriers.
363 Overall, this data together with a previous pilot study[19] further confirms the *gr/gr*

deletion as a significant risk factor in the Spanish population (OR=4.8; 95% CI=1.863-12.623; P<0.001; **Supp. Table 6**), providing additional support of its clinical relevance in Caucasians, consistently with the meta-analyses published, so far.[14–17] The clinical implication of this finding in the Spanish population reinforces the idea that the *gr/gr* deletion screening should gain more consideration when dealing with infertile couples. This issue is of particular importance considering that, in some populations, partial deletions were shown to favor the occurrence of complete deletions.[38,39]

The majority of *gr/gr* deletion carriers belonged, as expected, to the P,Q,R branches in both patients and controls, supporting that the phenotypic variability of *gr/gr* deletion is independent of Y-chromosomal background in Europeans.[40] Interestingly, we found the *b2/b3* deletion only in the patient cohort (4/296; 1.3%) and only one carrier belonged to hgr N and thus had the constitutive *b2/b3* deletion. The remaining three *b2/b3* deletion carriers belonged to hrgs C,F,G,H,I which is a frequent Y hgr in Moroccan population for which *b2/b3* deletion has been recently reported in association with male infertility.[28,29] The significant association observed indicates a Y background-dependent deleterious effect of *b2/b3* on spermatogenesis.

Here, we present the first screening for partial *AZFc* duplications in a large series of patients and controls of Spanish origin. Our data, consistently with those regarding the Italian population,[20] confirmed that both partial and complete *AZFc* duplications do not represent any risk for spermatogenic failure in the Caucasian population.

When considering the *DAZ* CN independently from the type of partial rearrangements *(gr/gr, b2/b3* or "atypical" deletion patterns, partial or complete *AZFc* duplications), we found that the reduced *DAZ* CN (0 and 2 copies) is associated with a significant reduction in semen quality in terms of SC and TMC as previously reported by Noordam

et al.[21] On the other hand, for the first time we showed that an increase (6 or 8 copies) in respect to the most common *DAZ* gene dosage (4 copies) does not affect significantly semen quality.

In conclusion, our 8-year experience together with the literature review allowed us to further clarify a number of debated issues concerning the routine Y chromosome microdeletions screening: i) the indication for routine Y-chromosome microdeletion screening may be eventually limited to subjects with ≤ 2 x10^6/ml; ii) in azoospermic *AZFc* deletion carriers classical TESE is associated to a low sperm recovery rate (9.1%), therefore microTESE, which allows better outcomes, should be regarded as the best option for sperm retrieval in these patients; iii) Y background could partially account for the differences in deletion frequency among populations. Finally, in our view, *gr/gr* deletion screening can be considered as part of the diagnostic work-up of idiopathic oligozoospermic men since it is a confirmed co-factor that contributes to impaired sperm production.[41] On the contrary, in line with the Italian data, partial *AZFc* duplication is unlikely to be involved in the etiopathogenesis of spermatogenic impairment in Caucasian populations.

Conflict of Interest Statement: The authors declare no conflict of interest.

Acknowledgments: The authors wish to thank the patients for participating in this study, the clinicians of the Andrology Service (Drs. G.Ortiz, S.Camarena) for providing samples, the staff of Seminology and Embriology Laboratory (Drs. A.Mata, O.Lopez, A.Garcia, O.Martinez, R.Gusta) for providing seminal phenotypes and information about ART outcomes and M.Vieri and A.Morales for technical support. A special thank to Esperança Martí i Salís, president of Fundació Puigvert, for her continuous support. We thank the IIB Sant Pau-Fundació Puigvert

Biobank for kindly providing some of the DNA samples. This work was funded by grants from the Spanish Health Ministry (FIS-08/1138 and FIS 11/02254) to CK and from the Spanish Ministry of Economy and Competitiveness (BFU 2009-07118) to RO.

References

1. Simoni M, Tüttelmann F, Gromoll J, Nieschlag E: Clinical consequences of microdeletions of the Y chromosome: the extended Münster experience. *Reprod Biomed Online* 2008; **16:** 289–303 .

2. Vogt PH, Edelmann A, Kirsch S *et al:* Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. *Hum Mol genet* 1996; **5:** 933–943 .

3. Repping S, Skaletsky H, Lange J *et al:*. Recombination between palindromes P5 and P1 on the human Y chromosome causes massive deletions and spermatogenic failure. *Am J Hum Genet* 2002; **71:** 906–922 .

4. Cruger DG, Agerholm I, Byriel L, Fedder J, Bruun-Petersen G: Genetic analysis of males from intracytoplasmic sperm injection couples. *Clin Genet* 2003; **64**: 198–203 .

5. Krausz C, Quintana-Murci L, Barbaux S *et al*:A high frequency of Y chromosome deletions in males with nonidiopathic infertility. *The J Clin Endocrinol Metab* 1999*;* **84**: 3606–3612.

6. Raicu F, Popa L, Apostol P, *et al:* Screening for microdeletions in human Y chromosome--AZF candidate genes and male infertility. *J Cell Mol Med* 2003;**7**: 43–48

7. Oliva R, Margarit E, Ballescá JL *et al:* Prevalence of Y chromosome microdeletions in oligospermic and azoospermic candidates for intracytoplasmic sperm injection. *Fertil Steril* 1998; **70:** 506–510 .

8. Martínez MC, Bernabé MJ, Gómez E *et al:* Screening for AZF deletion in a large series of severely impaired spermatogenesis patients. *J Androl* 2000; **21:** 651–655

9. Skaletsky H, Kuroda-Kawaguchi T, Minx PJ *et al:* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 2003; **423:** 825–837.

10. Jobling MA: Copy number variation on the human Y chromosome. *Cytogenet Genome Res* 2008*;* **123:** 253–262.

11. Repping S, Skaletsky H, Brown L *et al:* Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. *Nat Genet* 2003*;* **35:** 247–251.

12. Lin YW, Hsu LC, Kuo PL *et al*: Partial duplication at AZFc on the Y chromosome is a risk factor for impaired spermatogenesis in Han Chinese in Taiwan. *Hum Mutat* 2007*;* **28:** 486–494.

13. Krausz C, Chianese C, Giachini C, Guarducci E, Laface I, Forti G: The Y chromosome-linked copy number variations and male fertility. *J Endocrinol Invest* 2011; **34:** 376–382.

14. Navarro-Costa P, Gonçalves J, Plancha CE: The AZFc region of the Y chromosome: at the crossroads between genetic diversity and male infertility. *Hum Reprod Update* 2010; **16:** 525–542

15. Visser L, Westerveld GH, Korver CM *et al*:Y chromosome gr/gr deletions are a risk factor for low semen quality. *Hum Reprod* 2009; **24:** 2667–2673.

16. Tüttelmann F, Rajpert-De Meyts E, Nieschlag E, Simoni M: Gene polymorphisms and male infertility--a meta-analysis and literature review. *Reprod Biomed Online* 2007; **15:** 643–658.

17. Stouffs K, Lissens W, Tournaye H, Haentjens P: What about gr/gr deletions and male infertility? Systematic review and meta-analysis. *Hum Reprod Update* 2011*;* **17:** 197–209

18. Rozen SG, Marszalek JD, Irenze K *et al:* AZFc deletions and spermatogenic failure: a population-based survey of 20,000 Y chromosomes. *Am J Hum Genet* 2012; **91:** 890–896.

19. De Llanos M, Ballescà JL, Gázquez C, Margarit E, Oliva R: High frequency of gr/gr chromosome Y deletions in consecutive oligospermic ICSI candidates. *Hum Reprod* 2005*;* **20,** 216–220.

20. Giachini C, Laface I, Guarducci E, Balercia G, Forti G, Krausz C:Partial AZFc deletions and duplications: clinical correlates in the Italian population. Hum Genet 2008; **124:** 399–410.

21. Noordam MJ, Westerveld GH, Hovingh SE *et al:* Gene copy number reduction in the azoospermia factor c (AZFc) region and its effect on total motile sperm count. *Hum Mol Genet* 2011; **20:** 2457–2463.

22. World Health Organization:*WHO Laboratory manual for the examinaltion of human semen and semen-cervical mucus interaction*, 4th ed. University Press, Cambridge (UK), 1999

23. Miller SA, Dykes DD, Polesky HF: A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res 1988;* **16:** 1215.

24. Simoni M, Bakker E, Krausz C: EAA/EMQN best practice guidelines for molecular diagnosis of y-chromosomal microdeletions. State of the art 2004. *Int J Androl* 2004; **27:** 240–249.

25. Noordam MJ, Repping S: The human Y chromosome: a masculine chromosome. *Curr Opin Genet Dev* 2006; **16:** 225–32 (2006).

26. Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF: New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res* 2008; **18:** 830–838.

27. Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, den Dunnen JT: LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat.* 2011;**32:** 557-63.

28 Eloualid A, Rhaissi H, Reguig A *et al:* Association of spermatogenic failure with the b2/b3 partial AZFc deletion. *PLoS One* 2012*;* **7:** e34902.

29. Imken L, El Houate B, Chafik A *et al:* AZF microdeletions and partial deletions of AZFc region on the Y chromosome in Moroccan men. *Asian J Androl* 2007; **9:** 674–678.

30. Hofherr SE, Wiktor AE, Kipp BR, Dawson DB, Van Dyke DL: Clinical diagnostic testing for the cytogenetic and molecular causes of male infertility: the Mayo Clinic experience. *J Assist Reprod Genet* 2011; **28:** 1091–1098.

31. Ferlin A, Arredi B, Speltra E *et al:* Molecular and clinical characterization of Y chromosome microdeletions in infertile men: a 10-year experience in Italy. *The J Clin Endocrinol Metab* 2007; **92:** 762–770.

32. Kleiman SE, Almog R, Yogev L *et al:* Screening for partial AZFa microdeletions in the Y chromosome of infertile men: is it of clinical relevance? *Fertil Steril* 2012; **98:** 43–47.

33. Maurer B, Gromoll J, Simoni M, Nieschlag E: Prevalence of Y chromosome microdeletions in infertile men who consulted a tertiary care medical centre: the Münster experience. *Andrologia* 2001; **33:** 27–33.

34. Peterlin, B., Kunej, T., Sinkovec, J., Gligorievska, N. & Zorn, B. Screening for Y chromosome microdeletions in 226 Slovenian subfertile men. *Hum Reprod* 2002; **17:** 17–24.

35. Hopps CV, Mielnik A, Goldstein M, Palermo GD, Rosenwaks Z, Schlegel PN: Detection of sperm in men with Y chromosome microdeletions of the AZFa, AZFb and AZFc regions. *Hum Reprod* 2003; **18:** 1660–1665.

36. Stahl PJ, Masson P, Mielnik A, Marean MB, Schlegel PN, Paduch DA: A decade of experience emphasizes that testing for Y microdeletions is essential in American men with azoospermia and severe oligozoospermia. *Fertil Steril* 2010; **94:** 1753–1756.

37. Choi JM, Chung P, Veeck L, Mielnik A, Palermo GD, Schlegel PN: AZF microdeletions of the Y chromosome and in vitro fertilization outcome. *Fertil Steril* 2004; **81:** 337–341.

38. Zhang F, Lu C, Li Z *et al*: Partial deletions are associated with an increased risk of complete deletion in AZFc: a new insight into the role of partial AZFc deletions in male infertility. *J Med Genet* 2007; **44:** 437–444.

39. Lu C, Zhang J, Li Y *et al:* The b2/b3 subdeletion shows higher risk of spermatogenic failure and higher frequency of complete AZFc deletion than the gr/gr subdeletion in a Chinese population. *Hum Mol Genet* 2009; **18**: 1122–1130.

40. Krausz C, Giachini C, Xue Y *et al:* Phenotypic variation within European carriers of the Y-chromosomal gr/gr deletion is independent of Y-chromosomal background. *J Med Genet* 2009; **46**:21–31.

41. Krausz C. Male infertility: pathogenesis and clinical diagnosis. *Best Pract Res Clin Endocrinol Metab.* 2011; **25**:271-85.

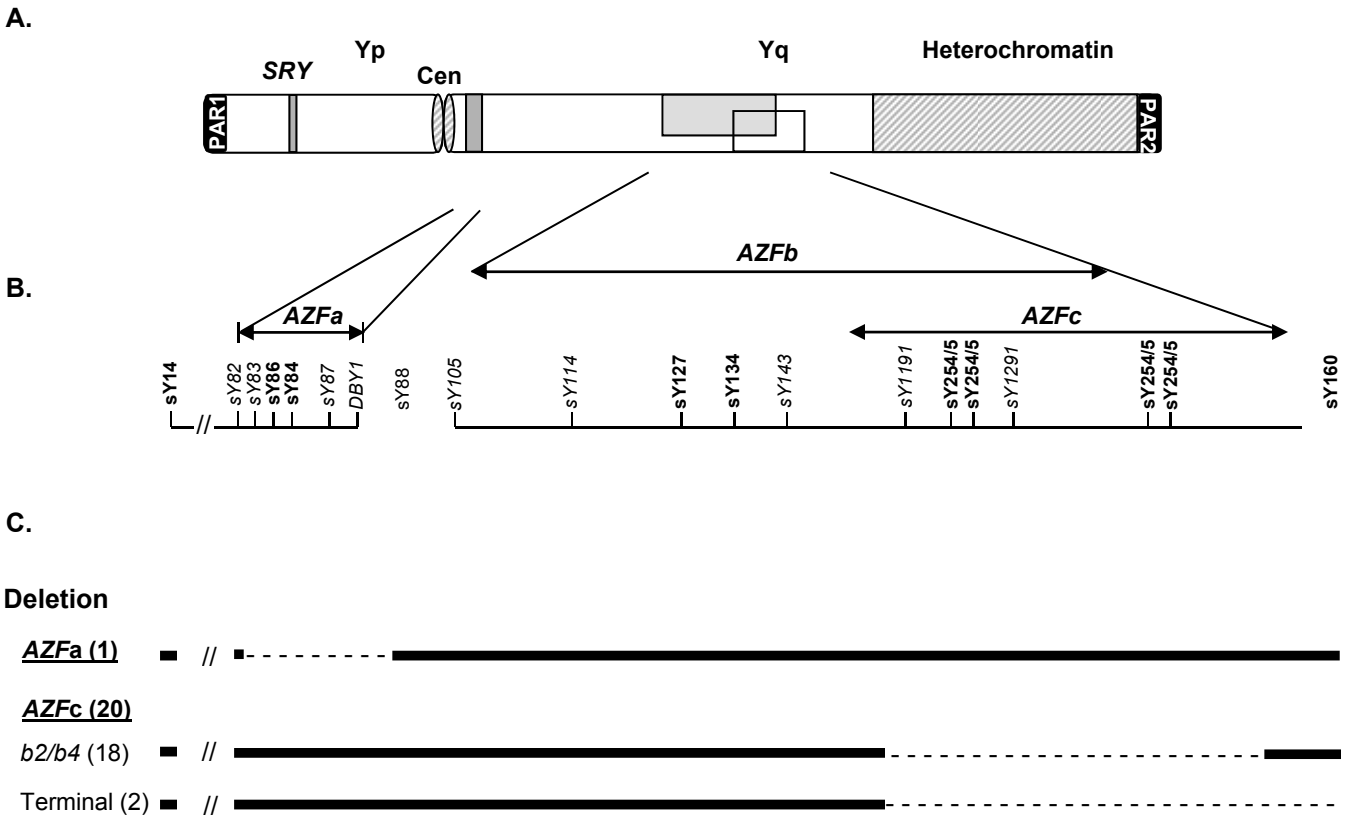
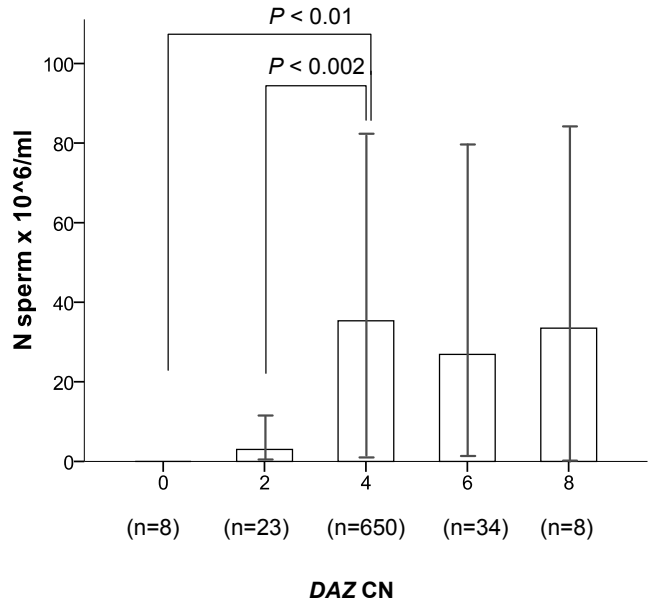


Figure 1. Representation of the 21 “genuine” Yq microdeletions detected. **A.** Schematic representation of the Y chrosome showing the three *AZF* regions. **B.** STS markers used for the diagnosis of Y microdeletions : in **bold** are the STSs used for the first step screening, in *italic* the ones used for the determination of the breakpoints of deletions.



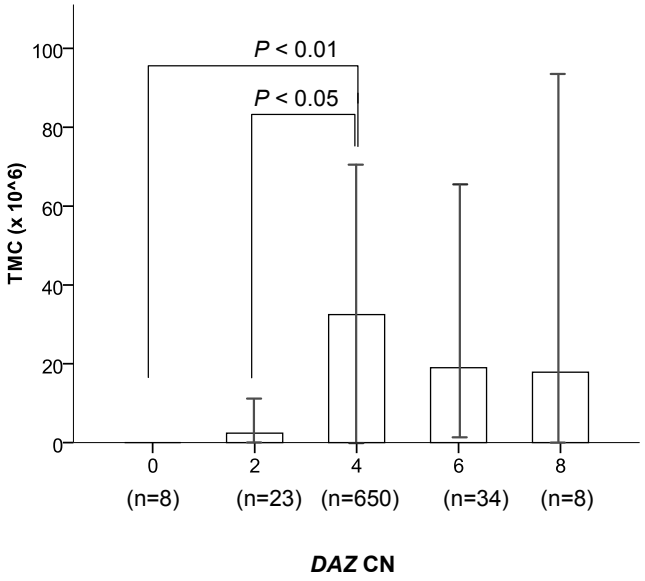


Figure 2. *DAZ* gene copy number effect on sperm concentration (A) and Total Motile sperm Count (TMC) (B) in the case/control study population. Data are presented as median (25th-75th percentiles). **(A)** Comparison between subjects with different *DAZ* gene copy number (*DAZ* CN), showing significantly reduced sperm concentration in men with 0 and 2 *DAZ* copies versus 4 *DAZ* copies (4 *DAZ* copies are considered as a "normal" gene dosage). **(B)** Men with 0 and 2 *DAZ* CN have significant lower TMC with respect to men with the "normal" gene CN (4 *DAZ*).

Table 1. Classification of the Yq (micro)deletion screening cohort according to the geographic origin of patients and to the etiology of spermatogenic disturbance.

Clinical findings	**N Patients with Y chr. (micro) deletion/Total (%)**		
	Spanish	Non Spanish	Total
Non idiopathic infertility	**5/239 (2.0)**	**3/86 (3.5)**	**8/325 (2.4)**
Karyotype abnormalities	3/41 (7.3)	3/17 (17.6)	6/58 (10.3)
Urogenital Obstructions	1/15 (0.0)	0/10 (0.0)	1/25 (4.0)
Cryptorchidism	1/127 (0.8)	0/32 (0.0)	1/159 (0.6)
*Testicular tumor**	0/6 (0.0)	0/1 (0.0)	0/7 (0.0)
Recurrent Infections/Inflammations	0/13 (0.0)	0/11 (0.0)	0/24 (0.0)
*Varicocele***	0/12 (0.0)	0/5 (0.0)	0/17 (0.0)
Other abnormalities	0/25 (0.0)	0/10 (0.0)	0/35 (0.0)
Idiopathic infertility	**9/328 (2.7)**	**10/126 (7.9)**	**19/454 (4.2)**
Unclassified	**0/21 (0.0)**	**0/6 (0.0)**	**0/27 (0.0)**
Total	**14/588 (2.3)**	**13/218 (5.9)**	**27/806 (3.3)**

***** 3 cryptorchid patients are included

****** Bilateral or unilateral varicocele grade 2 or 3 as the only andrological anomaly found.

Additional 54 patients with varicocele associated with other abnormal andrological findings or karyotype anomalies are included in the above etiologic categories.

Idiopathic infertility: no abnormal andrological or genetic findings. Patients with varicocele grade 1 or other mild andrological findings are included

Other abnormalities: includes systemic diseases and testis trauma.

Unclassified: patients whose medical history was insufficient for an etiologic classification.

Table 2. Frequency of Yq microdeletions in idiopathic and non-idiopathic patients with normal karyotype in Spanish and non-Spanish cohorts based on the sperm concentration.

Sperm Concentration (SC) x 10^6/ml	Spanish			Non Spanish			Total infertile men
	Non-idiopathic infertility	*Idiopathic Infertility*	*Total*	*Non-idiopathic infertility*	*Idiopathic Infertility*	*Total*	
	Frequency						
SC = 0	2/66 (3.0)	7/96 (7.3)	**9/162 (5.5)**	0/25 (0.0)	6/56 (10.7)	**6/81 (7.4)**	**15/243 (6.2)**
0 < SC ≤ 1	0/67 (0.0)	2/87 (2.3)	**2/154 (1.3)**	0/18 (0.0)	3/41 (7.3)	**3/59 (5.1)**	**5/213 (2.3)**
1 < SC ≤ 5	0/36 (0.0)	0/87 (0.0)	**0/123 (0.0)**	0/15 (0.0)	1/17 (5.9)	**1/32 (3.1)**	**1/155 (0.6)**
5 < SC < 20	0/19 (0.0)	0/42 (0.0)	**0/61 (0.0)**	0/8 (0.0)	0/9 (0.0)	**0/17 (0.0)**	**0/78 (0.0)**
SC ≥ 20*	0/8 (0.0)	0/16 (0.0)	**0/24 (0.0)**	0/3 (0.0)	0/3 (0.0)	**0/6 (0.0)**	**0/30 (0.0)**
	Cumulative frequency						
SC ≤ 1	2/133 (1.5)	9/183 (4.9)	**11/316 (3.5)**	0/43 (0.0)	9/97 (9.2)	**9/140 (6.4)**	**20/456 (4.4)**
SC ≤ 5	2/169 (1.2)	9/270 (3.3)	**11/439 (2.5)**	0/59 (0.0)	10/114 (8.7)	**10/173 (5.8)**	**21/612 (3.4)**
SC < 20	2/188 (1.0)	9/312 (2.9)	**11/500 (2.2)**	0/66 (0.0)	10/123 (8.1)	**10/189 (5.3)**	**21/689 (3.0)**
Total	2/196** (1.0)	9/328 (2.7)	**11/524 (2.1)**	0/69 (0.0)	10/126 (7.9)	**10/195 (5.1)**	**21/719 (2.9)**

* Subjects with normal sperm concentration but low motility (asthenozoospermia) or <4% normal morphology (teratozoospermia) or with combined anomalies: asthenoteratozoospermia.

** Two Spanish non-idiopathic infertile patients with unknown sperm parameters are excluded.

Table 3. Genotype, phenotype description and ART results of 27 patients with *AZF* deletion

Patient ID	Deletion type	Geographic origin	Y Haplogroup	SC* (x 10^6 /ml)	Karyotype	FSH (U/L)	LH (U/L)	T (ng/ml)	Mean testis volume (ml)	Testis histology/sperm recovery	ART/ Pregnancy
10-452	*AZFa*	Spanish	P,Q,R	0	46,XY	8.28	4.43	6.23	n.a	(SCOS) / sp-	2x IUI-D / 1 baby
08-221	*AZFb,c*	Spanish	n.a	0	(45,X[50], 46,X,idic(Y) q(11.1),ishY (DYZ3,SRY)++	21.3	20.3	4.30	n.a	n.p.	2x IUI-D / 1 baby
10-041	*AZFb,c*	Spanish	n.a	0	45,X[10]/46,X,idic(Y) (q11.22) [40]	11.7	n.p	n.p	13.5	n.p.	IUI-D / 1 pregnancy
04-143	*AZFc b2/b4*	Moroccan	E	0.01	46,XY	28.3	3.67	4.41	13.5	n.p	2x ICSI/ no pregnancy
05-070	*AZFc b2/b4*	Spanish	P,Q,R	0	46,XY	24.7	5.3	3.7	18	n.p.	n.p
07-179	*AZFc b2/b4*	Spanish	P,Q,R	0	46,XY	30.4	n.p	n.p	15	(R: mixed atrophy with no mature spermatids; L: SCOS) / sp+	ICSI ** / 1 baby
07-534	*AZFc b2/b4*	Romanian	C,F,G,H,I	0	46,XY	27	11.8	2.65	15	(80-90% SCOS. 10% Sclero Hialynosis) / sp-	n.p
08-254	*AZFc b2/b4*	Spanish	P,Q,R	0	46,XY	27.8	7.44	5.45	15	n.p.	n.p
09-029	*AZFc b2/b4*	Spanish	P,Q,R	0	46,XY	20.9	9	8.05	10.50	(90% SCOS. 10% HS) / sp-	ICSI ** / no pregnancy
10-465	*AZFc b2/b4*	Moroccan	E	0	46,XY	15.6	n.p	n.p	15	(SCOS) / sp-	n.p
06-167	*AZFc b2/b4*	Lithuanian	K,L,M,N,O,S,T	0	46,XY	8.82	1.82	4.87	22.5	(HS) / sp-	n.p
11-529	*AZFc b2/b4*	Spanish	P,Q,R	0	46,XY	14.2	4.15	2.68		(SCOS) / sp-	IUI-D / no pregnancy
07-339	*AZFc b2/b4*	Spanish	P,Q,R	0	46,XY	21.2	7.6	1.28	12.5	(SCOS) / sp-	IUI-D / 1 baby
07-026	*AZFc b2/b4*	Spanish	P,Q,R	0.01	46,XY	11.2	4.18	17.50	17.5	n.p.	2x ICSI/ no pregnancy
06-192	*AZFc b2/b4*	Moroccan	J	0.01	46,XY	23.8	n.p	1.96	16.00	n.p.	ICSI /no pregnancy

Patient ID	Deletion type	Geographic origin	Y Haplogroup	SC* (x 10^6/ml)	Karyotype	FSH (U/L)	LH (U/L)	T (ng/ml)	Mean testis volume (ml)	Testis histology/sperm recovery	ART/ Pregnancy
06-012	*AZFc b2/b4*	Spanish	P,Q,R	0.095	46,XY	18.4	n.p	n.p	12.50	n.p.	2x ICSI/ no pregnancy
08-039	*AZFc b2/b4*	Moroccan	J	0.19	46,XY	n.p	n.p	n.p	20	n.p.	ICSI / 1 female baby
07-313	*AZFc b2/b4*	Bolivian	P,Q,R	0.325	46,XY	n.p	n.p	n.p		n.p.	IUI-D / pregnancy
09-067	*AZFc b2/b4*	Peruvian	C,F,G,H,I	1.166	46,XY	n.p	n.p	n.p	15	n.p.	ICSI / 1 female baby
08-389	*AZFc term.*	Bolivian	P,Q,R	0	46,XY	14.7	7.02	25.3	n.a	(HS) / sp-	n.p.
10-489	*AZFc term.*	Cameroon	E	0	46,XY	14	7.31	16.2	n.a	(R: 50% Sclero Hialynosis;50% SCOS. L:pure SCOS) / sp-	n.p.
08-107	*AZFa,b,c*	Bolivian	n.a	0	46,XX ish, Yp11.3 (SRY+)	22	n.p	8.2	8	n.p.	IUI-D / 1 pregnancy
09-420	*AZFa,b,c*	Spanish	n.a	0	46, XX.ish, Yp11.3 (SRY+)	16.7	10.55	n.a	10	n.p.	n.p.
10-098	*AZFa,b,c*	Moroccan	n.a	0	(46,X,i(Y) (q11.22), ish Yp11.3(SRY++)	21.1	19.45	2.82	n.a	n.p.	n.p.
09-530	*AZFa,b,c*	Slovak	n.a	0	46,XX	n.p	3.59	5.50		n.p.	n.p.
12-221	*AZFc b2/b4*	Spanish	P,Q,R	0	46,XY	10.6	n.p	n.p	9	(SCOS) / sp-	n.p.
13-124	*AZFc b2/b4*	Spanish	P,Q,R	0	46,XY	22.8	n.p	n.p	11	n.p.	n.p.

SC: Sperm Concentration**; *** If more than one semen sample was available, the median of SC was used.
T: Testosterone**;**
Testis Histology: **R**: right testis; **L**: left testis; **SCOS**: Sertoli Cell Only Syndrome**; MA**: Maturation arrest; **HS**: Hypospermatogenesis;
sp+: spermatozoa recovered upon TESE; **sp-**: spermatozoa not recovered upon TESE.
n.a: not available data; **n.p**: not performed;
IUI-D : IntraUterine Insemination by donor sperm.
ICSI: Intracytoplasmic Sperm Injection
** ICSI performed by using donor sperm

Table 4. *AZFc* (partial) rearrangements in the Spanish study population: idiopathic infertile patients versus normozoospermic controls.

A. Comparison of the frequency of *gr/gr*, all partial deletions and (partial) *AZFc* duplications.

Phenotype	***gr/gr* deletion N/tot (%)**	**P [OR (95% CI)]**	**All partial deletions N/tot (%)***	**P [OR (95% CI)]**	***b2/b4* (partial) *AZFc* duplications N/tot (%)**	**P [OR (95% CI)]**
Patients	12/302 (3.9)	0.032 [2.853 (1.017-8.007)]	18/308 (5.8)	0.001 [4.196 (1.576-11.170]	15/305 (4.9)	0.440 [1.388 (0.671-2.872)]
Controls	5/359 (1.4)		5/359 (1.4)		13/367 (3.5)	

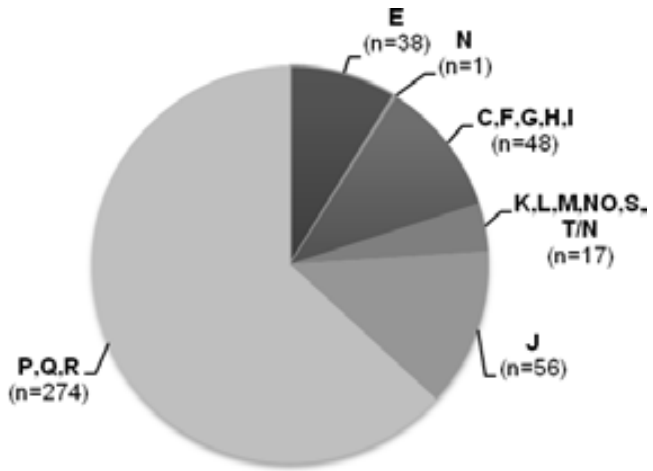
* Included *gr/gr*, *b2/b3* and "atypical" deletions
N/tot: Number/Total, frequency of subjects bearing a specific *AZFc* rearrangement (*gr/gr* deletion, any partial *AZFc* deletion, *b2/b4* (partial) duplication)

B. Distribution of subjects with different *DAZ* gene copy number (CN) based on *AZFc* architecture. Only Spanish idiopathic patients are considered.

***DAZ* CN**	**0**	**2**				**4**				**6**			**8**			
***AZFc* architecture**	***AZFc* del.**	***gr/gr* del.**	***b2/b3* del.**	**Atyp.del**	**Total**	**wt**	**del/dup**	**isolated *CDY1* ampl**	**Total**	***b2/b4* dup.**	**isolated *DAZ* ampl.**	**Total**	***b2/b4* dup.**	**isolated *DAZ* ampl.**	**del/dup**	**Total**
N pat/Tot. (%)	8/327 (2.4)	12/330 (3.6)	4/330 (1.2)	2/330 (0.6)	**18/330 (5.4)**	292/330 (88.5)	1/330 (0.3)	0/330 (0.0)	**293/330 (88.7)**	12/330 (3.6)	4/330 (1.2)	**16/330 (4.8)**	3/330 (0.9)	0/330 (0.0)	0/330 (0.0)	**3/330 (0.9)**
N contr./Tot. (%)	–	5/385 (1.3)	0/385 (0.0)	0/385 (0.0)	**5/385 (1.3)**	355/385 (92.2)	0/385 (0.0)	2/385 (0.5)	**357/385 (92.7)**	12/385 (3.1)	6/385 (1.5)	**18/385 (4.7)**	1/385 (0.2)	3/385 (0.8)	1/385 (0.2)	**5/385 (1.3)**
Total	8/327 (2.4)	17/715 (2.4)	4/715 (0.5)	2/715 (0.3)	**23/715 (3.2)**	647/715 (90.5)	1/715 (0.1)	2/715 (0.3)	**650/715 (90.9)**	24/715 (3.3)	10/715 (1.4)	**34/715 (4.7)**	4/715 (0.5)	3/715 (0.4)	1/715 (0.1)	**8/715 (1.1)**

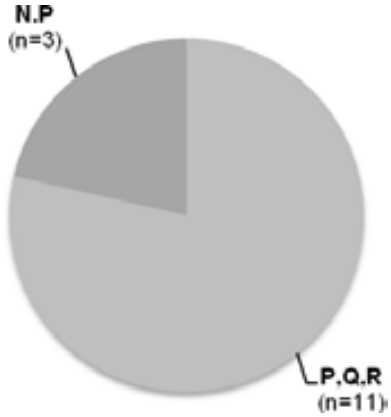
"*DAZ* CN=0" category includes only Spanish idiopathic patients screened for classical *AZF* deletions. "*DAZ* CN=2,4,6,8" categories include patients screened for partial *AZFc* rearrangments. ***AZFc* del.*:*** classical *AZFc* deletion**; wt**: normal *AZFc* architecture***; gr/gr* del**.: *gr/gr* deletion***; b2/b3* del**.: *b2/b3* deletion**; Atyp.del.:** Atypical deletions**; del/dup**: *gr/gr* or *b2/b3* deletion followed by a duplication restoring or increasing the reference *DAZ* dosage**; isolated *CDY1* ampl**.: isolated *CDY1* amplification with normal *DAZ* dosage***; b2/b4* dup**.: *b2/b4* duplication increasing reference *DAZ* dosage to 6 and 8 copies**; isolated *DAZ* ampl**.: isolated *DAZ* amplification

Supplementary Figure 1. Y hgr distribution in the whole study population.

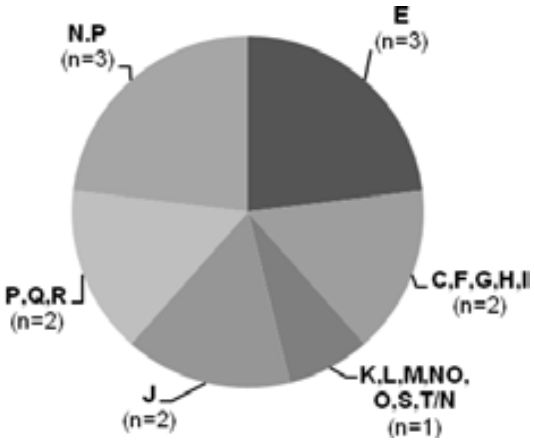


Supplementary Figure 2. Y hgr distribution in *AZF* deletion carriers.

A. Spanish *AZF* deletion carriers

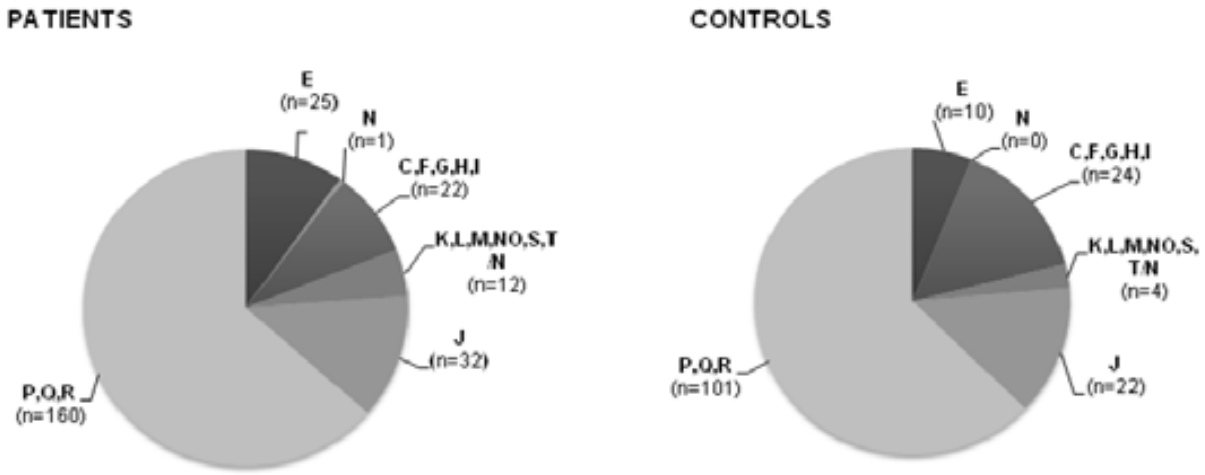


B. Non-Spanish AZF deletion carriers



N.P: Y hgr definition Not Performed in 2 Spanish *AZFb,c* and 4 (1 Spanish and 3 Non-Spanish) *AZFa,b,c* deleted men due to technical limitations.

Supplementary Figure 3. Case-control study on partial *AZFc* rearrangements.



n: number of subjects

Supplementary Table 1. Frequency of Y (micro) deletions in N=213 azoospermic patients who underwent bilateral testicular biopsy grouped on the basis of their histological findings. The number of patients with successful sperm recovery (sp+) in each group is indicated.

Testis Histology	**Frequency *AZF* microdeletion in each group**	**Type of microdeletion**
Complete maturation *	0/15 (0.0)	
SCOS type 1	5/69 (7.2)	4 *AZF*c b2/b4 1 *AZF*a
SCOS type 2	4/37 (10.8)	3 *AZF*c b2/b4 1 *AZF*c term.
Hypospermatogenesis Type1	2/19 (10.5)	1 *AZF*c b2/b4 1 *AZF*c term.
Hypospermatogenesis Type2	0/13 (0.0)	
Maturation Arrest Type1	0/31 (0.0)	
Maturation Arrest Type2	1/16 (6.2)	1 *AZF*c b2/b4
Sclero Hyalinosis (≥ 50% tubules)	0/13 (0.0)	
Total	12/213 (5.6)	

*obstructive azoospermia
Type 1: pure histological pattern
Type 2: predominant pattern in mixed histology

Supplementary Table 2. Distribution of normal and abnormal karyotype in the study population.

Karyotype	No. of patients (%)	Comment
Normal (46, XY)	**689 (85.5)**	
Abnormal	**58 (7.0)**	
Klinefelter Syndrome (47, XXY)	24 (41.2)	
Mosaic Klinefelter syndrome	3 (5.2)	
47, XYY	2 (3.5)	
Reciprocal Translocations	6 (10.3)	
Robertsonian Translocations	9 (15.8)	
Sex Reversal 46, XX, SRY+	2 (3.5)	Y absent **(deletion *AZFabc*)** in both patients
46, XX, SRY-	1 (1.7)	Y absent **(deletion *AZFabc*)**
Mosaic Turner Syndrome	3 (5.3)	
45,X [2]/46, XY [48], SRY+	1 (1.7)	
45,X[10]/46,X,idic(Y) (q11.22) [40], SRY+	1 (1.7)	**Terminal deletion *AZFbc* (sY160 -)**
(45,X[50], 46,X,idic(Y) q(11.1).ishY (DYZ3,SRY)++8.	1 (1.7)	**Terminal deletion *AZFbc* (sY160 -)**
46, X,i(Y)(q11.22), SRY++	1 (1.7)	**Deletion *AZFabc* (sY160 -)**
Other Abnormalities	7 (12.0)	
Unknown karyotype	**59 (7.3)**	

Supplementary Table 3. Phenotype of patients and controls bearing *gr/gr* deletion, *b2/b3* deletion and partial *AZF*c duplication with the indication of the type of *CDY1* and *DAZ* gene copy loss and Y hgr.

Code	Phenotype	Y hgr	Deleted gene copy		Sperm parameters		
			DAZ	*CDY1*	n.of sp/ml x 10^6	Motility A+B (%)	Normal morphology (%)
Infertile men with gr/gr deletion							
04-178	Idiopathic	E	3/4	1b	0.016	0	_
07-098	Idiopathic	E	1/2	1a	0.09	45	_
07-361	Unilateral Varicocele grade II	J	3/4	1a	1.6	17.6	5.2
08-079	Idiopathic	P,Q,R	3/4	1a	0	0	_
08-201	Idiopathic	P,Q,R	1/2	1b	7.7	31.7	2.5
08-220	Bilateral Varicocele grade I and Hydrocele	P,Q,R	1/2	1a	4.0	41	_
08-501	Idiopathic	P,Q,R	1/2	1a	2.70	60	_
08-504	Idiopathic	P,Q,R	3/4	1b	1.0	46	3.5
08-533	Bilateral Migrating Testis	P,Q,R	1/2	1a	15	40	2
10-079	Idiopathic	P,Q,R	1/2	1a	2.0	12	3
11-041	Idiopathic	P,Q,R	1/2	1a	7.3	32	0
12-049	Idiopathic	P,Q,R	1/2	1b	0	0	
Normozoospermic men with gr/gr deletion							
07-367		P,Q,R	3/4	1a	104	40	21
10-342		P,Q,R	1/2	1a	27	55	6
FT 05-62		P,Q,R	3/4	1a	81	44	_
FT 03-33		K,L,M,NO, S,T / N	1/2	1a	100	50	_
4993_C		C,F,G,H,I	1/2	1b	128.9	87.7	_
Infertile men with b2/b3 deletions							
08-205	Bilateral Varicocele grade I	C,F,G,H,I	3/4	1b	2.2	19	_
10-259	Unilateral Varicocele grade I	N	3/4	1b	10	30	3
12-058	Unilateral Varicocele grade II	C,F,G,H,I	3/4	1b	6	5	0
12-511	Idiopathic	C,F,G,H,I	3/4	1b	2	11	1

Supplementary Table 3. Continued

Code	Phenotype	Y hgr	Deleted gene copy		Sperm parameters		
			DAZ	*CDY1*	n.of sp/ml x 10^6	Motility A+B (%)	Normal morphology (%)
Infertile men with partial AZFc duplication							
06-201	Idiopathic	P,Q,R			0	0	_
04-272	Idiopathic	K,L,M,NO, S,T / N			0.04	36.7	_
07-450	Bilateral Varicocele grade I	P,Q,R			2.83	13.5	4
08-077	Idiopathic	J			2.06	12	9
08-261	Idiopathic	C,F,G,H,I			1.3	24.5	5
09-023	Unilateral Varicocele grade III	P,Q,R			0.03	14	_
09-050	Idiopathic	C,F,G,H,I			1.4	38	17
09-124	Idiopathic	P,Q,R			5.6	25.3	8.5
09-195	Idiopathic	P,Q,R			0	0	_
09-332	Idiopathic	J			2.23	15.6	4
09-399	Idiopathic	K,L,M,NO, S,T / N			1.6	24.3	3
09-501	Idiopathic	J			1.9	22	0
09-532	Idiopathic	K,L,M,NO, S,T / N			4.5	10	_
10-430	Idiopathic	C,F,G,H,I			6.7	27	2
11-574	Trauma	C,F,G,H,I			0	0	
Normozoospermic men with partial AZFc duplication							
09-013		J			118	57	30
09-034		K,L,M,NO, S,T / N			40	70	13
09-479		P,Q,R			45	54	21
FT 03-52		P,Q,R			34	72	_
08-271		P,Q,R			34	56	13
FT 04-23		C,F,G,H,I			38	20	_
FQ 01-528		P,Q,R			22	26	10
12-309		P,Q,R			43	37	30
12-337		C,F,G,H,I			86	42	10
3612		P,Q,R			101.3	68.1	_
3699		P,Q,R			123.7	60.3	_
3718		J			206.7	94.1	_
4470		J			236.6	68.7	_

Supplementary Table 4. Number of Y microdeletion carriers with SC within 0-2x10^6/ml and 2-5x10^6/ml in different populations.

Country	N.of patients	N.of *AZF* del. carriers	SC (x10^6/ml) range		Reference
			0<SC≤2	2<SC<5	
China	295	14	14	0	[42]
Italy	3073	99	97	2	[31]
German	3179	39	39	0	[1]
Sri Lanka	96	7	7	0	[43]
Romania	30	2	2	0	[6]
Slovenia	226	7	7	0	[34]
Denmark	342	9	9	0	[44]
Spain	128	6	6	0	[8]
Australia	50	2	2	0	[45]
France	131	2	2	0	[5]
Spain	186	9	8	1	[7]
Subtotal	***7870***	***196***	***193***	***3***	
Spain	806	21	21	0	This study
Total	**8676**	**217**	**214**	**3**	

Only studies in which *AZF* deletion carriers had an explicitly indicated sperm count are considered.
SC: Sperm Concentration.
* Only “genuine” and complete deletions are considered.

Supplementary Table 5. Sperm recovery upon TESE and pregnancy rate in *AZF*c deletion carriers reported in the literature. Only men with *AZF*c deletion who underwent TESE and for whom testis biopsy is available are considered.

Biopsy method	*AZF*c del.carriers	Findings in testis histology	*AZF*c deleted men with sp+ at TESE (%)	N of achieved pregnancy/sp+ men (%)	Reference
Classical TESE	11	**4** complete SCOS (TYPE1); **4** SCOS TYPE2; **1** MA TYPE2; **2** HS TYPE2	1 (9.09%)	n.p	This study
Not described	7	All SCOS	1 (14.3%)	n.r	[33]
Bilateral testicular FNA. FNA+bilateral TESE when FNA negative	25	**16** SCOS; **7** HS; **2** MA (SC)	7 (28%)	n.r	[31]
MicroTESE	32	n.r	18 (56.2%)	n.r	[35]
MicroTESE	10	**1** MA (SC)+SCOS; **1** reduced spermatogenesis+MA (SC); **1** MA (RS); **1** SCOS+MA (SG); **1** MA (SC); **1** MA (RS); **1** complete SCOS; **1** Reduced spermatogenesis+ complete SCOS; **1** Reduced spermatogenesis+Residual spermatogenesis; **1** MA (SC)+ Residual spermatogenesis.	6 (60%)	0/2 (0%)	[1]
MicroTESE	6	**1** with 95% SCOS; **1** MA; **1** SCOS+MA (n.s); **3** n.p	5 (83.3%)	4/5 (80%)	[46]
MicroTESE	3	**2** MA (n.s)+HY; **1** SCOS;	2 (66.6%)	0/1 (0%)	[34]
MicroTESE	20	**9** SCOS; **7** MA (n.s); **4** SCOS+MA	14 (70%)	2/12* (16.6%)	[47]
MicroTESE	21	n.r	15 (71.4%)	10/15 (66.7%)	[36]
As described in Devroy et and Silber et al (1995a &1996)	10	**5** SCOS; **3** MA (n.s); **2** SCOS+MA	8 (80%)	(20%)	[48]
Multiple TESE	17	**5** SCOS; **11** MA (n.s); **1** HS	10 (59%)	6/9 (66.6%)	[49]
Total	**142**		**74 (52%)**		

sp+: spermatozoa present in testis
FNA: Fine Needle Aspiration

SCOS: Sertoli Cell Only Syndrome; HS: Hypospermatogenesis; MA: Maturation Arrest; SC:Spermatocytes; RS:Round Spermatids; SG:Spermatogonia; HY:Germinal Hypoplasia
n.r: **n**ot **r**eported; n.s: maturation stage **n**ot **s**pecified; n.p: **n**ot **p**erformed
*: poor oocyte quality noted in 2 cases in which no pregnancy was achieved

Supplementary Table 6. Frequency of *gr/gr* deletion in Spanish population. Data from de Llanos et al 2005 are included.

Phenotype	*gr/gr* N/tot (%)	P1 [OR1 (95% confidence interval)]	P2 [OR2 (95% confidence interval)]
Patients	24/585 (4.1)		
Controls (1)	5/393 (1.2)	0.007 [3.225 (1.241-8.380)]	
Controls (2)	5/591 (0.8)		2.1×10^{-4} [4.849 (1.863-12.623)]

Controls (1): Only normozoospermic subjects are considered
Controls (2): Whole control cohort

References Supplementary Table 4 and Supplementary Table 5

42. Ng PP, Tang MH, Lau ET *et al*: Chromosomal anomalies and Y-microdeletions among Chinese subfertile men in Hong Kong. *Hong Kong Med J.* 2009 **15**: 31–38.

43. Fernando L, Gromoll J, Weerasooriya TR, Nieschlag E, Simoni M *et al* :Y-chromosomal microdeletions and partial deletions of the Azoospermia Factor c (AZFc) region in normozoospermic, severe oligozoospermic and azoospermic men in Sri Lanka. *Asian J Androl. 2006;* **8**:39–44 (2006).

44. Krausz C, Rajpert-De Meyts E, Frydelund-Larsen L, Quintana-Murci L, McElreavey K, Skakkebaek NE *et al:* Double-blind Y chromosome microdeletion analysis in men with known sperm parameters and reproductive hormone profiles: microdeletions are specific for spermatogenic failure. *The J Clin Endocrinol Metab* 2001*;* **86**:2638–2642.

45. Cram DS, Ma K, Bhasin S *et al*: Y chromosome analysis of infertile men and their sons conceived through intracytoplasmic sperm injection: vertical transmission of deletions and rarity of de novo deletions. *Fertil Steril* 2000*;* **74**: 909–915.

46. Choi JM, Chung P, Veeck L, Mielnik A, Palermo GD, Schlegel PN *et al*: AZF microdeletions of the Y chromosome and in vitro fertilization outcome. *Fertil Steril* 2004; 81:337-341

47. Oates RD, Silber S, Brown LG, Page DC *et al*: Clinical characterization of 42oligospermic or azoospermic men with microdeletion of the AZFc region of the Y chromosome, and of 18 children conceived via ICSI. *Hum Reprod* 2002; 17:2813-2824.

48. Silber SJ, Alagappan R, Brown LG, Page DC *et al*: Y chromosome deletions in azoospermic and severely oligozoospermic men undergoing intracytoplasmic sperm injection after testicular sperm extraction. *Hum Reprod* 1998; 13:3332-3337.

49. Stouffs K, Lissens W, Tournaye H, Van Steirteghem A, Liebaers I: The choice and outcome of the fertility treatment of 38 couples in whom the male partner has Yq microdeletion. *Hum Reprod* 2005; 7:1887-1896

PAPER 4

High Resolution X Chromosome-Specific Array-CGH detects New CNVs in infertile males.

Krausz C, Giachini C, Lo Giacco D, Daguin F, Chianese C, Ars E, Ruiz-Castane E, Forti G, Rossi E. *PLoS One*. 2012;7:e44887. (PMID: 23056185).

SUMMARY 4

Background:
The etiology of altered spermatogenesis remains unknown in about 40% of cases (''idiopathic infertility'') and it is likely that a large proportion of them are caused by still unknown genetic factors. CNVs have progressively drawn the attention in relation to various complex human diseases, whereas their role in male infertility is poorly defined. In fact only Y chromosome-linked CNVs have been demonstrated as important contributors to impaired sperm production. Although it has been predicted that the X chromosome is also enriched in spermatogenesis genes, no clinically relevant gene mutations have been identified so far.

Aims:
In order to advance our understanding of the role of X-linked genetic factors in male infertility, we used a high resolution X chromosome specific array-CGH (a-CGH) approach to identify novel X-linked variants potentially linked to male infertility.

Materials and Methods:
A total of 96 infertile men with different grade of spermatogenic impairment (49 azoospermic, 25 cryptozoospermic and 22 oligozoospermic men) and 103 normozoospermic men were analyzed with a high resolution a-CGH platform customized for the X chromosome. Selected patient-specific deletions and losses were validated by PCR (+/-) or quantitative PCR and then screened in a case/control setting (359 infertile vs 370 normozoospermic men).

Results:
We identified 73 CNVs (29 losses and 44 gains) the majority of which (75.3%) were novel, providing the largest collection of X-linked CNVs in relation to spermatogenesis. The majority of gains (63%) were found also in the normozoospermic group whereas deletions were less frequent in controls since only 38% were present also in this group. One of the most relevant findings of our study is represented by the significantly higher global burden of deletions in patients compared to controls due to an excessive rate of deletions/person (0.57 versus 0.21, respectively; $p = 8.785 \times 10^{-6}$) and to a higher mean sequence loss/person (11.79 Kb and 8.13 Kb, respectively; $p = 3.435 \times 10^{-4}$). A significantly lower sperm concentration and total sperm count was found in patient with >1 CNV with respect to those with ≤1 CNV.
We found 12 patient-specific deletions with potential clinical implication; more than 90% of them were private or rare (frequency <1%). X-linked Cancer Testis Antigen (CTA) genes were recurrently affected.

Conclusions:
We observed a significantly increased deletion burden in relation to spermatogenic impairment, which suggests its involvement in the etiology of spermatogenic impairment. We observed the lack of highly recurrent deletions on the X chromosome and identified a number of potentially important patient-specific that represent novel targets for future investigations.
X -linked CTA family members are recurrently affected and their dosage variation may play a role in X-linked CNV-related spermatogenic failure.

OPEN ACCESS Freely available online

PLOS ONE

High Resolution X Chromosome-Specific Array-CGH Detects New CNVs in Infertile Males

Csilla Krausz[1,2]*, Claudia Giachini[1], Deborah Lo Giacco[2,3], Fabrice Daguin[1], Chiara Chianese[1], Elisabet Ars[3], Eduard Ruiz-Castane[2], Gianni Forti[4], Elena Rossi[5]

1 Unit of Sexual Medicine and Andrology, Molecular Genetic Laboratory, Department of Clinical Physiopathology, University of Florence, Florence, Italy, **2** Andrology Service, Fundació Puigvert, Barcelona, Spain, **3** Molecular Biology Laboratory, Fundació Puigvert, Universitat Autònoma de Barcelona, Barcelona, Spain, **4** Endocrinology Unit, Department of Clinical Physiopathology, University of Florence, Florence, Italy, **5** Biology and Medical Genetics, University of Pavia, Pavia, Italy

Abstract

Context: The role of CNVs in male infertility is poorly defined, and only those linked to the Y chromosome have been the object of extensive research. Although it has been predicted that the X chromosome is also enriched in spermatogenesis genes, no clinically relevant gene mutations have been identified so far.

Objectives: In order to advance our understanding of the role of X-linked genetic factors in male infertility, we applied high resolution X chromosome specific array-CGH in 199 men with different sperm count followed by the analysis of selected, patient-specific deletions in large groups of cases and normozoospermic controls.

Results: We identified 73 CNVs, among which 55 are novel, providing the largest collection of X-linked CNVs in relation to spermatogenesis. We found 12 patient-specific deletions with potential clinical implication. Cancer Testis Antigen gene family members were the most frequently affected genes, and represent new genetic targets in relationship with altered spermatogenesis. One of the most relevant findings of our study is the significantly higher global burden of deletions in patients compared to controls due to an excessive rate of deletions/person (0.57 versus 0.21, respectively; $p = 8.785 \times 10^{-6}$) and to a higher mean sequence loss/person (11.79 Kb and 8.13 Kb, respectively; $p = 3.435 \times 10^{-4}$).

Conclusions: By the analysis of the X chromosome at the highest resolution available to date, in a large group of subjects with known sperm count we observed a deletion burden in relation to spermatogenic impairment and the lack of highly recurrent deletions on the X chromosome. We identified a number of potentially important patient-specific CNVs and candidate spermatogenesis genes, which represent novel targets for future investigations.

Citation: Krausz C, Giachini C, Lo Giacco D, Daguin F, Chianese C, et al. (2012) High Resolution X Chromosome-Specific Array-CGH Detects New CNVs in Infertile Males. PLoS ONE 7(10): e44887. doi:10.1371/journal.pone.0044887

Editor: Laszlo Orban, Temasek Life Sciences Laboratory, Singapore

Received February 29, 2012; **Accepted** August 15, 2012; **Published** October 9, 2012

Funding: This work was supported by the Tuscan Regional Health Research Program (“Progetto Salute 2009”) to GF, the Italian Ministry of University (grant PRIN 2010-2012 to CK), the Italian Telethon (Grant GGP08204 to CK) and the Spanish Ministry of Health (grant FIS-08/1138, grant FIS-11/02254). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: c.krausz@ dfc.unifi.it

Introduction

Male factor infertility affects about 7% of men in the general population and the etiology of altered spermatogenesis remains unknown in about 40% of cases (“idiopathic infertility”) and it is likely that a large proportion of them are caused by still unknown genetic factors [1]. Nevertheless, besides abnormal karyotype and Y chromosome microdeletions no other recurrent genetic anomalies have been identified in men with primary testicular failure, raising questions about the appropriateness of the investigative approaches used so far [2–4]. The first innovative study applying whole-genome analysis of SNPs and the successive follow-up study failed in leading to the identification of recurrent genetic factors with large effect size [5,6]. Recently, high resolution array Comparative Genomic Hybridisation (array-CGH) studies identified new spermatogenesis candidate genes on autosomes and on the X chromosome and some recurring and private patient-specific CNVs with potential clinical interest [7,8].

Both sex chromosomes are enriched with genes prevalently or exclusively expressed in the testis [9,10]. Nevertheless, only Y chromosome-linked Copy Number Variants (CNVs) and Y-linked genes have been demonstrated as important contributors to impaired sperm production in humans [for review see [11,12]). In particular, the so called AZoospermia Factor (AZF) regions on the Yq have been found deleted in about 5–10% of azoospermic men (absence of spermatozoa in the ejaculate) and 2–5% of severe oligozoospermic men (<5 millions spermatozoa in the ejaculate). Data on the potential role of X-linked gene products in spermatogenesis derive mainly from model organisms and a higher than expected number of X-linked spermatogenesis genes have been identified [10,13]. The apparent paucity of information in humans is probably related to the scarcity of X-linked genes studied (only eight), none of which yet described as causative, except for the *AR* gene [14]. Similarly, the question whether the X chromosome contains AZF-like regions has not been sufficiently explored so far.

In order to advance the understanding of the role of X-linked CNVs and genes in male infertility, we applied an innovative approach based on high resolution X chromosome specific array-CGH. Given that such a detailed analysis of the X chromosome has not been published until now and the testicular function of subjects included in the Genomic Variant Database is unknown (except for 30 X-linked CNVs (23 duplications and 7 deletions) reported in the recent paper by Tuttelmann et al. [7]), ours is the first study providing a detailed analysis of X-linked losses and gains in several hundred subjects with known sperm parameters.

Materials and Methods

Subjects

The local Ethical Committees of the University Hospital Careggi and the Fundació Puigvert approved the study. All participants signed an informed consent. We analyzed with array-CGH 96 idiopathic infertile subjects with different grade of spermatogenic impairment (49 azoospermic, 25 cryptozoospermic and 22 oligozoospermic men) and 103 normozoospermic men. Infertile patients were selected on the basis of a comprehensive andrological examination including medical history, semen analysis, scrotal ultrasound, hormone analysis, karyotype and Y chromosome microdeletion screening. Patients with mono- or bilateral cryptorchidism, varicocele grades 2 and 3, obstructive azoospermia, recurrent infections, iatrogenic infertility, hypogonadotrophic hypogonadism, karyotype anomalies, Y chromosome microdeletions including partial deletions of the *AZFc* region, and partial *AZFc* duplications and patients with non-Italian or non-Spanish origin were excluded. Testis histology was available for 47 men. Controls in the Spanish cohort were fertile normozoospermic men undergoing pre-vasectomy, whereas the Italian control cohort included normozoospermic volunteers not belonging to infertile couples (60% with proven fertility). The ethnic/geographic composition was similar in the control and patient groups (40% Spanish and 60% Italians).In the second part of the study, we performed a case-control association study reaching a total of 359 patients and 370 normozoospermic controls on 13 selected CNVs which appeared to be specific to infertile men based on the array-CGH analysis. Detailed phenotypic data relative to the study populations are provided in Table 1.

Methods

Germline DNA was extracted from peripheral blood samples in all the participants with standard methods.

Array-CGH. Customized array-CGH platforms (custom 8×60 K, Agilent Technologies, Santa Clara, CA, USA) were generated using the eArray software (http://earray.chem.agilent.com/); 53069 probes (60-mer oligonucleotides) were selected from those available in the Agilent database and cover the whole chromosome X, including Xp and Xq pseudoregions, with a medium resolution of 4 Kb. Four replicate probe groups, with every probe present in two copies on the platform, were designed in regions containing mouse infertility-associated genes i.e. sperm protein associated with the nucleus, X-linked family members (*SPANX*); testis expressed 11 *TEX11*, TAF7-like RNA polymerase II, TATA box binding protein (TBP)-associated facto *(TAF7L) and*). In these regions, the medium resolution is 2 Kb. The array also included, for the normalization of copy number changes, Agilent control clones spread along all autosomes (6842 probes). As a reference DNA, we used the same normozoospermic subject for all the study population. This control DNA was already characterized for CNV content in previous array-CGH experiments against eight different normospermic controls and presented one private gain of 27 Kb mapping to Xcentr which was not considered for the frequency analyses. 300 ng of test DNA and control DNA were double-digested with RsaI and AluI (Promega) for 1 hour at 37°C. After digestion, samples were incubated at 65°C for 20 minutes to inactivate the enzymes, and then labeled by random priming (Agilent Technologies) for 2 hours using Cy5-dUTP for the test DNA and Cy3-dUTP (Agilent Technologies) for the control DNA. Labeled DNAs were incubated at 65°C for 10 minutes and then purified with Microcon YM-30 filter units (Millipore, Billerica, USA). Every purified sample was brought to a total volume of 9.5 µl in 1xTE (pH 8.0, Promega), and yield and specific activity were determined for each sample using a NanoDrop ND-1000 UV-VIS Spectrophotometer (Labtech International LTD). The appropriate cyanine 5- and cyanine 3-labeled samples were combined in a total volume of 16 µl. After sample denaturation and pre-annealing with 5 ul of Human Cot-1 DNA (Invitrogen, Carlsbad, CA), hybridization was performed at 65°C with shaking for 24 hours. After two washing steps, the array was analyzed through the Agilent scanner and the Feature Extraction software (v10 1.1.1). Graphical overview was obtained using the DNA Analytics (v4.0.73). All the array experiments were analysed using the ADM-2 algorithm at threshold 5. Aberrant signals including 3 or more adjacent probes were considered as genomic CNVs (Figure S1). The positions of oligomeres refer to the Human Genome March 2006 assembly (hg18). All experimental data was submitted to GEO repository with the following Series accession number: GSE37948.

Molecular genetic analyses for confirmation of array-CGH data and for the case-control study

Molecular analysis of deletions. For the first step screening as for the confirmatory step, we performed PCR protocol in a final volume of 10 µl containing 70 ng of genomic DNA, 3 mM MgCl2, 400 µM deoxynucleotides triphosphates, 10 pmol of specific primers, 50 U/ml of Taq DNA Polymerase (Promega PCR MASTER MIX 2X). All the primers for the first step screening had an optimal annealing temperature between 58–60°C and suspected deletions were further confirmed by i) lowering the annealing temperature (55°C); ii) performing additional PCRs with alternative primers (see details in the Table S1).

Molecular analysis of gains and the loss CNV31. Gains and loss CNV31 screening were performed using pre-designed TaqMan® Copy Number Assays or Custom TaqMan® Copy Number Assays (Applied Biosystems). All assays were conducted using three or four replicates for each sample (on the basis of the assay quality), in a final volume of 20 ul according to the manufacturer's instructions. The reaction mix components were: 1X TaqMan® Genotyping Master Mix, 1X TaqMan® Copy Number Assay, 1X TaqMan® Copy Number Reference Assays, 10 ng of genomic DNA. Briefly, the TaqMan® Copy Number Assay – containing two specific primers and a FAMTM dye-labeled MGB probe to detect the genomic DNA target sequence – is run in duplex with the TaqMan® Copy Number Reference Assays – containing two primers and a VIC® dye-labeled TAMRATM probe to detect the genomic DNA reference sequence. On each plate the same normozoospermic control used as reference DNA for array-CGH experiments (calibrator sample), the DNA sample of the CNV carrier and the No Template Control (NTC) were run. The CopyCaller SoftwareTM was used for post-PCR data analysis for all the copy number quantitation experiments. Information about qPCR probes are provided in Table S2.

Table 1. Clinical description of the study population.

A

SPERM COUNT		PATIENTS (n = 359)	CONTROLS (n = 370)
Total sperm count (10^6)	median (25^{th}–75^{th} percentile)	2.6 (0.00–13.62)	263.20 (159.00–405.50)
	mean ± SD	8.77±12.72	311.79±199.99
Sperm concentration (10^6/ml)	median (25^{th}–75^{th} percentile)	0.90 (0.00–4.40)	76 (50.00–117.50)
	mean ± SD	2.56±3.27	91.32±59.64

B

GEOGRAPHIC ORIGIN	SEMEN PHENOTYPE	ARRAY-CGH STUDY (n = 96)	CASE-CONTROL STUDY (n = 263)	TOTAL (n = 359)
ITALIAN (n = 233)	azoospermic	15	42	57
	cryptozoospermic	20	17	37
	oligozoospermic	22	117	139
	severe oligozoospermic ($<5\times10^6$/ml)	*17*	*47*	-
	moderate oligozoospermic($<20\times10^6$/ml)	*0*	*12*	-
SPANISH (n = 126)	azoospermic	34	15	49
	cryptozoospermic	5	36	41
	oligozoospermic	0	36	36
	severe oligozoospermic ($<5\times10^6$/ml)	*0*	*24*	-
	moderate oligozoospermic($<20\times10^6$/ml)	*0*	*12*	-

C

HORMONAL PARAMETERS	ARRAY-CGH STUDY (n = 96)	CASE-CONTROL STUDY (n = 263)	TOTAL (n = 359)
FSH (U/L)			
Mean ± SD	13.40±9.09	11.13±10.08	11.74±9.84
Median (25^{th}–75^{th} percentile)	11.38 (5.30–19.0)	8.10 (4.27–14.30)	8.70 (4.40–15.0)
LH (U/L)			
Mean ± SD	5.02±2.34	5.46±4.02	5.36±3.72
Median (25th–75th percentile)	4.45 (3.30–6.21)	4.60 (3.20–6.75)	4.60 (3.20–6.70)
TESTOSTERONE (ng/ml)			
Mean ± SD	5.18±2,.94	4.49±2.08	4.64±2.30
Median (25th–75th percentile)	4.80 (3.51–6.40)	4.20 (3.40–5.20)	4,.20 (3.40–5.40)
TESTIS VOLUME (ml)			
Mean ± SD	11.55±4.55	12.89±4.32	12.54±4.42
Median (25th–75th percentile)	11.50 (8.0–14.0)	13.0 (9.0–14.87)	13.0 (9.0–14.50)

A) Semen phenotype of the entire study population (array-CGH and case-control study); B) Description of all analyzed patients (array-CGH and case-control study) according to their geographic origin and semen phenotype; C) Hormonal levels and testis volumes of all analyzed patients (array-CGH and case-control study).
doi:10.1371/journal.pone.0044887.t001

Statistical analysis

Statistical analyses were performed using the statistical package SPSS (version 17.0.1, Chicago, IL, USA). Non-parametric Mann-Whitney U test was performed for comparisons of: i) median values of CNV number and DNA change between patients and controls; ii) median values of sperm concentration and total sperm count in relationship with CNV number. Frequencies were compared by Fisher exact test.

Results

Characterization of X-chromosome linked CNVs

We performed a high resolution array-CGH analysis using a microarray containing probes densely covering the complete human X chromosome (average resolution: 4 kb). Of the 199 subjects analyzed (96 idiopathic infertile subjects and 103 normozoospermic men), 97 (36 patients and 61 controls) showed the lack of CNVs, whereas the remaining 102 samples were found to carry 73 CNVs (44 gains and 29 losses) (Tables 2, 3, and 4). Thirty-two CNVs intersected genes/transcription units based on data available in genomic databases. As shown in Figure 1, CNVs were evenly distributed along the X chromosome with higher density in the PAR1.

Since homologous sequences at the border of a CNV may act as a substrate for non-allelic homologous recombination (NAHR), we checked the nature of regions flanking (between the minimum-maximum size of the CNV and approximately up to 1 Mb from the maximum size) the identified CNVs in order to understand whether NAHR is likely to occur (UCSC Genome Browser). Highly homologous sequences were identified only in 19% of CNVs, indicating that NAHR is not involved in the majority of observed CNVs. This figure was concordant with other observa-

Table 2. List of the 31 patient-specific (not found in normozoospermic controls) CNVs detected by array-CGH and their description according to type, gene location (*NO* = no gene found within) and occurrence in the Database of Genomic Variants (DGV).

CNV type	CNV code	Region	Size (Kb)	Start position	End position	Coding sequences within the CNV*	DGV	Frequency
LOSSES	**17**	Xp22.31	31.70	6,594,834	6,626,533	*NO*		1
	18	Xp22.31	82.00	6,756,310	6,838,310	*NO*	Variation_8908 Variation_53018 Variation_34619	1
	22	Xp22.11	24.35	22,969,648	22,993,997	*NO*		1
	23	Xp21.3	6.69	25,274,024	25,280,712	*NO*		1
	24	Xp21.3	67.33	26,891,769	26,959,101	*NO*		1
	31	Xp11.23	81.13	47,766,391	47,847,516	*ZNF630*	Variation_96640 Variation_9861 Variation_53005	1
	32	Xp11.22	4.15	52,065,798	52,069,943	*NO*		1
	50	Xq22.1	45.36	101,803,578	101,848,935	*ARMCX5-GPRASP2*		1
	54	Xq24	44.85	118,281,024	118,325,874	*NO*		1
	56	Xq25	86.07	124,632,886	124,718,959	*NO*		1
	57	Xq25	188.03	124,929,673	125,117,699	*NO*	Variation_52924	1
	61	Xq27.2	4.67	140,773,893	140,778,561	*MAGEC3*		1
	66	Xq27.3	7.35	145,030,566	145,037,917	*NO*		1
	67	Xq28	5.42	148,456,474	148,461,889	*NO*		1
	71	Xq28 (PAR)	290.99	154,586,913	154,877,901	*SPRY3 VAMP7*		1
GAINS	**1a**	Xp22.33 (PAR)	224.83	1,544	226,372	*PLCXD1 GTPBP6 PPP2R3B*		1
	5	Xp22.33 (PAR)	160.10	302,644	462,740	*NO*		1
	11	Xp22.33 (PAR)	39.73	1,347,599	1,387,328	*CSF2RA*		1
	14	Xp22.33 (PAR)	1.40	1,896,197	1,897,608	*NO*	Variation_31542	2
	19	Xp22.31	245.03	7,002,649	7,247,676	*STS HDHD1A*		1
	20	Xp22.2	602.10	11,104,518	11,706,614	*ARHGAP6 AMELX MSL3*		1
	25	Xp21.3	60.40	27,277,529	27,337,933	*NO*		1
	26	Xp21.1	88.57	37,168,387	37,256,960	*PRRG1*		1
	27	Xp21.1	9.61	37,242,364	37,251,969	*NO*		1
	30	Xp11.3	81.13	47,766,391	47,847,516	*ZNF630*	Variation_53003 Variation_9343 Variation_83491	1
	38	Xq13.2	8.98	72,202,996	72,211,976	*NO*		1
	39	Xq21.1	21.98	76,992,067	77,014,050	*MAGT1*		1
	40	Xq21.1	5.28	80,112,246	80,117,526	*NO*	Variation_83611	2
	55	Xq24	206.90	118,691,020	118,897,917	*SEPT6 ANKRD58 RPL39 SNORA69 UPF3B RNF113A NDUFA1*		1
	58	Xq25	9.68	125,143,278	125,152,957	*NO*		1
	60	Xq26.3	42.30	134,585,636	134,627,936	*NO*	Variation_52936	1

*CNV minimum size.

doi:10.1371/journal.pone.0044887.t002

Table 3. List of the 33 control-specific (not found in idiopathic patients) CNVs detected by array-CGH and their description according to type, gene location (*NO* = no gene found within) and occurrence in the Database of Genomic Variants (DGV).

CNV type	CNV code	Region	Size (Kb)	Start position	End position	Coding sequences within the CNV*	DGV	Frequency
LOSSES	**5.B**	Xp22.33	12.63	701,071	713,696	NO		1
	25.A	Xp21.2	9.69	31,282,923	31,292,613	*DMD*		1
	25.B	Xp21.1	28.26	33,953,232	33,981,492	*NO*	Variation_7783	2
	33.A	Xp11.21	58.89	56,403,390	56,462,278	*NO*		1
	53.A	Xq24	170.73	118,278,913	118,449,646	*SLC25A43*		1
	58.A	Xq25	12.68	125,198,109	125,210,792	*NO*		1
	60.A	Xq26.3	50.84	134,801,361	134,852,198	*SAGE1*		1
	60.D	Xq27.1	217.83	140,175,103	140,392,930	*NO*		1
	66.A	Xq28	37.12	147,393,583	147,430,698	*AFF2*		1
	71.A	Xq28 (PAR)	122.36	154,755,542	154,877,901	*VAMP7*		1
GAINS	**4.A**	Xp22.33 (PAR)	237.08	153,373	390,452	*PLCXD1 GTPBP6 PPP2R3B*		1
	5.A	Xp22.33 (PAR)	241.98	674,222	916,206	*NO*		1
	5.C	Xp22.33 (PAR)	420.72	747,358	1,168,080	*NO*		1
	12.A	Xp22.33 (PAR)	6.61	1,693,897	1,700,511	*ASMT*		1
	12.B	Xp22.33 (PAR)	683.74	1,716,023	2,399,766	*ASMT DHRSX*		1
	15.A	Xp22.33 (PAR)	27.94	2,382,699	2,410,643	*DHRSX*	Variation_83270	1
	15.B	Xp22.33/22.32	280.09	4,206,493	4,486,580	*NO*		1
	16.A	Xp22.31	1609.42	6,487,238	8,096,662	*HDHD1 STS VCX PNPLA MIR651*		1
	19.A	Xp22.31	129.96	7,961,788	8,091,751	*MIR651*	Variation_9337	1
	19.B	Xp22.31	177.54	8,411,159	8,588,699	*KAL1*		1
	20.A	Xp22.2	665.88	14,590,604	15,256,487	*GLRA2 FANCB MOSPD2 ASB9 ASB11 PIGA*		1
	20.B	Xp22.13	13.34	18,018,894	18,032,238	*NO*		1
	25.C	Xp21.1	185.02	34,931,807	35,116,827	*NO*		1
	25.D	Xp21.1	215.00	35,269,628	35,484,626	*NO*		1
	31.A	Xp11.23	78.87	48,021,982	48,100,848	*SSX3*		1
	34.A	Xp11.12	48.06	56,870,427	56,918,489	*NO*		1
	36.A	Xq11	716.03	63,925,948	64,641,977	*ZC4H2 ZC3H12B*		1
	38.A	Xq13.2	192.04	74,375,875	74,567,915	*UPRT ZDHHC15*	Variation_74012	1
	38.B	Xq13.3	153.23	75,123,387	75,276,621	*NO*		1
	55.A	Xq25	53.80	120,385,787	120,439,584	*NO*		1
	59.A	Xq26.3	24.69	134,151,039	134,175,725	*NO*		1
	60.B	Xq26.3	13.45	136,050,422	136,063,872	*NO*		1
	60.C	Xq26.3	91.26	137,089,527	137,180,783	*NO*		1

*CNV minimum size.

doi:10.1371/journal.pone.0044887.t003

tions reporting a similar frequency of potential NAHR targets [15]. It is interesting to note that in some areas (Xp11.12-q21.1) only duplications were found, whereas from Xq27.1-q27.3 only deletions were detected. One of the PAR1-linked losses (CNV15) was found in 23 patients and only once in controls (Figure 1b). This small CNV has already been described in the Database of Genomic Variants (DGV) both as loss and gain. This CNV was situated inside a 3914 bp Simple Tandem Repeat which included two Segmental Duplications (respectively of 1498 bp and 1444 bp) that therefore may act as substrate for NAHR. This mechanism may have lead also to reciprocal duplication and in fact CNV14, identified in our study, is the reciprocal duplication of CNV15. No genes were identified inside or nearby CNV14/15 which made it difficult to attribute a pathogenic role to this loss. Moreover, the same sequence was present also on the Y chromosome which further complicated the interpretation of the results.

Considering the size of detected CNVs, which ranged from 1.4 Kb to 1609 Kb (Tables 2, 3, and 4), we noticed that losses were typically of small/medium size and only 17% of them were large (Figure 2). Conversely, large gains represented 48% of the total CNVs and the difference between frequencies of losses and gains of >100 Kb was statistically significant ($p = 0,012$). Small

Table 4. List of CNVs found by array-CGH considering their occurrence in controls and in patients with their description according to type, gene location (NO = no gene found within) and presence in the Database of Genomic Variants (DGV).

	CNV code	CNV type	Region	Size (Kb)	Start position	End position	Coding sequences within the CNV*	DGV	Frequency in patients	Frequency in controls
Control-enriched CNVs	12	GAIN	Xp22.33 (PAR)	17.30	1,693,897	1,711,194	*ASMT*	Variation_83259	1	3
Patient-enriched CNVs	15	LOSS	Xp22.33 (PAR)	1.40	1,896,197	1,897,608	*NO*	Variation_104545	23	1
	64	LOSS	Xq27.3	3.92	143,436,347	143,440,268	*NO*	Variation_115340	8	5
	69	LOSS	Xq28	11.77	154,044,877	154,056,645	*NO*		7	3
Common CNVs	16	LOSS	Xp22.32	7.76	4,250,413	4,258,174	*NO*	Variation_52995	2	1
	35	GAIN	Xp11.1	117.14	57,318,438	57,435,573	*FAAH2*		1	1
	49	GAIN	Xq22.1	5.30	100,942,190	100,947,490	*NO*		2	2
	51	GAIN	Xq22.2	34.69	103,152,319	103,187,013	*H2BFWT H2BFM*	Variation_3254	5	4
	68	GAIN	Xq28	105.66	148,686,631	148,792,286	*MAGEA8*	Variation_31571	2	2

*CNV minimum size.

doi:10.1371/journal.pone.0044887.t004

CNVs (<10 Kb) were more frequently found in patients in respect to controls whereas large gains have been found mainly in controls (Figure 2).

According to the Database of Genomic Variants (DGV) website, losses/gains were divided into "known" and "novel", identifying 21 novel losses and 34 novel gains (Tables 2, 3, and 4). Among the 73 CNVs, 31 (15 losses and 16 gains) were found only in patients, "patient-specific" (Table 2) and 33 (10 losses and 23 gains) were found only in the control group, "control-specific" (Table 3). Of the remaining 9 CNVs, only one gain (CNV12) was found more frequently among controls whereas those resulting more frequent among patients ("patient-enriched") were deletions. The rest (4 gains and 1 deletion) were found to equally occur in both patients and controls (Table 4). These data suggest that gains are less likely to affect spermatogenesis since 63% of them (28/44) were found also in normozoospermic controls. On the contrary, deletions were less frequent in controls (11/29; 38%) indicating that in the presence of a deletion an abnormal sperm phenotype is more likely to occur. A general outline of the array-CGH findings with phenotypic description is provided in Table S3.

CNV burden

In order to assess the potential impact of CNVs in cases versus controls, we used two primary measures of CNV burden: the mean size and the mean number of CNVs/individual (Table 5A). The mean value of losses bp was significantly higher in patients than in controls (11.79 Kb and 8.13 Kb, respectively; $p = 3.435 \times 10^{-4}$). All losses were confirmed by PCR plus/minus or Real Time PCR, except for PAR-linked losses (n = 4), for which no suitable assay could be designed. The number of CNVs/person was significantly higher in patients compared to controls ($p = 0.002$) and depended on the overrepresentation of losses in the former group (0.57 versus 0.21; $p = 8.785 \times 10^{-6}$) (Table 5). CNV15, the most frequently found loss appears to be the major contributor to the deletion burden, however even without this loss the number of losses/person is significantly higher in the patient's group ($p = 0.041$). Phenotypic description of patients (loss-carriers and no CNV-carriers) is provided in Table S4. Although the frequency of patients with more than one CNV (n = 19; 19.8%) was nearly twice that of controls (n = 11; 10.7%), the difference did not reach statistical significance ($p = 0.078$). On the other hand, comparing the frequencies of subjects with ≥1 CNV in cases versus controls, we observed a highly significant difference when considering the total number of CNVs ($p = 0.003$) and of losses ($p < 0.001$) (Table 5B).

CNVs and semen parameters

A significant association with sperm concentration and total sperm number was observed among patients when considering the total CNV number (Table 6). Patients with more than 1 CNV had a significantly lower sperm concentration and total sperm count than those with ≤1 CNV ($0.2 \pm 0.6 \times 10^6$/ml versus $1.0 \pm 2.0 \times 10^6$/ml; $p < 0.022$; $2.3 \pm 4.6 \times 10^6$ versus $1.0 \pm 3.3 \times 10^6$; $p < 0.032$). The maximum number of CNVs/subject was three, and of the five patients with three CNVs four were azoospermic and one was severely oligozoospermic with <1 million spermatozoa/ejaculate (Table S6). All of them had at least one private CNV (uniquely found in this patient), and only one patient (07-170) shared two recurrent CNVs with two others (07-13, 07-30). Given that the selection of patients was based on the absence of known causes of spermatogenetic failure, subjects with multiple CNVs did not show any additional andrological anomaly or other relevant diseases. Semen parameters and testis histology of patients and controls with >1 CNVs are reported in Table S5, 6.

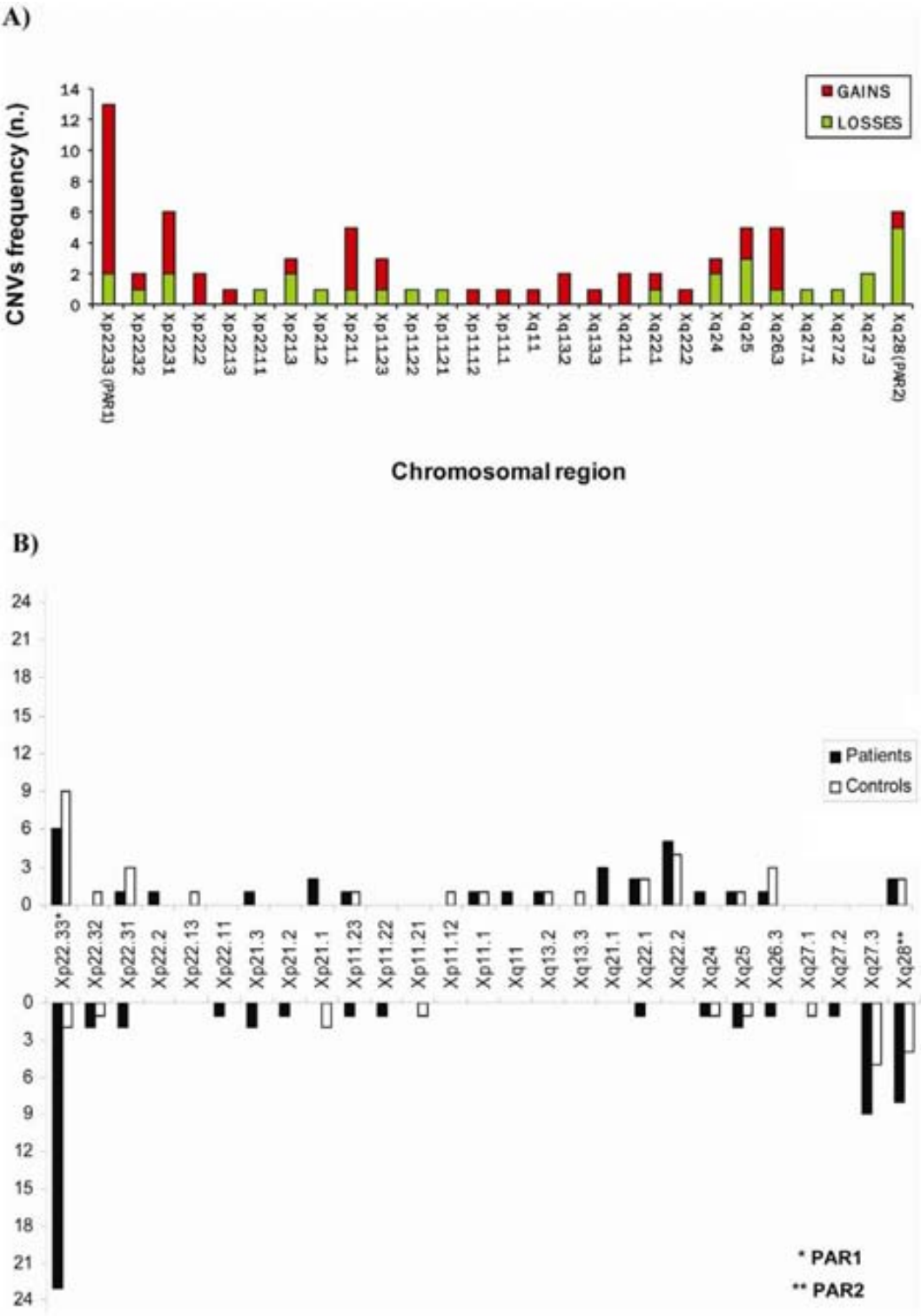
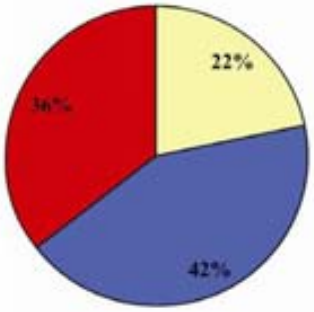


Figure 1. Schematic representation of the distribution of the 73 CNVs (44 gains and 29 losses) along the X chromosome identified by high resolution X chromosome specific array-CGH analysis. A) The histogram shows that the 73 CNVs were evenly distributed along the X chromosome but displayed a higher density in the pseudoautosomal region 1, PAR1 (Xp22.33). B) The frequency of gains (upwards) and losses (downwards) per X chromosome region in patients and controls are indicated.
doi:10.1371/journal.pone.0044887.g001

Screening for selected deletions

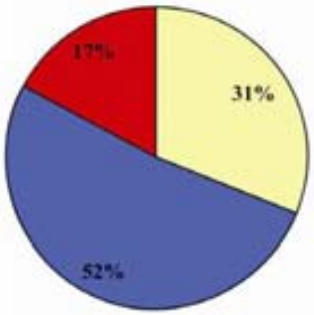
To further investigate the potential clinical implications of losses, 13 patient-specific deletions were subsequently screened in a large group of infertile and normozoospermic men: excluding CNV66, they all remained patient-specific (Table 7). Due to the rarity of the 12 patient-specific losses, statistically significant

A. Gains + Losses



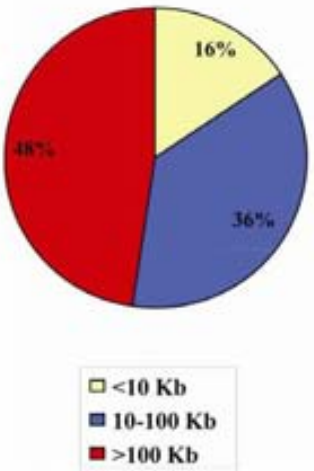
A. Gains + Losses (n=73)				
	<10kb	10-100 kb	>100 kb	Novel
Patient-specific	10	14	7	23
Control-specific	2	14	17	29
Patient-enriched	2	1	-	1
Control-enriched	-	1	-	-
Common	2	1	2	2

B. Losses



B. Losses (n=29)				
	<10kb	10-100 kb	>100 kb	Novel
Patient-specific	5	8	2	11
Control-specific	1	6	3	9
Patient-enriched	2	1	-	1
Control-enriched	-	-	-	-
Common	1	-	-	-

C. Gains



C. Gains (n=44)				
	<10kb	10-100 kb	>100 kb	Novel
Patient-specific	5	6	5	12
Control-specific	1	8	14	20
Patient-enriched	-	-	-	-
Control-enriched	-	1	-	-
Common	1	1	2	2

Figure 2. Array-CGH study.: distribution of the 73 CNVs according to their size: small (<10 Kb), medium (10–100 Kb) and large (>100 Kb) referred to A) all CNVs (44 gains and 29 losses); B) losses; C) gains. Losses were typically of small/medium size (52%) whereas gains are generally of larger size (48%). On the side, tables display the number of A) all CNVs; B) losses; C) gains of different size and categorized according to their occurrence in patients/controls: i) "patient-specific" when found only in patients; ii) "control-specific" when found only in controls; iii) "patient-enriched" when found predominantly in patients; iv) "control- enriched" when found predominantly in controls; v) "common" when found at a similar frequency in patients and controls.
doi:10.1371/journal.pone.0044887.g002

differences were not observed in their frequencies compared to the control group. In fact, 8/12 were private (found in a single individual) whereas only 4 were recurrent with a still relatively low frequency (0.5–1.1%).

Recurrent patient-specific CNVs. Among the patient-specific recurrent CNVs, three deletions are of major interest. CNV67, observed in 1.1% of patients may remove (considering its maximum size) the melanoma antigen family A, 9B (*MAGEA9B*), which belongs to the Cancer Testis Antigens (CTAs) gene family,

Table 5. Array-CGH study: Comparison between patients and controls of the mean number and mean extension of CNVs (A) as well as the number of all subjects bearing more than one CNV (B).

A	PATIENTS (n = 96)		CONTROLS (n = 103)		p	
	Mean CNV number ± sd	Mean CNV extension (Kb) ± sd	Mean CNV number ± sd	Mean CNV extension (Kb) ± sd	p1	p2
LOSSES+GAINS	0.87±0.85	36.21±85.4	0.54±0.76	73.87±222.08	2.095×10^{-3}	0.113
LOSSES	0.57±0.64	11.79±38.43	0.21±0.46	8.13±32.30	8.785×10^{6}	3.435×10^{-4}
GAINS	0.30±0.54	24.42±76.50	0.33±0.62	65.74±220.07	0.862	0.733

B	PATIENTS (n = 96)	CONTROLS (n = 103)	p	OR (95% CI)
≥1 CNV/subject	60	42	0.003	1.5 (1.2–2.0)
≥1 loss/subject	47	20	<0.001	2.5 (1.6–3.9)
≥1 gain/subject	25	28	0.874	-

sd = standard deviation. OR = odds ratio. CI = confidence interval. *p*1 refers to the mean number of CNV/subject. *p*2 refers to the mean DNA change/subject.
doi:10.1371/journal.pone.0044887.t005

expressed exclusively in the testis with the highest expression level in spermatocytes and in some tumour cell lines [16]. This deletion may also affect additional genes with prevalent or exclusive expression in the testis such as other CTAs and the following: transmembrane protein 185A (*TMEM185A*), chromosome X open reading frame 40A (*CXorf40A*), X linked heat shock transcription factor family (*HSFX*) all situated at <1 Mb from the deletion. Phenotypes of patients with this deletion ranged from azoospermia due to Sertoli Cell Only Syndrome (SCOS, [17]) to oligozoospermia. CNV 31 presents a reciprocal duplication (CNV30, Table 2) and was observed in 4 patients (two found by array-GH and two by qPCR) and 0/325 controls. CNVs 30/31 affect the dosage of zinc finger protein 630 (*ZNF630*), a gene with unknown function; however, considering their maximum extension, additional genes with exclusive expression in the testis such as the sperm acrosome associated 5 *SPACA5,/SPACA5b*) are also involved. CNV32 does not remove any gene directly, but it is situated within an area abundant in CTA genes. In order to define whether the underlying mechanism of these deletions is NAHR we analyzed the flanking regions. Only CNV 30/31 showed Segmental Duplications (SD) which may explain the recurrence of deletion/duplication events. Although also CNV67 was found in 4 patients, this deletion does not have a reciprocal duplication and it is not flanked by SDs. An alternative mechanism for the formation of CNV67 could be Non Homologous End Joining (NHEJ), since substrates for this mechanism are highly represented in this area (many LINE and Alu elements). However this hypothesis requires further confirmation by the fine mapping of the breakpoints.

Private patient-specific CNVs. Concerning private patient-specific deletions, which were found only in single patients, we observed two deletions directly affecting gene dosage. CNV50 removes the ARMCX5-GPRASP2 read-through (*ARMCX5-GPRASP2*) genes for which no testis expression data are available. The carrier of this deletion suffers from azoospermia due to SCOS. CNV61, observed in one azoospermic man, removes another CTA family member, the melanoma antigen family C, 3 *MAGEC3*. This deletion may also affect other neighbouring CTA genes, such as the melanoma antigen family C, 1 *MAGEC1* and Sperm protein associated with the nucleus, X-linked, family

Table 6. Array-CGH study: comparison of patients' semen parameters according to the number of CNVs.

PATIENTS (n = 96)				
	SPERM CONCENTRATION ($n\times10^6$/ml)	p	TOTAL SPERM NUMBER ($n\times10^6$)	p
0 CNV (n = 36)	1.2±2.4 (0.01; 0.0–12.0)		2.9±5.7 (0.01; 0.0–30)	
≥1 CNV (n = 60)	0.6±1.3 (0.0; 0.0–6.2)	0.068	1.6±3.4 (0.0; 0.0–17.4)	0.075
0 LOSS (n = 49)	1.0±2.1 (0.01; 0.0–12.0)		2.7±5.0 (0.01; 0.0–30.0)	
≥1 LOSS (n = 47)	0.6±1.5 (0.0; 0.0–6.2)	0.053	1.4±3.6 (0.0; 0.0–17.4)	0.051
0 GAIN (n = 71)	1.0±2.1 (0.0; 0.0–12.0)		2.4±4.9 (0.0; 0.0–30.0)	
≥1 GAIN (n = 25)	0.4±0.7 (0.0; 0.0–2.3)	0.185	1.3±2.3 (0.0; 0.0–6.4)	0.215
≤1 CNV (n = 77)	1.0±2.0 (0.0; 0.0–12.0)		2.3±4.6 (0.0; 0.0–30.0)	
>1 CNV (n = 19)	0.2±0.6 (0.0; 0.0–2.0)	0.022	1.0±3.3 (0.0; 0.0–13.4)	0.032
≤1 LOSS (n = 88)	0.9±1.9 (0.0; 0.0–12.0)		2.1±4.4 (0.0; 0.0–30.0)	
>1 LOSS (n = 8)	0.2±0.6 (0.0; 0.0–1.8)	0.230	1.7±4.7 (0.0; 0.0–13.4)	0.309
≤1 GAIN (n = 92)	0.8±1.9 (0.0; 0.0–12.0)		2.2±4.5 (0.0; 0.0–30.0)	
>1 GAIN (n = 4)	0.0±0.0 (0.0; 0.0–0.01)	0.293	0.0±0.0 (0.0; 0.0–0.01)	0.29

Sperm concentration and total sperm number are expressed as: mean ± standard deviation (median; range). Significance is depicted by a *p* value<0.05.
doi:10.1371/journal.pone.0044887.t006

Table 7. Case-control study of selected losses preliminarily identified by array-CGH as patient-specific (not found in normozoospermic controls).

CNV code	X chr. band	Patients (frequency)	Controls (frequency)	p value	Carriers code	Carriers phenotype (total sperm count $\times 10^{\wedge 6}$ and/or testis histology)	Genes inside and nearby (<500 Kb)
17	Xp22.31	2/359 (0.55%)	0/370 (0%)	0.244	A448, A828	A448: Oligozoosp. (13.4); A828: Oligozoosp. (20)	*NLGN4X, VCX3A*[§], *HDHD1A*[†], *STS*
18	Xp22.31	1/359 (0.27%)	0/370 (0%)	0.492	07-96	Azoosp. (0.0; SCOS)	*VCX3A*[§], *HDHD1A*[†], *STS*
22	Xp22.11	1/359 (0.27%)	0/370 (0%)	0.492	MMP718	Oligozoosp. (6.4)	*DDX53*[§], *PTCHD1*[†], *PRDX4*
23	Xp21.3	1/359 (0.27%)	0/370(0%)	0.492	A142	Cryptozoosp. (0.01)	*POLA1, SCARN23, ARX*[§]
31	Xp11.23	2/270 (0.74%)	0/325(0%)	0.206	A630, 09-126	A630: Oligozoosp. (7.2) 09-126: Oligozoosp. (8)	*ARAF, SYN1, TIMP1, CFP, ELK1, UXT, LOC100133957, CXXC1P1, ZNF81, ZNF182*[§], *SPACA5**, ***ZNF630,*** *SPACA5B**, *LOC100509575**, *SSX5*[†], *SSX1, SSX3, SSX4, SSX4B, SLC38A5,FTSJ1,PORCN,EBP, TBC1D25, RBM3, WDR13*
32	Xp11.22	2/359 (0.55%)	0/370(0%)	0.244	A162, 08-190	A162: Cryptozoosp. (0.01); 08-190 Azoosp (0.0)	*MAGED1, SNORA11D, SNORA11E, MAGED4B/MAGED4*[§], *XAGE2*[†], *XAGE2B, XAGE1B, XAGE1A, XAGE1D, XAGE1C, XAGE1E*
50	Xq22.1	1/359 (0.27%)	0/370 (0%)	0.492	07-22	Azoosp. (SCOS)	*NXF2B, NXF2, TMSB15A, NXF4 ARMCX5, GPRASP1*[§], ***ARMCX5-GPRASP2,*** *GPRASP2*[†], *BHLHB9, RAB40AL, BEX1, NXF3*
54	Xq24	1/359 (0.27%)	0/370 (0%)	0.492	MM550	Oligozoosp. (0.24)	*LONRF3, KIAA1210, PGRMC1*[§], *SLC25A43*[†], *LOC100303728, SLC25A5, CXorf56, UBE2A, NKRF, SEPT6, MIR766*
56	Xq25	1/359 (0.27%)	0/370 (0%)	0.492	06-188	Azoosp. (0.0; SCOS)	*LOC100129520, DCAF12L2*[§]
57	Xq25	1/359 (0.27%)	0/370(0%)	0.492	05-238	Cryptzoosp. (0.22)	*DCAF12L2*[†], *DCA12L1*
61	Xq27.2	1/359 (0.27%)	0/370 (0%)	0.492	07-30	Azoosp. (0.0; SCOS)	*SPANXA2,SPANXA1,SPANXD,SPANXE*[§], ***MAGEC3,*** *MAGEC1*[†], *MAGEC2*
66	Xq27.3	1/359 (0.27%)	1/370 (0.2%)	1.000	07-516, CS67	07-516: Azoosp. (0.0; mixed SCOS-hypospermatogenesis); CS67: Normozoosp. (235)	*CXorf1, MIR890, MIR888, MIR892A, MIR892B, MIR891B, MIR891A*[§]
67	Xq28	4/359 (1.11%)	0/370 (0%)	0.058	05-196, MMP676, MMP687, MMP704	05-196: Azoosp. (0.0; SCOS); MMP676: Oligozoosp. (21.5); MMP687: Oligozoosp.(57.2); MMP704: Oligozoosp (1.02).	*IDS, LOC100131434, CXorf40A*[§], *MAGEA9B**, *HSFX2*[†], *HSFX1*[†], *TMEM185A, MAGEA11, HSFX1, HSFX2, MAGEA9, MAGEA8, CXorf40B*

Genes inside the CNV minimum size are depicted in bold;
*genes inside the CNV maximum size;
[§]the first proximal flanking gene;
[†]the first distal flanking gene; the remaining genes are situated <500 Kb from the minimum size border. Azoosp = Azoospermia; Oligozoosp = Oligozoospermia; Cryptozoosp = Cryptozoospermia; SCOS = Sertoli Cell Only Syndrome; SGA = Spermatogenic Arrest.
doi:10.1371/journal.pone.0044887.t007

member E (*SPANXE*). Four deletions (CNV22, 54, 56 and 57) contained several (from 4–32) conserved transcription factor binding sites, but the neighbouring genes were relatively distant (from 8 Kb to 400 Kb).

Discussion

The diffusion of assisted reproductive techniques as a therapeutic option in severe male factor infertility raised several questions about the short and long-term consequences on the offspring, since infertile men are at higher risk of being carriers of genetic anomalies in both their genomic DNA and gametes. Although the importance of diagnosing genetic factors in this category of future fathers is fully recognized, the diagnostic workup of infertile men is still limited to a few genetic tests. Our working hypothesis was that, similarly to Y chromosome-linked CNVs (AZF and gr/gr deletions), we would be able to identify recurrent, pathogenic deletions on the X chromosome. First, an X-chromosome specific high resolution array-CGH analysis was carried out in 199 men with known sperm count and was followed by a screening of selected CNVs in several hundred infertile patients and normozoospermic controls. Our array-CGH analysis showed that 50% of subjects presented at least one CNV, and the majority of these CNVs (55/73) were not reported in currently available databases of genomic variants. Among the few X-linked CNVs reported in subjects with known sperm count [7] only six partially or completely overlapping CNVs were found. This can be due to both technical issues (different array resolution, different criteria used for the interpretation of data, lack of validation in the Tuttelmann paper) and/or due to the patient selection criteria (azoospermic men were selected for a specific histology, called SCOS, in the Tuttelmann et al paper [7]). Interestingly, a small deletion, CNV 69 on Xq28 was observed in 7 patients and 3 controls and it maps inside a CNV reported by Tuttelmann et al [7] as patient-specific, present in a single oligozoospermic German man ("private"). This discrepancy is likely due to the larger size (34 Kb) of the patient-specific deletion in the German patient compared to our 10 subjects (11.7 kb). On the contrary, a reciprocal deletion/duplication (CNV31/CNV30) was observed exclusively in patients (n = 4) in our study, whereas Tuttelmann et al. found two normozoospermic carriers of the duplication and one carrying the deletion [7]. However, the deletion encountered in the above German study was 25 Kb smaller than CNV31/30. An other interesting finding concerns two partially overlapping gains detected in both studies, which affect the dosage of two genes (*H2BFWT* and *H2BFM*). In our study this CNV (CN51) has been found both in controls (n = 4) and patients (n = 5), whereas in the German study [7] it was found only in an oligozoospermic patient. Given that the larger CNV reported in the German study [7] duplicates also two other genes (*TMSB15B*, *H2BFXP*), the combined analysis of the results suggests that it is more likely that the not shared genes, situated in the larger duplication, are responsible for the observed oligozoospermic phenotype.

The further analysis of patient-specific deletions (n = 13) revealed that >90% of them are unique or rare (frequency <1%). These data are in line with the previous whole genome array-CGH study [7] in which among the 27 patient-specific CNVs only one recurrent duplication was found in two oligozoospermic men. Similarly in the paper by Stouffs et al, among the 10 patient specific autosomal CNVs only two were recurrent [8]. The role of rare CNVs has already been established for other multifactorial diseases [18,19] and since mutations causing spermatogenic failure are unlikely transmitted to the next generation, we can predict that *de novo* mutations probably play a major role in primary testicular failure. It remains difficult to ascertain the importance of rare patient-specific CNVs in spermatogenesis through family analysis, since analysis on maternal X-chromosome would not be informative and brothers (with a 50% chance of sharing the same X chromosome) were not available for analysis. The difficulty to obtain DNA from relatives in relationship with infertility studies is related to the delicate nature of this condition and for this reason the two previous array-CGH studies were also unable to define the *de novo* nature of the identified CNVs. As an alternative way to explore their potential clinical relevance, we performed a search for functional genomic regions (protein coding genes, microRNAs, conserved transcription binding sites) mapping inside or nearby the 13 deletions of interest. Since men are hemizygous for X-linked genes, their CNV-dependent altered expression cannot be compensated by a normal allele and could potentially lead to a direct pathological effect. Ours is the first study suggesting that X-linked CTA family members are recurrently affected and their dosage variation may play a role in CNV-related spermatogenic failure. CTA genes comprise more than 240 members from 70 families and are generally divided into two broad categories: X-linked (mostly multicopy genes) and non-X CTA genes (mainly single copy genes located on autosomes) [for review see [16,20]]. These genes are normally expressed only in germ cells but aberrant activation has also been reported in a number of malignant tumors. The exclusive physiological expression in germ cells strongly suggests a role in spermatogenesis hence human CTA gene family members are largely unexplored and no clinical data is available. Interestingly, by tracing the evolutionary history of CTA genes, it has been demonstrated that CTA genes in general and the X chromosome linked CTA genes in particular are under strong diversifying pressure and amongst the fastest-evolving genes in the human genome [21]. Consequently, many of the human X-linked CTA genes do not have easily identifiable orthologues in the mouse or rat genomes, which makes it difficult to study the role of these genes in animal models. Clues regarding functionality of CTAs for many of these proteins point to a role in cell cycle regulation or transcriptional control [for review see [22]]. Data obtained in the 103 controls (array-CGH analysis) indicates that in this group only one control-specific deletion contained a CTA gene, the sarcoma antigen 1 *SAGE1*, which indicates that this gene is unlikely a spermatogenesis candidate gene. In support of such a statement, the expression of this gene is extremely low in the testis. On the contrary, for the patient-related CTA genes expression levels in the testis and germ cells were substantially higher. Apart from CTA family members we identified other potential candidate genes in the patient group which deserve further genetic screening. On the contrary, we can conclude that those genes which are deleted in control subjects, are unlikely to be spermatogenesis candidate genes since their absence is compatible with normal spermatogenesis. Among the 6 gene-containing control–specific losses, with the exception of vesicle-associated membrane protein 7 (*VAMP7*), the level of testicular expression is either absent or very low. *VAMP7* is situated in PAR2 and it has been described as strongly expressed in the testis, especially in spermatids. Our data indicates that *VAMP7* haploinsufficiency (i.e. one copy of the gene is still retained on the Y-linked PAR2) does not impair spermatogenesis.

One of the most stimulating findings of our article is related to the CNV burden observed in the patients' group in relationship with loss of genetic material. The relatively high frequency of Y chromosome deletions (4–7% in severe spermatogenic failure) already suggested that infertile men are more prone to the loss of genetic material [11]. The mechanism by which Y chromosome

deletions lead to spermatogenetic failure is not fully clarified and they may act either by removing genes involved in spermatogenesis or by affecting meiosis. Here we found an excess of X-linked CNV number and DNA loss in patients with reduced sperm count, which was only partially related to direct gene removal, hence the majority of deletions mapped close to gene-rich areas. We also found a significant association between CNV number and sperm count in the infertile group, which further reinforces the potential link between deletion burden and spermatogenic failure. Similarly to our data, in the paper by Tuttelmann et al [7] a significant inverse correlation has been found between sperm count and CNV number at the whole genome level.

Whether the observed deletions are directly responsible for the phenotype (either affecting gene expression or interfering with sex chromosome pairing for those mapping to the PAR regions) or simply arise due to increased genomic instability, remains a puzzling question. Some previous observations suggest a possible relationship between genomic instability and male infertility and are related to microsatellite instability [23] as well as to the presence of multiple CNVs on the Y chromosome in men with AZF deletions [24] and an excessive CNV number in azoospermic men with SCOS [7]. Previously, we also observed a significant effect of multiple rearrangements in the AZFc region on sperm production, suggesting a potential link between a less stable genome and spermatogenic efficiency [25]. Additionally, epidemiological observations showing a higher incidence of morbidity (including cancer) and lower life expectancy [22,26] in infertile men would support a potential link between altered spermatogenic function and genomic instability. Our study suggests a potential involvement of increased X-linked deletion burden in the aetiology of impaired spermatogenesis and stimulates further research to better define its implication in primary testicular failure and on general health issues for both the patient and his future offspring.

In conclusion, by the analysis of the X chromosome, at the highest resolution available to date, in a large group of subjects with known sperm count we were able to provide evidence about the lack of highly recurrent deletions, which suggest that an AZFc-like region does not exist on this sex chromosome. Our investigation gives an important contribution both to the field of genetics and reproductive medicine since we identified a large number of novel CNVs, and by our second step analysis, we confirmed 12 deletions as being specific to men with impaired spermatogenesis. The analysis of gene-containing CNVs in patients and in controls allows to discern between those that merit future research and those which are unlikely to be involved in spermatogenesis.

Supporting Information

Figure S1 Array-CGH profiles of two CNVs detected by customed oligonucleotide-based X microarray. Magnified view of CNV 30 (left) and CNV 50 (right) in cases 08-79 and 07-22, respectively. The shaded areas indicate a gain in DNA copy number (duplication, average log2 ratios: +1) detected by red dots (left) and a deletion (average log2 ratios: −4) detected by green dots (right). Arrows indicate the first and the last oligonucleotide duplicated (left) or deleted (right), respectively.
(TIF)

Table S1 List of primers used for the validation of array-CGH results and for the case-control study.
(DOC)

Table S2 List of TaqMan Copy number assay codes used for the validation process.
(DOC)

Table S3 A general outline of the array-CGH findings with phenotypic description of patients and controls.
(XLS)

Table S4 Phenotypic features according to the presence/absence of losses in patients, including the comparison between carriers and no-CNV carriers of hormonal parameters and testis volumes (A) as well as the description of patients with losses detected during both the array-CGH and case-control studies (B).
(DOC)

Table S5 Array-CGH study: comparison of semen parameters according to the number of CNVs in the control group.
(DOC)

Table S6 Array-CGH study: Spermatogenic characteristics of patients and controls carrying more than one CNV.
(DOC)

Acknowledgments

We thank Prof O. Zuffardi, D. Conrad, K. Aston and D. Carrell for helpful discussions. A special thank to Mrs Esperancia Marti from the Fundacio Puigvert for her continues support. We also thank all the clinicians (M. Maggi, A. Magini, F. Lotti) who provided samples for this study from the Andrology Unit of the University of Florence and from the Fundacio Puigvert (L. Bassas, O. Rajmil, J. Sarquella, A. Vives, J. Sanchez-Curbelo).

Author Contributions

Conceived and designed the experiments: CK ER. Performed the experiments: CG DLG FD CC. Analyzed the data: CK ER EA. Contributed reagents/materials/analysis tools: EA ER-C GF. Wrote the paper: CK. Patient reclutement: ER-C GF CK.

References

1. Krausz C (2011) Male infertility: pathogenesis and clinical diagnosis. Best Pract Res Clin Endocrinol Metab 25:271–285.
2. Nuti F, Krausz C (2008) Gene polymorphisms/mutations relevant to abnormal spermatogenesis. Reprod Biomed Online 16:504–513.
3. Tuttelmann F, Rajpert-De Meyts E, Nieschlag E, Simoni M (2007) Gene polymorphisms and male infertility–a meta-analysis and literature review. Reprod Biomed Online 15:643–658.
4. Matzuk MM, Lamb DJ (2008) The biology of infertility: research advances and clinical challenges. Nat Med 14:1197–1213.
5. Aston KI, Carrell DT (2009) Genome-wide study of single-nucleotide polymorphisms associated with azoospermia and severe oligozoospermia. J Androl 30:711–725.
6. Aston KI, Krausz C, Laface I, Ruiz-Castane E, Carrell DT (2010) Evaluation of 172 candidate polymorphisms for association with oligozoospermia or azoospermia in a large cohort of men of European descent. Hum Reprod 25:1383–1397.
7. Tuttelmann F, Simoni M, Kliesch S, Ledig S, Dworniczak B, et al. (2011) Copy number variants in patients with severe oligozoospermia and Sertoli-cell-only syndrome. PLoS One 6:e19426.
8. Stouffs K, Vandermaelen D, Massart A, Menten B, Vergult S, et al. Array comparative genomic hybridization in male infertility. Hum Reprod 27:921–929.
9. Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, et al. (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423:825–837.
10. Wang PJ, McCarrey JR, Yang F, Page DC (2001) An abundance of X-linked genes expressed in spermatogonia. Nat Genet 27:422–426.

11. Krausz C, Chianese C, Giachini C, Guarducci E, Laface I, et al. (2011) The Y chromosome-linked copy number variations and male fertility. J Endocrinol Invest 34:376–382.
12. Tyler-Smith C, Krausz C (2009) The will-o′-the-wisp of genetics–hunting for the azoospermia factor gene. N Engl J Med 360:925–927.
13. Mueller JL, Mahadevaiah SK, Park PJ, Warburton PE, Page DC, et al. (2008) The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. Nat Genet 40:794–799.
14. Stouffs K, Tournaye H, Liebaers I, Lissens W (2009) Male infertility and the involvement of the X chromosome. Hum Reprod Update 15:623–637.
15. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, et al. (2006) Global variation in copy number in the human genome. Nature 444:444–454.
16. Fratta E, Coral S, Covre A, Parisi G, Colizzi F, et al. (2011) The biology of cancer testis antigens: putative function, regulation and therapeutic potential. Mol Oncol 5:164–182.
17. Del Castillo EB, Trabucco A, De La Balze FA (1947) syndrome produced by absence of the germinal epithelium without impairment of the Sertoli or Leydig cells. J Clin Endocrinol Metab 7: 493–502.
18. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, et al. (2009) Finding the missing heritability of complex diseases. Nature 461:747–753.
19. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, et al. (2010) Functional impact of global rare copy number variation in autism spectrum disorders. Nature 466:368–372.
20. Almeida LG, Sakabe NJ, deOliveira AR, Silva MC, Mundstein AS, et al. (2009) CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens. Nucleic Acids Res 37:D816–819.
21. Stevenson BJ, Iseli C, Panji S, Zahn-Zabal M, Hide W, et al. (2007) Rapid evolution of cancer/testis genes on the X chromosome. BMC Genomics 8:129.
22. Salonia A, Matloob R, Gallina A, Abdollah F, Sacca A, et al. (2009) Are infertile men less healthy than fertile men? Results of a prospective case-control survey. Eur Urol 56:1025–1031.
23. Maduro MR, Casella R, Kim E, Levy N, Niederberger C, et al. (2003) Microsatellite instability and defects in mismatch repair proteins: a new aetiology for Sertoli cell-only syndrome. Mol Hum Reprod 9:61–68.
24. Jorgez CJ, Weedin JW, Sahin A, Tannour-Louet M, Han S, et al. (2011) Aberrations in pseudoautosomal regions (PARs) found in infertile men with Y-chromosome microdeletions. J Clin Endocrinol Metab 96:E674–679.
25. Krausz C, Giachini C, Xue Y, O'Bryan MK, Gromoll J, et al. (2009) Phenotypic variation within European carriers of the Y-chromosomal gr/gr deletion is independent of Y-chromosomal background. J Med Genet 46:21–31.
26. Jensen TK, Jacobsen R, Christensen K, Nielsen NC, Bostofte E (2009) Good semen quality and life expectancy: a cohort study of 43,277 men. Am J Epidemiol 170:559–565.

Table S1. Primers used for the validation of array-CGH results (in bold) and for the case-control study

LOSS code	**PRIMERS FOR FIRST STEP SCREENING** (5'→ 3') Forward. Reverse		**PRIMERS FOR CONFIRMATION** (5'→ 3') Forward. Reverse	
16	**del Xp22.32-F1**	**TTATTGGTGGCGGGGTATTA**	del Xp22.32-F2	TCTTTGCTCTCCTCGCAAAT
	del Xp22.32-R1	**TATGTTTGGCAGGCATTTGA**	del Xp22.32-R2	GCGAATTTGGTGAATGTGTG
17	**del Xp22.31B-F4**	**GTCCTGCCTTTGACCACATT**	del Xp22.31B-F3	TTCATGGGAAACAACTGCAA
	del Xp22.31B-R4	**CGTGGACAGGGTTCTTCATT**	del Xp22.31B-R3	TGACAAATGCAAGGTGGAAA
18	**del Xp22.31-F1**	**CTGGTGTTAGGCCGTGAAAT**	del Xp22.31-F2	GCTCCCAATGGATTTGAGAA
			del Xp22.31-R2	AGGCTAAAAGAGGCCCAGAG
	del Xp22.31-R1	**GCATGAACCTGAACAGAGCA**	del Xp22.31-F3	CAACAGGAAGCGAGTTGTCA
			del Xp22.31-R3	CCTGGTGGTGGAGACAGTTT
22	**del Xp22.11-F1**	**ACTTTTCCCAGCTTTGCTCA**	del Xp22.11-F2	CCCCACGTATTGGTTAATGG
			del Xp22.11-R2	TGAGATGCATTCCATTCCAA
	del Xp22.11-R1	**GGCAGGAGCTAAGAATGCAC**	del Xp22.11-F3	ACACACACACCCTGACTCCA
			del Xp22.11-R3	CGTGTTTGGAACCTCCATCT
23	**del Xp21.3-F1**	**ATTTTGGTTCCTCTGCATGG**	del Xp21.3-F2	AAGCCCACACCACATTCTTC
			del Xp21.3-R2	TGACCTGGGAGCAGTTCTCT
	del Xp21.3-R1	**CCCTGGTAGCCACCATTCTA**	del Xp21.3-F3	CTCTGAGTCACCCACCGTCT
			del Xp21.3-R3	ACCCCTGTTAAAGGGCTGAT
24	**Xp21.3b-F3**	**AAGGGAGGGAGGAAGTTTGA**	delXp21.3b-F1	CACGCTGAAGTCATCCAGAA
	Xp21.3b-R3	**GCTGGAGCAGATCACAACAA**	delXp21.3b-R1	CAAAAGCAATGGGGTTCACT
25.A	**delXp21.2-F1**	**TGCTGAAGTCTGCAACAACC**		
	delXp21.2-R1	**GTCACTTGCCAGAGGAGGAG**		
25.B	**delXp21.1-F1**	**CCTGTGGAGCTGTGAGTCAA**		
	delXp21.1-R1	**AAATGCCAGTAGCACCCAAC**		
32	**delXp11.22B-F2**	**GAAATTGTGGCACACCACTG**	delXp11.22B-F3	TCCTGATGGATCTTTTGACCA
	delXp11.22B-R2	**GGGGATCCATAGAACAGCAA**	delXp11.22B-R3	CTCAGTGGTGCCTTGGAGAT
33.A	**delXp11.21-F1**	**CGTTTCATGGTTTCCGAGTT**		
	delXp11.21-R1	**CCCATAAAATCCCCAGACCT**		
50	**del Xq22.1-F1**	**GGTGCTTGTGTTTGCTGATG**	del Xq22.1-F2	ACTGGAGCCTTATCCCACCT
	del Xq22.1-R1	**TTGTTTGAGCCAGGGATTCT**	del Xq22.1-R2	TTTGTGAGCTTTTGCCCATT
53.A	**delXq24-F1**	**CGTGCGTATGTGTTTCATCC**		
	delXq24-R1	**CCACTGGCCCACATCTTAGT**		
54	**del Xq24-F1**	**CTTCTGCACGTTTGAGGACA**	del Xq24-F2	TGCTGGAACAATGACCTTGA
			del Xq24-R2	GACACTTTCGGTTTGCCCTA
	del Xq24-R1	**TCCTCTGTGCTGAAGGGTCT**	del Xq24-F3	AGTGAACAGGGAGAGGCAGA
			del Xq24-R3	CCCAAAAGCTGAAGTCCAAG
56	**delXq25B-F2**	**ATTGCAGTCAGAGCCTTGGT**	delXq25B-F3	TGGGTGGTGGTATGACATTG
			delXq25B-R3	TTGGCTCCCAAATAAGATGG
	delXq25B-R2	**GCAGATTGGCTGGTCATCTT**	delXq25B-F4	TATTTCCATGGACAGCAGCA
			delXq25B-R4	TCACAGACACACATGGCTCA
57	**del Xq25-F1**	**GCAGGTCAGCTTGAAACTCC**	del Xq25-F2	GGAGCCAATAACAGCTCTGC
			del Xq25-R2	CAGAATCTGCGTGGAGAACA
			del Xq25-F3	TGCTCCAGTGATTGTCCTCA
	del Xq25-R1	**AGGGCAGCAAGTTTCTTTCA**	del Xq25-R3	AGGGAGATGGAGAACGTGTG
			del Xq25-F4	GCAGGTCAGCTTGAAACTCC
			del Xq25-R4	TGGAGCCTGGAATCGTAGAC
58.A	**delXq25B1-F1**	**ACTGAATGCCCAATGAGAGG**		
	delXq25B1-R1	**ACAACCTTGGAGACCAATGC**		

LOSS code	**PRIMERS FOR FIRST STEP SCREENING** (5'→ 3') Forward. Reverse		**PRIMERS FOR CONFIRMATION** (5'→ 3') Forward. Reverse	
60.A	**delXq26.3-F1**	**TGACACAAGTGTTGGGGAAA**		
	delXq26.3-R1	**TGGACAGCTGCATAGTCCTG**		
60.D	**delXq27.1-F1**	**TTGGAATCCTTTCTCGGATG**		
	delXq27.1-R1	**CCGGTTTTCTGAGTGACCAT**		
61	**del Xq27.2-F1**	**CAGATAGGGCACTGCAGACA**	del Xq27.2-F3	ACACCAGAGACTGGGGAGTG
	del Xq27.2-R1	**CCTGCTTTTGCCAGGTAGAG**	del Xq27.2-R3	CTGGTGCATTTCATGCTCTC
64	**del Xq27.3-F1**	**CGGCCAAAGTATTCTGGGTA**	del Xq27.3-F3	ACCAACTGGTGTGGCTAAGG
			del Xq27.3-R3	CACGTGACTCTGACCAGCAT
	del Xq27.3-R1	**CAGGGGAGGTGTGATAGCAT**	del Xq27.3-F4	GGCACTGAAAAGATGGGAAA
			del Xq27.3-R4	GGTTTCACTGCTACCCTCCA
66	**del Xq27.3B-F3**	**TATGCACGTTGAAGCCTGAG**	del Xq27.3B-F4	TGGCACATGGTAGGCATTTA
	del Xq27.3B-R3	**CGAGGAAACTGAAGCCACTC**	del Xq27.3B-R4	GGTTTAGCTCCAGGATGCAG
66.A	**delXq28B-F1**	**TCCAGAGAGCCACCTTGACT**		
	delXq28B-R1	**GGGCAGAACACAAAGGACAT**		
67	**del Xq28-F1**	**ACAGCATAGGCTCTGGAGGA**	del Xq28-F2	GCCAGGCAGAAGAGTACCTG
			del Xq28-R2	CTGTGCTCCTGTTCCAGTGA
	del Xq28-R1	**CCCTGAGCTAGGTGCTTCAC**	del Xq28-F3	CACTGGAACAGGAGCACAGA
			del Xq28-R3	TGTGGAGCAACAGGTGAGAG
69	**del Xq28B2-F1**	**TTACAGGTGTGTGCCACCAT**	del Xq28B2-F2	GCCAGAGGAAATGGCTTGTA
			del Xq28B2-R2	GATGGGTAGGTGGAGAACCA
	del Xq28B2-R1	**TGTCAACTGGTTGGAACAGG**	del Xq28B2-F4	AGGGAAATGGAAACCAAAGG
			del Xq28B2-R4	TCTGAAGGCAAAGGATCACA

Table S2. List of TaqMan Copy number assay codes used for the validation process

CNV CODE	TaqMan Copy Number Assay ID
CNV19	Hs00131100_cn
CNV20	Hs02640673_cn
CNV25	Hs05664369_cn
CNV26	Hs04101901_cn
CNV27	Hs04101901_cn
CNV30/31	Hs04119215_cn
CNV39	Hs05614245_cn
CNV40	Hs04510684_cn
CNV55	Hs05675355_cn
CNV58	Hs05642887_cn
CNV60	Hs07514404_cn

CNV CODE	LOSS or GAIN	Region	Size (Kb)	Start	End	Coding sequences within the CNV	DGV	SD	Frequency in patients (n=96)	Patient Code CNV carrier	Semen Phenotype (total sperm count in x10^6)	Control Code CNV carrier	Semen Phenotype (total sperm count in x10^6)
1a	GAIN	Xp22.33	224.83	1.544	226.372	*PLCXD1, GTPBP6, PPP2R3B*		N	1	A800	Oligozoospermia (2)		
4.A	GAIN	Xp22.33	237.08	153.373	390.452	*PLCXD1, GTPBP6, PPP2R3B*		N	0			08-373	Normozoospermic (75)
5	GAIN	Xp22.33	160.10	302.644	462.740	NO		N	1	A326	Oligozoospermia (6.3)		
5.A	GAIN	Xp22.33	241.98	674.222	916.206	NO		N	0			07-507[2]	Normozoospermic (188)
5.B	LOSS	Xp22.33	12.63	701.071	713.696	NO		N	0			08-103	Normozoospermic (240)
5.C	GAIN	Xp22.33	420.72	747.358	1.168.080	NO		N	0			08-648	Normozoospermic (56)
11	GAIN	Xp22.33	39.73	1.347.599	1.387.328	*CSF2RA*		N	1	MMP718[2]	Oligozoospermia (6.4)		
12	GAIN	Xp22.33	17.30	1.693.897	1.711.194	*ASMT*	Variation_73908 Variation_83259	N	1	06-111[2]	Azoospermia (0)	C17[2] C104 08-18	Normozoospermic (570) Normozoospermic (280) Normozoospermic (355)
12.A	GAIN	Xp22.33	6.61	1.693.897	1.700.511	*ASMT*		N	0			08-29	Normozoospermic (294)
12.B	GAIN	Xp22.33	683.74	1.716.023	2.399.766	*ASMT, DHRSX*		Y	0			C11[2]	Normozoospermic (78.70)
14	GAIN	Xp22.33	1.40	1.896.197	1.897.608	NO	Variation_31542	Y	2	M8[2] 07-170[3]	Oligozoospermia (0.6) Oligozoospermia (0.53)		
15	LOSS	Xp22.33	1.40	1.896.197	1.897.608	NO	Variation_104545	Y	23	05-123 06-111[2] 06-157 06-201 07-99[2] 07-341[2] 07-505 08-26 08-92[2] 08-259[2] 08-262 08-282 09-263 A371[2] 09-223 A833 A448[2] A456 05-172 05-205[2] A688 09-137 05-236[2]	Azoospermia (0) Azoospermia (0) Azoospermia (0) Azoospermia (0) Azoospermia (0) Azoospermia (0) Azoospermia (0) Azoospermia (0) Azoospermia (0) Azoospermia (0) Azoospermia (0) Azoospermia (0) Azoospermia (0) Oligozoospermia (0.01) Oligozoospermia (0.79) Oligozoospermia (0.6) Oligozoospermia (13.4) Oligozoospermia (2.7) Azoospermia (0) Azoospermia (0) Oligozoospermia (6.1) Azoospermia (0) Azoospermia (0)	C73	Normozoospermic (384)
15.A	GAIN	Xp22.33	27.94	2.382.699	2.410.643	*DHRSX*	Variation_83270	Y	0			08-471	Normozoospermic (157.50)
15.B	GAIN	Xp22.33/22.32	280.09	4.206.493	4.486.580	NO		N	0			07-462	Normozoospermic (197.50)
16	LOSS	Xp22.32	7.76	4.250.413	4.258.174	NO	Variation_52995	N	2	07-13[3] 07-170[3]	Azoospermia (0) Oligozoospermia (0.53)	08-196	Normozoospermic (185.50)
16.A	GAIN	Xp22.31	1609.42	6.487.238	8.096.662	*HDHD1, STS, VCX, PNPLA4, MIR651*		N	0			C53[3]	Normozoospermic (271)
17	LOSS	Xp22.31	31.70	6.594.834	6.626.533	NO		N	1	A448[2]	Oligozoospermia (13.4)		
18	LOSS	Xp22.31	82,00	6.756.310	6.838.310	NO	Variation_8908 Variation_53018 Variation_34619	N	1	07-96	Azoospermia (0)		
19	GAIN	Xp22.31	245.03	7.002.649	7.247.676	*STS; HDHD1A*		N	1	A761	Oligozoospermia (2.4)		
19.A	GAIN	Xp22.31	129.96	7.961.788	8.091.751	*MIR651*	Variation_39337	N	0			C85	Normozoospermic (215.60)

CNV CODE	LOSS or GAIN	Region	Size (Kb)	Start	End	Coding sequences within the CNV	DGV	SD	Frequency in patients (n=96)	Patient Code CNV carrier	Semen Phenotype (total sperm count in x10^6)	Control Code CNV carrier	Semen Phenotype (total sperm count in x10^6)
19.B	GAIN	Xp22.31	177.54	8.411.159	8.588.699	KAL1		N	0			08-340	Normozoospermic (117)
20	GAIN	Xp22.2	602.10	11.104.518	11.706.614	*ARHGAP6, AMELX, MSL3*		N	1	M9	Azoospermia (0)		
20.A	GAIN	Xp22.2	665.88	14.590.604	15.256.487	*MOSPD2, GLRA2, FANCB, ASB9, ASB11, PIGA*		N	0			07-507[2]	Normozoospermic (188)
20.B	GAIN	Xp22.13	13.34	18.018.894	18.032.238	NO		N	0			C115	Normozoospermic (760)
22	LOSS	Xp22.11	24.35	22.969.648	22.993.997	NO		N	1	MMP718[2]	Oligozoospermia (6.4)		
23	LOSS	Xp21.3	6.69	25.274.024	25.280.712	NO		N	1	A142	Cryptozoospermia (0.01)		
24	LOSS	Xp21.3	67.33	26.891.769	26.959.101	NO		N	1	M4	Azoospermia (0)		
25	GAIN	Xp21.3	60.40	27.277.529	27.337.933	NO		N	1	05-223	Azoospermia (0)		
25.A	LOSS	Xp21.2	9.69	31.282.923	31.292.613	*DMD*		N	0			08-395	Normozoospermic (50)
25.B	LOSS	Xp21.1	28.26	33.953.232	33.981.492	NO	Variation_7783	N	0			07-272	Normozoospermic (512,50)
												C119[2]	Normozoospermic (394,40)
25.C	GAIN	Xp21.1	185.02	34.931.807	35.116.827	NO		N	0			08-414[3]	Normozoospermic (138)
25.D	GAIN	Xp21.1	215.00	35.269.628	35.484.626	NO		N	0			08-414[3]	Normozoospermic (138)
26	GAIN	Xp21.1	88.57	37.168.387	37.256.960	*PRRG1*		N	1	07-516[2]	Azoospermia (0)		
27	GAIN	Xp21.1	9.61	37.242.364	37.251.969	NO	N	Y	1	07-13[3]	Azoospermia (0)		
30	GAIN	Xp11.23	81.13	47.766.391	47.847.516	*ZNF630*	Variation_53003 Variation_9343 Variation_83491	Y	1	08-79[2]	Azoospermia (0)		
31	LOSS	Xp11.23	81.13	47.766.391	47.847.516	*ZNF630*	Variation_96640 Variation_9861 Variation_53005	Y	1	A630	Oligozoospermia (7.2)		
31.A	GAIN	Xp11.23	78.87	48.021.982	48.100.848	*SSX3*		Y	0			07-531	Normozoospermic (167.50)
32	LOSS	Xp11.22	4.15	52.065.798	52.069.943	NO		N	1	A162	Cryptozoospermia (0,01)		
33.A	LOSS	Xp11.21	58.89	56.403.390	56.462.278	NO		N	0			C7	Normozoospermic (416)
34.A	GAIN	Xp11.12	48.06	56.870.427	56.918.489	NO		N	0			08-105	Normozoospermic (550)
35	GAIN	Xp11.1	117.14	57.318.438	57.435.573	*FAAH2*		N	1	07-451	Oligozoospermia (0.21)	C17[2]	Normozoospermic (570)
36.A	GAIN	Xq11	716.03	63.925.948	64.641.977	*ZC4H2; ZC3H12B*		N	0			C87	Normozoospermic (368)
38	GAIN	Xq13.2	8.98	72.202.996	72.211.976	*NO*		N	1	A339	Oligozoospermia (4)		
38.A	GAIN	Xq13.2	192.04	74.375.875	74.567.915	*UPRT, ZDHHC15*	Variation_74012	N	0			08-414[3]	Normozoospermic (138)
38.B	GAIN	Xq13.3	153.23	75.123.387	75.276.621	NO		N	0			C113	Normozoospermic (332.40)
39	GAIN	Xq21.1	21.98	76.992.067	77.014.050	*MAGT1*		N	1	A2[3]	Cryptozoospermia (0.01)		
40	GAIN	Xq21.1	5.28	80.112.246	80.117.526	NO	Variation_83611	N	2	05-238[3]	Oligozoospermia (0.22)		
										07-02	Azoospermia (0)		
49	GAIN	Xq22.1	5.30	100.942.190	100.947.490	NO		N	2	08-141	Azoospermia (0)	09-01	Normozoospermic (270)
										07-99[2]	Azoospermia (0)	C81[2]	Normozoospermic (354)
50	LOSS	Xq22.1	45.36	101.803.578	101.848.935	*ARMXC5-GPRASP2*		Y	1	07-22	Azoospermia (0)		
51	GAIN	Xq22.2	34.69	103.152.319	103.187.013	*H2BFWT, H2BFM*	Variation_3254	Y	5	A404	Oligozoospermia (5.8)	10-457	Normozoospermic (772)
										M6	Azoospermia (0)	C53[3]	Normozoospermic (271)
										M8[2]	Azoospermia (0)	C81[2]	Normozoospermic (354)
										07-30[3]	Azoospermia (0)	C110[3]	Normozoospermic (259.60)
										08-259[2]	Azoospermia (0)		
53.A	LOSS	Xq24	170.73	118.278.913	118.449.646	*SLC25A43*		N	0			C119[2]	Normozoospermic (394.40)
54	LOSS	Xq24	44.85	118.281.024	118.325.874	NO		N	1	MMP550	Oligozoospermia (0.24)		
55	GAIN	Xq24	206.90	118.691.020	118.897.917	*SEPT6, ANKRD58, RPL39, SNORA69, UPF3B, RNF113A, NDUFA1*		N	1	07-13[3]	Azoospermia (0)		
55.A	GAIN	Xq25	53.80	120.385.787	12.043.959	NO		N	0			C124	Normozoospermic (169.75)

CNV CODE	LOSS or GAIN	Region	Size (Kb)	Start	End	Coding sequences within the CNV	DGV	SD	Frequency in patients (n=96)	Patient Code CNV carrier	Semen Phenotype (total sperm count in x10^6)	Control Code CNV carrier	Semen Phenotype (total sperm count in x10^6)
56	LOSS	Xq25	86.07	124.632.886	124.718.959	NO		N	1	06-188	Azoospermia (0)		
57	LOSS	Xq25	188.03	124.929.673	125.117.699	NO	Variaiton_52924	N	1	05-238[3]	Oligozoospermia (0.22)		
58	GAIN	Xq25	9.68	125.143.278	125.152.957	NO		N	1	A320	Oligozoospermia (5.8)		
58.A	LOSS	Xq25	12.68	125.198.109	125.210.792	NO		N	0			C32[2]	Normozoospermic (480)
59.A	GAIN	Xq26.3	24.69	134.151.039	134.175.725	NO		Y	0			C53[3]	Normozoospermic (271)
60	GAIN	Xq26.3	42.30	134.585.636	134.627.936	NO	Variaiton_52936	Y	1	A2[3]	Cryptozoospermia (0.01)		
60.A	LOSS	Xq26.3	50.84	134.801.361	134.852.198	*SAGE1*		Y	0			C11[2]	Normozoospermic (78.70)
60.B	GAIN	Xq26.3	13.45	136.050.422	136.063.872	NO		N	0			08-322	Normozoospermic (348)
60.C	GAIN	Xq26.3	91.26	137.089.527	137.180.783	NO		N	0			08-119[2]	Normozoospermic (154)
60.D	LOSS	Xq27.1	217.83	140.175.103	140.392.930	NO		Y	0			C110[3]	Normozoospermic (259.60)
61	LOSS	Xq27.2	4.67	140.773.893	140.778.561	*MAGEC3*		N	1	07-30[3]	Azoospermia (0)		
64	LOSS	Xq27.3	3.92	143.436.347	143.440.268	NO	Variation_115340	N	8	07-341[2]	Azoospermia (0)	C110[3]	Normozoospermic (259.60)
										08-79[2]	Azoospermia (0)	C60	Normozoospermic (200)
										M14	Azoospermia (0)	C45	Normozoospermic (122.40)
										M15	Oligoastenozoospermia (24.3)		
										A2[3]	Cryptozoospermia (0.01)	C43	Normozoospermic (137)
										A616	Cryptozoospermia (0.01)		
										A1020	Oligozoospermia (2.4)	C35[2]	Normozoospermic (122)
										MMP597	Oligozoospermia (17.4)		
66	LOSS	Xq27.3	7.35	145.030.566	145.037.917	NO		N	1	07-516[2]	Azoospermia (0)		
66.A	LOSS	Xq28	37.12	147.393.583	147.430.698	*AFF2*		N	0			C77	Normozoospermic (125.40)
67	LOSS	Xq28	5.42	148.456.474	148.461.889	NO		N	1	05-196[2]	Azoospermia (0)		
68	GAIN	Xq28	105.66	148.686.631	148.792.286	*MAGEA8*	Variaiton_31571	N	2	05-196[2]	Azoospermia (0)	C32[2]	Normozoospermic (480)
										05-238[3]	Oligozoospermia (0.22)	C35[2]	Normozoospermic (122)
69	LOSS	Xq28	11.77	154.044.877	154.056.645	NO		N	7	07-30[3]	Azoospermia (0)	C59	Normozoospermic (125.40)
										08-92[2]	Azoospermia (0)		
										A371[2]	Oligozoospermia (2.8)		
										07-170[3]	Oligozoospermia (0.53)	08-119[2]	Normozoospermic (154)
										A955	Oligozoospermia (2.5)	C88	Normozoospermic (396)
										05-205[2]	Azoospermia (0)		
										A609	Oligozoospermia (8)		
71	LOSS	Xq28	290.99	154.586.913	154.877.901	*VAMP7, SPRY3*		N	1	05-236[2]	Azoospermia (0)		
71.A	LOSS	Xq28	122.36	154.755.542	154.877.901	*VAMP7*		N	0			C71	Normozoospermic (270)

Semen phenotypes are reported for both patients and controls, whereas testis histology and information about sperm recovery (when TESE performed) are described only for patients. N= Not present; Y=Yes, present.[2] refers to subjects that displayed two CNVs; [3] refers to subjects that displayed three CNVs. Sp-: no spermatozoa recovered by TESE; Sp+: spermatozoa recovered by TESE; SCOS: Sertoli Cell Only Syndrome. The PAR regions are highlighted in grey.

Table S4. Phenotypic features according to the presence/absence of losses in patients, including the comparison between carriers and no-CNV carriers of hormonal parameters and testis volumes (A) as well as the description of patients with losses detected during both the array-CGH and case-control studies (B).

A.			
	Loss Carriers (n=47)	**No-CNV Carriers (n=36)**	**p value**
FSH (U/L)			
Mean Value ± SD	12.9 ± 8.37	14.21 ± 10.02	0.712
Median (25^{th}-75^{th} percentile)	11.29 (5.80 - 18.55)	13.0 (5.20 – 21.80)	
LH (U/L)			
Mean Value ± SD	5.28 ± 2.64	4.40 ± 1.85	0.243
Median (25^{th}-75^{th} percentile)	5.22 (3.60 - 6.90)	3.85 (3.20 – 5.62)	
TESTOSTERONE.(ng/ml)			
Mean Value ± SD	4.78 ± 1.65	6.07 ± 4.50	0.454
Median (25^{th}-75^{th} percentile)	4.57 (3.27 - 6.64)	5.00 (3.50 - 6.90)	
TESTIS VOLUME (ml)			
Mean Value ± SD	12.45 ± 4.57	9.96 ± 3.85	0.014
Median (25^{th}-75^{th} percentile)	13.0 (9.0 - 16.0)	10.0 (7.00-13.00)	

B.					
Patient ID	**CNV code**	**Total sperm count (n x 10^6)**	**FSH (U/L)**	**Mean testicular volume (ml)**	**Testis histology**
05-123	15	0	14.1	15	MA
05-172	15	0	4.9	11.5	SCOS Type2; sp+
06-111	15	0	18.5	7	SCOS Type1; sp-
06-157	15	0	3.7	18	MA; sp+
06-201	15	0	13	7	MA; sp-
07-99	15	0	6.8	18	SCOS Type1; sp-
07-505	15	0	14.1	8	SCOS Type1; sp-
08-26	15	0	23.6	9	SCOS Type1; sp-
08-259	15	0	12.7	4.5	SCOS Type1; sp-
08-262	15	0	7.23	13	MA + HS; sp-
08-282	15	0	18.7	11.5	SCOS Type1; sp-
09-263	15	0	11.2	8	SCOS Type1; sp-
09-137	15	0	19	13	HS
A833	15	0.6	7.68	14	n.p
09-233	15	0.79	22.8	11.5	SCOS + HS
A456	15	2.7	9.8	10.5	n.p
A688	15	6.1	10.16	8	n.p
07-13	16	0	10.1	16	MA; sp+
A828	17	20.00	13.3	11.5	n.p
07-96	18	0	29.9	13	SCOS Type1; sp-
MMP718	22	6.4	5.3	14	n.p
A142	23	0.01	2.77	18	n.p
M4	24	0	26	0.5	n.p
A630	31	7.20	3.94	23	n.p
09-126	31	8		22	n.p
08-190	32	0	22.7	5	n.p
A162	32	0.01	8.6	13	n.p
07-22	50	0	14.6	15	SCOS Type1; sp-
MMP550	54	0.24	13.1	8	n.p
06-188	56	0	15.5	10.5	SCOS Type1; sp-
05-238	57	0.22	5	11.5	HS
07-30	61	0	34	16	SCOS Type1; sp-
08-79	64	0	3.75	20	MA; sp-
M14	64	0	12.5	9	SCOS Type1; sp-
M15	64	0	24.3	5	n.p
A2	64	0.01	2.15	16	n.p
A616	64	0.01	7.2	19.5	n.p
A1020	64	2.4	3.5	12	n.p
MMP597	64	17.4	3.3	14	n.p
07-516	66	0	27.7	11.5	SCOS + HS; sp+
05-196	67	0	30.7	16	SCOS Type1
MMP704	67	1.02	9.7	14	n.p
MMP676	67	21.5	3.6	14	n.p
MMP687	67	57.20	6.9	14	n.p
A955	69	2.5		15	n.p
A609	69	8.64	7.96	13	n.p
A448	15,17	13.44	9.35	18	n.p
07-341	15,64	0	15	13	MA; sp-
08-92	15,69	0	23	10	SCOS Type1; sp-
05-205	15,69	0	13.2	13	SCOS Type1; sp-

B. continue					
Patient ID	**CNV Code**	**Total sperm count (n x 10^6)**	**FSH (U/L)**	**Mean testicular volume (ml)**	**Testis histology**
A371	15,69	0.01	11.38	16	n.p
05-236	15,71	0	5.8	10.5	SCOS + MA
07-170	16,69	0.53	5.8	15	MA + HS
07-30	61,69	0	34	14	SCOS Type1; sp-

Non-parametric test was performed for all comparisons, except for testis volume for which Student t-test was used. sp-/sp+: no spermatozoa/spermatozoa recovered by Testicular Sperm Extraction (TESE); n.p.: not performed. MA: Maturation Arrest; SCOS Type1: Pure SCOS; SCOS Type2: SCOS + Complete maturation in rare tubules; HS:Hypospermatogenesis

Table S5. Array-CGH study: comparison of semen parameters according to the number of CNVs in the control group

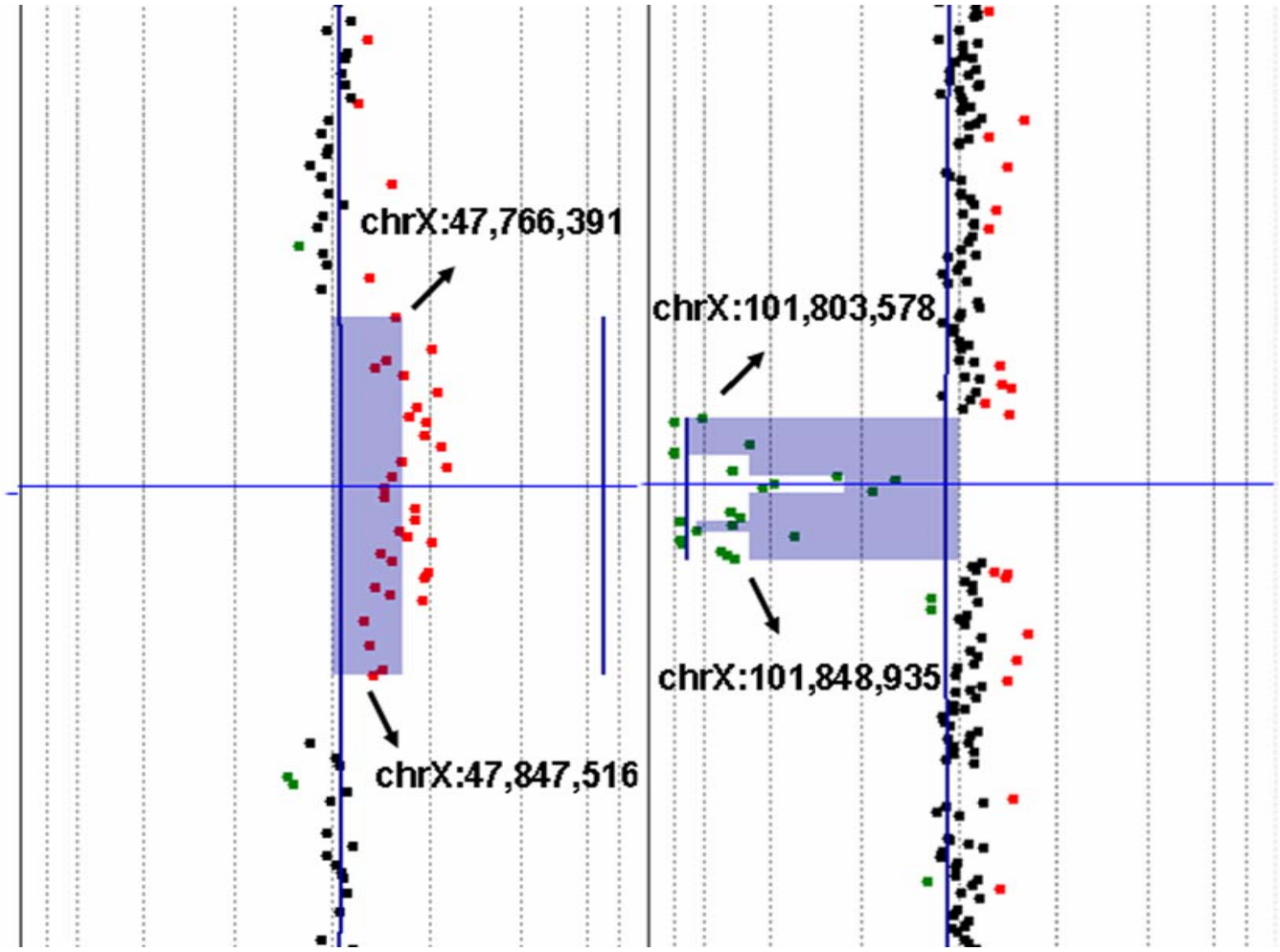
CONTROLS (n=103)				
	Sperm concentration (n x 10^6 / ml)	**p**	**Total sperm number (n x 10^6)**	**p**
0 CNV (n=61)	**94.7±46.9 (88.0; 27.0-250.0)**	**0.869**	**330.7±209.5 (295.0; 39.5-957.0)**	**0.334**
≥ 1 CNV (n=42)	**95.7±53.1 (82.0; 24.0-220.0)**		**289.6±180.4 (269.7; 50.0-772.0)**	
0 LOSS (n=83)	**94.4±48.2 (85.0; 27.0-250.0)**	**0.877**	**320.6±204.4 (280.0; 39.5-957.0)**	**0.653**
≥ 1 LOSS (n=20)	**97.8±54.9 (96.0; 24.0-205.0)**		**286.2±172.8 (254.7; 50.0-617.4)**	
0 GAIN (n=75)	**97.1±48.4 (90.0; 25.0-250.0)**	**0.352**	**322.9±201.9 (283.5; 39.5-957.0)**	**0.339**
≥ 1 GAIN (n=28)	**89.6±52.1 (79.0; 24.0-220.0)**		**289.7±189.8 (269.7; 56.0-772.0)**	
≤ 1 CNV (n=92)	**95.0±49.8 (84.5; 25.0-250.0)**	**0.785**	**318.6±202.7 (280.0; 39.5-957.0)**	**0.579**
> 1 CNV (n=11)	**96.2±47.3 (113.0; 24.0-168.0)**		**274.5±158.9 (269.5; 78.7-570.0)**	
≤1 LOSS(n=101)	**95.1±49.7 (87.0; 24.0-250.0)**	**0.886**	**313.5±200.1 (271.0; 39.5-957.0)**	**0.616**
>1 LOSS(n=2)	**96.5±27.6 (96.5; 77.0-116.0)**		**331.9±88.3 (331.9; 269.5-394.4)**	
≤ 1 GAIN (n=99)	**94.0±49.7 (84.0; 24.0-250.0)**	**0.129**	**317.4±201.0(280.0; 39.5-957.0)**	**0.528**
> 1 GAIN (n=4)	**123.2±31.6 (115.5; 94.0-168.0)**		**237.7±94.9 (229.5; 138.0-354.0)**	

Sperm concentration and total sperm number are expressed as: mean ± SD (median; range). Significance is depicted by a p value< 0.05

Table S6. Array-CGH study: Spermatogenic characteristics of patients and controls carrying more than one CNV.

Patient Code	n of CNVs (n Losses + n Gains)	Sperm concentration (n x 10^6 / ml) – Total sperm count (n x 10^6)	Testis Histology
05-196	2 (1+1)	0.00-0.00	Pure SCOS; sp-
05-205	2 (2+0)	0.00-0.00	Pure SCOS; sp-
05-236	2 (2+0)	0.00-0.00	SCOS + Spermatocytic arrest; sp-
06-111	2 (1+1)	0.00-0.00	Pure SCOS; sp-
07-99	2 (1+1)	0.00-0.00	Pure SCOS; sp-
07-341	2 (2+0)	0.00-0.00	Scleroialinosis. Spermatogonic arrest.with rare spermatocytes; sp-.
07-516	2 (1+1)	0.00-0.00	50% SCOS. Hypospermatogenesis; sp+.
08-79	2 (1+1)	0.00-0.00	Spermatocytic arrest; sp-
08-92	2 (2+0)	0.00-0.00	Pure SCOS; sp-.
08-259	2 (1+1)	0.00-0.00	Pure SCOS; sp-.
M8	2 (0+2)	0.00-0.00	n.p.
A371	2 (2+0)	0.01-0.01	n.p.
A448	2 (2+0)	1.80-13.44	n.p.
MMP718	2 (1+1)	2.00-6.4	n.p.
07-13	3 (1+2)	0.00-0.00	Maturation Arrest.
05-238	3 (1+2)	0.22-0.22	Hypospermatogenesis; sp+
07-30	3 (2+1)	0.00-0.00	Pure SCOS; sp-
07-170	3 (2+1)	0.15-0.53	Incomplete Spermatogenic Arrest (spermatogonia) + Hypospermatogenesis
A2	3 (1+2)	0.01-0.01	n.p.
Control Code	**n of CNVs (n Losses + n Gains)**	**Sperm concentration (n x 10^6 / ml)**	**Total sperm count (n x 10^6)**
07-507	2 (0+2)	94	188
08-119	2 (1+1)	44	154
C11	2 (1+1)	31.5	78.7
C17	2 (1+1)	150	570
C32	2 (1+1)	123	480
C35	2 (1+1)	24	122
C81	2 (0+2)	118	354
C119	2 (2+0)	116	394.4
08-414	3 (0+3)	168	138
C53	3 (0+3)	113	271
C110	3 (2+1)	77	269.5

Testis histologies were available only for azoospermic patients. sp-/ sp+: no spermatozoa/spermatozoa recovered by Testicular Sperm Extraction (TESE); n.p.: not performed.


chrX:47,766,391
chrX:47,847,516
chrX:101,803,578
chrX:101,848,935

PAPER 5

Recurrent X chromosome-linked deletions: discovery of new genetic factors in male infertility.

Lo Giacco D., Chianese C., Ars E., Ruiz-Castañé E., Forti G. and Krausz C.
Journal of Medical Genetics. 2013 (Under review).

SUMMARY 5

Background:
In a previously published study, based on high-resolution X-chromosome-specific a-CGH platforms, we provided the largest collection of X-linked CNVs related to spermatogenesis. Among the 29 deletions identified by a-CGH, 3 recurrent deletions (frequency > 1%) on Xq caught our attention for their exclusive (CNV67) or prevalent (CNV64 and CNV69) presence in patients.

Aims:
We performed the molecular and clinical characterization of these 3 recurrent X-linked deletions in order to investigate their role in the etiology of spermatogenic impairment.

Methods:
An in-depth analysis of the three deletions was performed including:

- The PCR-based screening of a large series of idiopathic patients (n=627) and normozoospermic controls (n=628) from two Mediterranean populations (Spanish and Italian);
- The molecular characterization of deletions. We used PCR +/- to map deletion junctions and Long- Range PCR and direct sequencing to define breakpoints;
- The exploration for functional elements in the region of interest;
- Genotype-phenotype correlation analysis.

Results:
CNV64 was found in 5.7% of patients and in 3.1% of controls (*p* value=0.013; OR=1.89; CI=1.1-3.3). Similarly, the frequency of CNV69 carriers was significantly higher in patients (3.5%) than controls (1.6%) (*p* value=0.023; OR=2.204 CI=1.05-4.62). The carriers of both deletions showed a significantly lower sperm concentration and total motile sperm number compared to non-carriers. The molecular characterization of CNV69 revealed the existence of at least two deletion patterns, named type A and type B. Only type B deletion was significantly more frequent in patients than controls (*p* value=0.011; OR=9.19 CI=1.16-72.8). No genes have been identified inside the maximum size of CNV64 and CNV69, nevertheless a number of regulatory elements, including weak and strong enhancers, insulators and weak promoters have been found to be potentially affected because of their proximity to the deletion. CNV67 deletion was exclusively found in patients at a frequency of 1,1% (*p*<0.01) and was maternally transmitted. The semen phenotype of one carrier (11-041) *versus* his normozoospermic non-carrier brother strongly indicates a pathogenic effect of the deletion on spermatogenesis. Although the highly repetitive nature of the genomic region involved, and the incomplete assembly of the currently available reference sequence of the human X chromosome, prevented the fine mapping of deletion breakpoints, chromosome walking allowed to better define the deletion and to hypothesize the involvement of the proximal copy of the CTA gene *MAGEA9* and/or of its regulatory elements. CNV67 may also affect regulatory elements *HSFX1/2* (Heat shock transcription factor family, X-linked 1/2) with testis-specific expression.

Conclusions:
The X chromosome-linked recurrent deletions CNV64, CNV67, CNV69 are significantly associated with spermatogenic failure. CNV67 specific to spermatogenic failure and with a frequency of 1,1% in oligo/azoospermic men resembles the *AZF* regions on the Y chromosome with potential clinical implications.

Recurrent X chromosome-linked deletions: discovery of new genetic factors in male infertility

D. Lo Giacco [1,2*], C. Chianese [3*], E. Ars [1,2], E.Ruiz-Castañé [2], G.Forti [3] and Csilla Krausz [2,3§]

[1] Molecular Biology Laboratory, Fundació Puigvert Universitat Autònoma de Barcelona, Barcelona, Instituto de Investigaciones Biomédicas Sant Pau (IIB-Sant Pau), Barcelona, Spain.

[2] Andrology Service, Fundació Puigvert, Universitat Autònoma de Barcelona, Instituto de Investigaciones Biomédicas Sant Pau (IIB-Sant Pau), Barcelona, Spain.

[3] Department of Experimental and Clinical Biomedical Sciences, University of Florence, Italy.

***** Equally contributing authors

[§] Corresponding author**: Csilla Krausz MD, PhD**

Department of Experimental and Clinical Biomedical Sciences, University of Florence;

Viale Pieraccini 6- 50139 Florence, Italy. Phone: +39 055 4271415 – Fax: +39 055 4271413 ; email: csillagabriella.krausz@unifi.it

Keywords: male infertility, X-linked CNVs, X chromosome, spermatogenesis

Word count: 2009

ABSTRACT

Background The role of X-linked genes and CNVs in male infertility remains poorly explored. Our previous array-CGH analyses revealed 3 recurrent deletions in Xq, exclusively (CNV67) and prevalently (CNV64, CNV69) found in patients. Molecular and clinical characterization of these CNVs was performed in this study.

Methods 627 idiopathic infertile patients and 628 controls were tested for each deletion with PCR +/-. We used PCR +/- to map deletion junctions and Long-Range PCR and direct sequencing to define breakpoints.

Results CNV64 was found in 5.7% of patients and in 3.1% of controls (*p* value=0.013; OR=1.89; CI=1.1-3.3), CNV69 in 3.5% of patients and 1.6% of controls (*p* value=0.023; OR=2.204 CI=1.05-4.62). For CNV69, we identified two breakpoints, type A and B, with the latter being significantly more frequent in patients than controls (*p* value=0.011; OR=9.19 CI=1.16-72.8). CNV67 was detected exclusively in patients (1.1%) and maternally transmitted. The semen phenotype of one carrier (11-041) *versus* his normozoospermic non-carrier brother strongly indicates a pathogenic effect of the deletion on spermatogenesis. *MAGEA9*, an ampliconic gene reported as independently acquired on the human X chromosome with exclusive physiological expression in the testis is likely to be involved in CNV67.

Conclusions We provide the first evidence for X chromosome-linked recurrent deletions associated to spermatogenic failure. CNV67, specific to spermatogenic failure and with a frequency of 1,1% in oligo/azoospermic men resembles the AZF regions on the Y chromosome with potential clinical implications.

Introduction

The etiology of altered spermatogenesis remains unknown in about 40% of cases representing the so-called "idiopathic infertility", of which a large proportion is likely related to still unknown genetic factors.[1] During the last years, Copy Number Variations (CNVs) have been proven as an interest-meriting aspect also in andrology. To date, the only known CNVs that actually cause spermatogenic failure are Y chromosome microdeletions; however, the high-resolution whole-genome approaches allowed the identification of new spermatogenesis candidate genes as well as recurrent and private patient-specific CNVs with potential clinical interest also on autosomes and the X chromosome.[2–4] In a previously published study,[4] based on high-resolution X-chromosome-specific array-CGH platforms (average resolution: 4 kb), we provided the largest collection of X-linked CNVs related to spermatogenesis and, more interestingly, we observed a deletion burden in relation to spermatogenic impairment, as idiopathic infertile men showed an excessive rate of deletions compared to normozoospermic controls. Among the 29 deletions identified by array-CGH, 3 recurrent deletions (frequency > 1%) on Xq caught our attention for their exclusive (CNV67) or prevalent (CNV64 and CNV69) presence in patients. Our previous publication included also a case-control follow-up study in which only "patient-specific" CNVs were incorporated and CNV67 was found in 4/359 (1.1%) patients and 0/370 controls.

To evaluate the potential role of these recurrent X-linked CNVs in male infertility, we screened over 1,200 men with known sperm parameters in two Mediterranean populations. All three deletions have been confirmed as significant risk factors for impaired spermatogenesis. In particular, CNV67 was confirmed as "patient-specific", being the first X-linked deletion with potential clinical implications.

Subjects and Methods

Germline DNA from a total of 1,255 subjects (627 strictly selected idiopathic infertile patients and 628 normozoospermic controls) coming from Spain (36.6%) and Italy (63.4%) was analyzed (for patient selection criteria see also

Supplementary Materials). The ethnic/geographic composition was similar between the control and patient groups.

Molecular analysis of deletions

A multi-step STS PCR +/- protocol was optimized in order to reliably identify and confirm the presence of deletions (Supplementary Methods). PCR +/- was also used to further restrict the deletion interval for CNV67 and CNV64.

Pedigree analysis

Segregation analysis in relatives was possible for CNV67 carriers 11-041 (Spanish) and MMP704 (Italian). To understand whether CNV67 occurred *de novo*, we analyzed each patient's mother and siblings (11-041's brother and MMP704's sister). Screening of CNV67 carriers' female relatives was performed through quantitative PCR (qPCR) using a TaqMan® Copy Number Assay (hs03323870_cn).

CNV69 Deletion Breakpoint Definition

Conventional PCR +/- using primers mapping to the flanking regions of the minimum CNV69 size was performed to confine breakpoints to smaller intervals. Long-Range (LR)-PCR was subsequently performed to amplify the junction fragment including the breakpoint (Figure 1A; Supplementary Table 2). Additional primers were then designed to sequence the obtained LR-PCR product (Figure 1B). To easily classify the type of breakpoint in all CNV69 carriers and provide a tool for a potential diagnostic screening, another pair of deletion-specific primers was designed (Supplementary Table 2).

Statistical Analysis

Statistical analysis was performed using SPSS software (version 17.0). Significance was tested by using Fisher's exact test and corrected by Holm's test for multiple testing (Supplementary Methods).

Results

Physical mapping and bioinformatic characterization:

All three deletions map to the long arm of the X chromosome in q27.3 (CNV64) and q28 (CNV67 and CNV69).

CNV64: this deletion, described in the Database of Genomic Variants (nsv829407), removes between 3.923-6.382 Kb of DNA considering its minimum (chrX: 143,436,346-143,440,268) and maximum (chrX: 143,434,786-143,441,167) CNV size, respectively. The presence of highly repetitive sequences in this region prevented a fine mapping of the deletion breakpoints. Nevertheless, based on the +/- STS pattern, we localized proximal and distal breakpoints within a region of 0.6 Kb upstream and 0.3 Kb downstream the proximal and distal edges of the minimum CNV size, respectively. No genes and regulatory elements are directly removed by this deletion. Nevertheless, a number of functional elements are located at <0.5 Mb from the deletion and may be affected (Table 1B).

CNV69: this deletion removes between 11.770-22.141 Kb of DNA, considering its minimum (chrX: 154,044,876-154,056,645) and maximum (chrX: 154,037,065-154,059,205) CNV size, respectively. Sequencing revealed the existence of at least two different deletion patterns referred to as type A and type B (Figure 1C). The alignment of the flanking sequences indicates that none of deletion types originates from homologous recombination thus the most likely mechanism is Non-homologous End Joining (NHEJ). This hypothesis is further supported by the presence of different additional nucleotides (NHEJ "molecular scars") compared to the reference sequence at deletion junction (Figure 1B). For instance, type A breakpoint was characterized by the insertion of a TAA tract (InsTAA), implying a deletion length of 16.06 Kb, whereas type B breakpoint displayed the insertion of a CATTCCATGTCCCC tract (InsCATTCCATGTCCCC) and the deletion measured 18.53 Kb. Breakpoint-specific primers allowed the detection of two fragments, each mirroring the deletion sizes: 700 bp for type A and 450 bp for type B (Figure 1C,D). Of 32 subjects (22 patients and 10 controls) carrying CNV69, 17 displayed type A (10 patients and 7 controls) and 10 displayed type B (9 patients and 1 control). In the remaining 5 subjects (3 patients and 2 controls), the deletion could not be classified neither as type A nor B and were thus referred to as type C.

A number of regulatory elements, but no genes, are directly involved (Table 1). Moreover, the region at <0.5 Mb surrounding the maximum CNV size includes several genes, of which *BRCC3* was reported as being associated with reproductive phenotypes.[5]

CNV67: based on array-CGH analysis, we previously reported that this deletion removes between 5.417-25.513 Kb of DNA, considering its minimum (chrX: 148,456,473-148,461,889) and maximum (chrX: 148,452,726-148,478,238) CNV size based on array-CGH, respectively.[4] Although the exact deletion breakpoints could not be defined, chromosome walking allowed us to better define the deletion extension, estimated to be 11.664 Kb. It is worth noticing that the highly repetitive nature of this region may predispose to inversions[6] and thus we propose two possible scenarios, represented in Supplementary Figure 2. Also in this case the most likely mechanism for deletion formation is NHEJ. The molecular characterization also highlighted that the proximal maximum size has been underestimated with array-CGH, probably due to the repetitive nature of the region. The distal deletion breakpoint was restricted to a 15.261 Kb region downstream the distal edge of the minimum size, which includes the proximal copy of *MAGEA9* and is 3.7 Kb far from the proximal copy of *HSFX1/2.*. Furthermore, the region at <0.5 Mb from the proximal and distal edges of the maximum CNV size includes also other genes and regulatory elements (Table 1).

Genotype-phenotype correlation

All three deletions were significantly more frequent in patients than controls (Table 1A). The carriers' phenotype is described in Supplementary Table 4. Estimating CNV69 deletion frequencies according to the type of breakpoint, we observed that type B breakpoint is significantly more frequent in patients (9/614; 1.5%) compared to controls (1/619; 0.2%) (*p* value=0.011; [OR=9.19 (CI=1.16-72.8)]). Concerning type A breakpoint, no statistically significant difference was found between patients (10/615; 1.6%) and controls (7/624; 1.1%) (Table 1A).

CNV64 and CNV69 carriers displayed a significantly lower Total Sperm Count (TSC) and Total Motile sperm Count (TMC) compared to non-carriers (*p* values <0.05; Supplementary Table 3).

CNV67 was found in 1.1% of patients and no controls with a *p*=0.008 (significant *p*<0.017 after correcting by the Holm's test) (Table 1A). Statistically significant difference was found between carriers and non-carriers for TMC (Supplementary Table 3).

Pedigree analysis: CNV67 was maternally inherited and both patients' mothers and MMP704's sister, also carrying CNV67, did not suffer from premature ovarian failure or anovulation. Interestingly, patient 11-041's brother did not carry CNV67 and was normozoospermic, differently from his oligozoospermic brother carrying CNV67 (Supplementary Figure 1).

Discussion

In this study, we provide evidence about the association of CNV64, CNV67 and CNV69 with spermatogenic failure through: i) a case-control study on a large study population; ii) the molecular characterization of deletions; iii) the exploration for functional elements in the region of interest and iv) a genotype-phenotype correlation analysis.

Concerning patient-enriched CNVs, CNV64 and CNV69, two clues are supporting their association with impaired spermatogenesis: i) deletion carriers had an increased probability (OR=1.9 and 2.2, respectively) of impaired spermatogenesis compared to non-carriers; ii) semen quality, in terms of TSC and TMC, was significantly impaired in carriers compared to non-carriers. Type B CNV69, the larger one, resulted significantly more represented in patients than controls suggesting that this deletion pattern may account for the potential deleterious effect of CNV69 on sperm production. This may be related to the closer position of type B deletion (7.8 Kb) to an upstream insulator compared to type A, the proximal breakpoint of which maps to 2.6 Kb downstream (Supplementary Figure 3). Importantly, through the breakpoint definition we

developed a simple diagnostic tool for type B deletion allowing other laboratories to further explore the role of this deletion in other ethnic groups.

We confirmed in >1,200 subjects the recurrent (1.1%) and "patient-specific" feature of CNV67. Unfortunately, it was impossible to obtain a fine mapping of CNV67 breakpoints for two reasons: i) presence of highly repetitive sequences; ii) incomplete assembly of the currently available reference sequence of the human X chromosome, derived from 16 different individuals.[6] However, we propose that the removal of the proximal *MAGE9A* gene (situated within the newly defined maximum size) and/or its regulatory elements are the most likely determinants of the observed infertile phenotype.

We previously suggested that Cancer Testis gene dosage variation may play a role in CNV-related spermatogenic failure;[4] accordingly *MAGEA9* belongs to this gene family. *MAGEA9* is an ampliconic gene reported as independently acquired on the human X chromosome, since no orthologs could be detected in the murine X chromosome.[6] Independently acquired X-linked genes are predominantly expressed in the testis with a specific expression of multi-copy genes in male germ cells.[6] Therefore, it can be speculated that the loss of *MAGE9A* copies would affect spermatogenesis. Moreover, according to *Gene Ontology (GO) MAGEA9* is mainly involved in 'Regulation of gene expression', 'DNA methylation', 'Reproduction' and 'Spermatogenesis', reflecting its potential involvement in transcriptional and epigenetic regulatory mechanisms of gametogenesis.[5] The pathogenic effect of CNV67 could depend not only on the removal of the proximal gene copy but also on its influence on regulatory elements of the distal copy. For instance, large-scale CNVs might change the 3D structure of chromatin, which is seemingly crucial for correct gene regulation.[7–9] *UniprotKb* database reports that MAGEA9 could be involved in embryo development, whereas the *Consensus pathway* database indicates a physical interaction with the transcription factor HNF1A. GO terms for HNF1A include "blastocyst development", therefore we could speculate that loss of *MAGEA9* might also affect embryo development, providing another explanation to the infertile status of our deletion carriers. Recently, a private control-specific deletion of the more distal *MAGEA9* copy was recently reported in a normozoospermic man. This may

indicate either that there could be functional differences between the two *MAGE9A* copies or that the normozoospermic carrier may still face fertility problems due to the putative involvement of the gene in early embryonic development (normozoospermia is not a synonym of fertility, and no data on his fertility status were reported by Lopes *et al*).[3] Furthermore, this deletion may also affect regulatory elements of another independently acquired X-linked multi-copy gene, *HSFX1/2* with testis-specific expression.

Pedigree analyses of two CNV67 carriers indicated that this deletion is maternally inherited, thus not affecting female fertility. This is in accordance with the lack of expression of *MAGE9A* in the ovary. Patient 11-041's family is especially informative since the pathological semen phenotype of the carrier (11-041) *versus* his normozoospermic non-carrier brother is a strong indicator for a pathogenic effect of the deletion on spermatogenesis.

For the first time, we provide evidence for a significant association between recurrent X-linked deletions and impaired sperm production. Strikingly, CNV67, specific to spermatogenic failure, resembles AZF deletions on the Y chromosome. This finding merits further studies in order to disentangle the structural complexity of this region and provide a feasible substrate for fine molecular characterization and large scale diagnostic testing.

Table1. Case-control study and analysis of regions involved in CNV64, CNV67 and CNV69

A. Case-control study and list of genes mapping inside and within CNVs flanking regions.

X chr. band	CNV Code	Patients (frequency)	Controls (frequency)	*p* value [OR (95%CI)]	Genes inside and nearby
Xq27.3	**64**	36/627 (5.7%)	19/628 (3.03%)	0.013 [1.89 (1.10-3.27)	-
Xq28	**67**	7/627 (1.1%)	0/628 (0.0%)	0.008	*IDS; LINC00893; CXorf40A*; **MAGEA9B**; HSFX2§; HSFX1§; TMEM185A; MAGEA11; HSFX1; HSRX2; MAGEA9B; MAGEA; CXorf40B*
Xq28	**69**	22/627 (.5%)	10/628 (1.59%)	0.033 [2.2 (1.05-4.62)	
	Type A	10/615 (1.6%)	7/624 (1.1%)	0.301(n.s.)	*GAB3; DKC1; SNORA36A; SNORA56; MPP1; SMIM9; F8; H2AFB3FUNDC2;CMC4; MTCP1; BRCC3*; VBP1§; RAB39B;CLIC2; TMLHE-AS1; H2AFB3; F8A1;F8A1;H2AFB1; TMLHE-AS1*
	Type B	9/614 (1.5%)	1/619 (0.2%)	0.011 (9.19; 1.16-72.8)	
	Type C	3/608 (0.5%)	2/620 (0.3%)	0.491 (n.s.)	

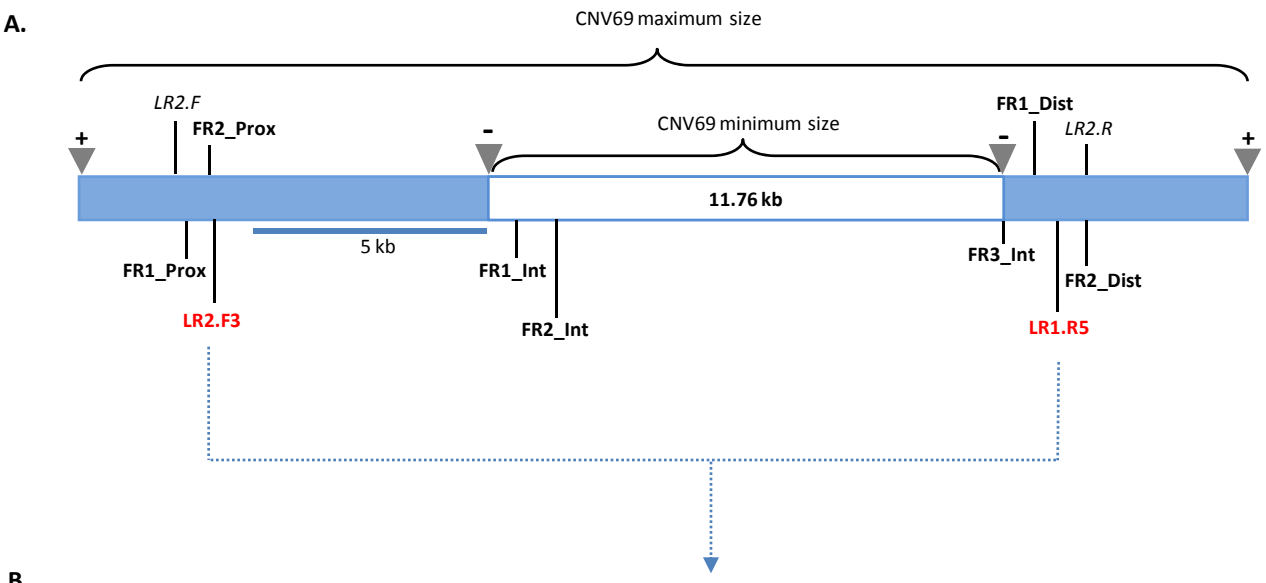
Only regulatory elements and genes located <500 Kb from proximal/distal maximum CNV size border are considered. n.s.= not significant.
*The first proximal flanking gene
§The first distal flanking gene

B. Regulatory elements detected within the three CNVs

CNV Code	Position	Number of Regulatory Elements†						
		weak enhancer	***strong enhancer***	***weak promoter***	***active promoter***	***weak transcribed region***	***transcription elongation***	***insulator***
64	*Upstream proximal edge*	10	-	-	-	2	-	4
	Inside the deletion	-	-	-	-	-	-	-
	Downstream distal edge	11	3	1	-	2	-	7
67	*Upstream proximal edge*	57	19	7	3	11	4	8
	Inside the deletion	6	2	-	-	1	-	-
	Downstream distal edge	70	17	10	4	16	1	16
69	*Upstream proximal edge*	110	38	21	7	28	7	16
	Inside the deletion	-	-	-	-	-	-	-
	Downstream distal edge	68	9	25	7	8	2	4

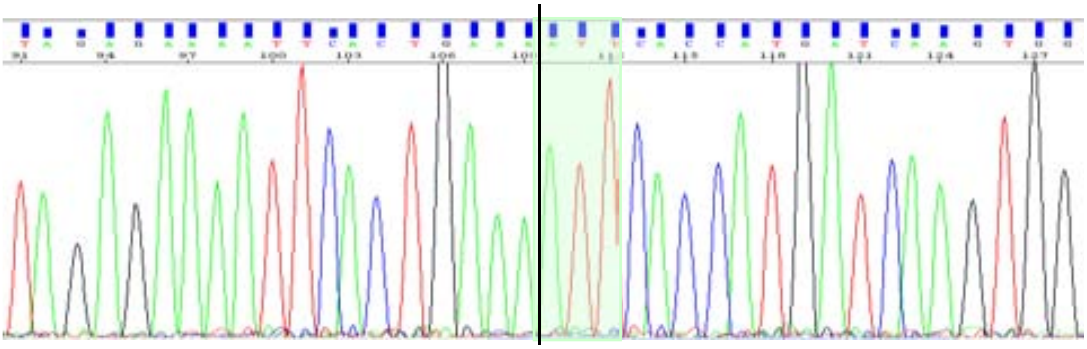
† Predicted functional elements in normal 4 human cell lines from ENCODE (cell lines: GM12878-B lymphocyte, lymphoblastoid, International HapMap Project, H1-hESC-embryonic stem cells, HUVEC-umbilical vein endothelial cells, NHER-epidermal keratinocytes).

Figure 1.



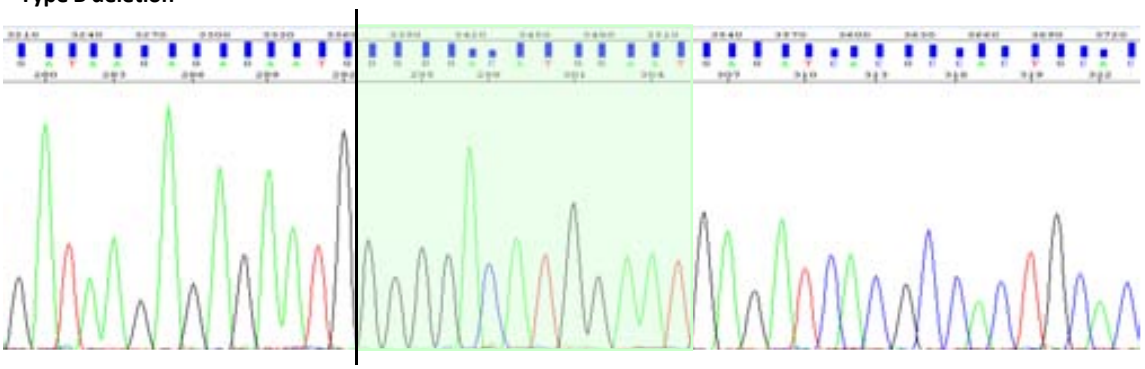
B.

Type A deletion



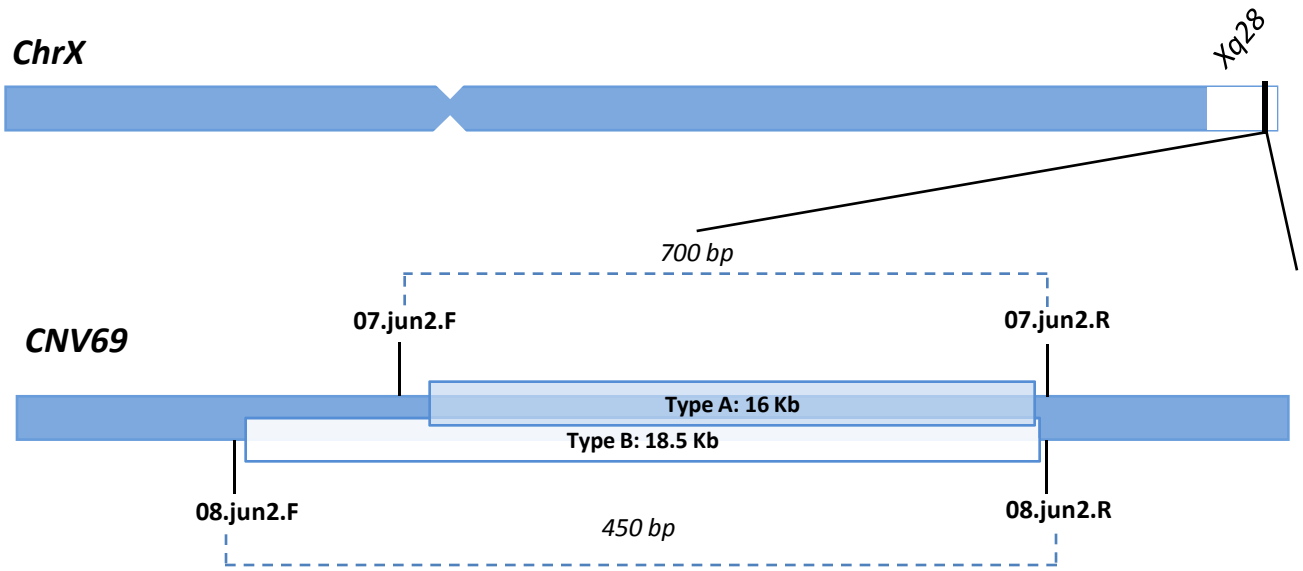
Proximal	TAGAGAAAATTCACTGAAACTAAAATTTATAATCCTTGT
Type A	TAGAGAAAATTCACTGAAAATTCACCATGATCAAGTGG
Distal	GCAGCACATCAAAAAGCTTACCCACCATGATCAAGTGG

Type B deletion



Proximal	GATAAGAGAGAATGTGAACCAAGCGAAAGGAGTTTCCTCTTATAAA
Type A	GATAAGAGAGAATG GGGGACATGGAATGAGATCACGCCACTGCAC
Distal	ACCCAGGTGGTGGAGGTTGCAGTGAGCCGAGATCACGCCACTGCAC

C.



D.

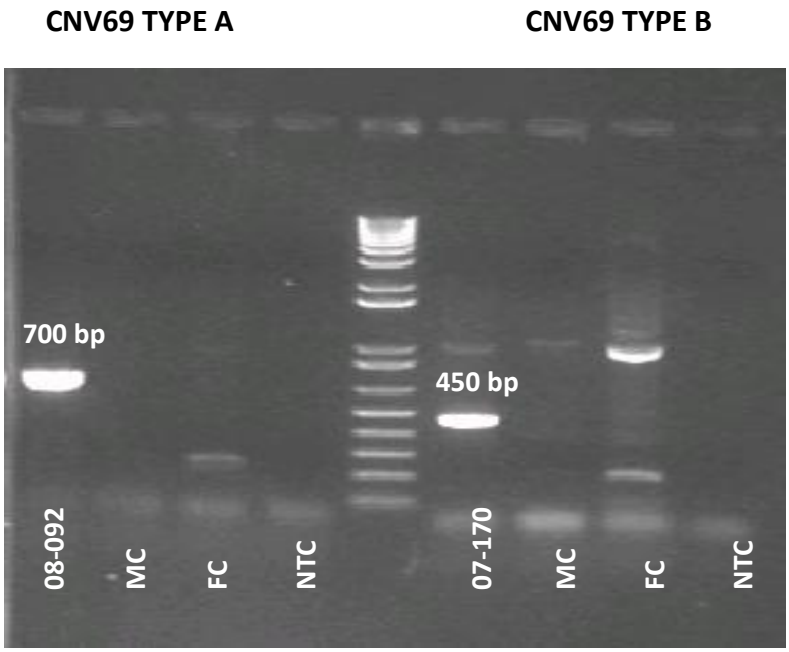


Figure 1. Fine mapping of CNV69 deletion breakpoint. (A) **Schematic representation of the minimum and maximum CNV size**. Triangles (+) depict the first and last positive array-CGH probes, delineating the maximum CNV size, and triangles (-) depict the first and last negative array-CGH probes, delineating the minimum CNV size. Primers used for chromosome walking are indicated in **bold; Prox:** primers located upstream the proximal edge of the minimum CNV size; **Dist:** primers located downstream the distal edge of the minimum size; **Int:** primers inside the minimum CNV size. Primers used for junction amplification and sequencing are indicated in *italic* and red, respectively. (B) **Sequencing of deletion junction** in Type A and Type B breakpoints. Sequences shown were obtained with LR.R5 reverse primer. Insertions compared to the reference sequence at the deletion junction are highlighted in green and the alignment of the sequences flanking the breakpoint is shown in the lower panel of each electropherogram. (C) **Type of deletions**. Breakpoint-specific primers used for screening of CNV69 carriers are indicated in **bold**. Dotted lines depict the amplification product obtained for each type of deletion. (D) **PCR product using breakpoint-specific primers.** Electrophoresis of the two fragments obtained, each mirroring the deletion sizes. No amplification was obtained in male and female controls. MC: male control; FC: female control; NTC: no template control.

Acknowledgments: The authors wish to thank the clinicians from the Andrology Unit of the University of Florence (M. Maggi, A. Magini, F. Lotti), from the Fundació Puigvert (L.Bassas, O.Rajmil, J.Sarquella, A.Vives, J.Sanchez-Curbelo) and from University of Ancona (G. Balercia) who helped in providing samples for this study. We also thank Dr. H. Skaletsky and Dr. F. Daguin for useful discussions and E. Rossi (University of Pavia, Italy) for her input on array-CGH data. A special thank to Esperança Martí, President of the Fundacio Puigvert for her continuous support. The IIB Sant Pau-Fundació Puigvert Biobank is acknowledged for kindly providing part of DNA samples.

Competing interests: None

Funding: This work was funded by grants from the Spanish Health Ministry FIS 11/02254, Italian Ministry University (Grant PRIN 2010-2012 to CK).

Patient Consent: All participants signed an informed consent.

Ethics Approval: The local Ethical Committees of the University Hospital Careggi and the Fundació Puigvert approved the study.

References

1 Krausz C. Male infertility : Pathogenesis and clinical diagnosis. *Best Pract Res Clin Endocrinol Metab* 2011; 25(2):271-85.

2 Tüttelmann F, Simoni M, Kliesch S, Ledig S, Dworniczak B, Wieacker P, Röpke A. Copy number variants in patients with severe oligozoospermia and Sertoli-cell-only syndrome. *PloS one* 2011;6(4):e19426.

3 Lopes AM, Aston KI, Thompson E, Carvalho F, Gonçalves J, Huang N, Matthiesen R, Noordam MJ, Quintela I, Ramu A, Seabra C, Wilfert AB, Dai J, Downie JM, Fernandes S, Guo X, Sha J, Amorim A, Barros A, Carracedo A, Hu Z, Hurles ME, Moskovtsev S, Ober C, Paduch DA, Schiffman JD, Schlegel PN, Sousa M, Carrell DT, Conrad DF. Human spermatogenic failure purges deleterious mutation load from the autosomes and both sex chromosomes, including the gene DMRT1. *PLoS genetics* 2013;9:e1003349.

4 Krausz C, Giachini C, Lo Giacco D, Daguin F, Chianese C, Ars E, Ruiz-Castañe E, Forti G, Rossi E. High resolution X chromosome-specific array-CGH detects new CNVs in infertile males. *PloS one* 2012;7:e44887.

5 Chalmel F, Lardenois A, Evrard B, Mathieu R, Feig C, Demougin P, Gattiker A, Schulze W, Jégou B, Kirchhoff C, Primig M. Global human tissue profiling and protein network analysis reveals distinct levels of transcriptional germline-specificity and identifies target genes for male infertility. *Human reproduction (Oxford, England)* 2012;27:3233–48.

6 Mueller JL, Skaletsky H, Brown LG, Zaghlul S, Rock S, Graves T, Auger K, Warren WC, Wilson RK, Page DC. Independent specialization of the human and mouse X chromosomes for the male germ line. *Nature Genetics* 2013.

7 Sanyal A, Lajoie BR, Jain G DJ. The long-range interaction landscape of gene promoters. *Nature* 2012; 489(7414):109-13.

8 Yip KY, Cheng C, Bhardwaj N, Brown JB, Leng J, Kundaje A, Rozowsky J, Birney E, Bickel P, Snyder M GM. Classification of human genomic regions based on experimentallY determined binding sites of more than 100 transcriptio-related factors. *Genome biology* 2012; 13(9):R48.

9 Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan K, Cheng C, Mu XJ, Khurana E, Rozowsky J, Alexander R, Min R, Alves P, Abyzov A, Addleman N, Bhardwaj N, Boyle AP, Cayting P, Charos A, Chen DZ, Cheng Y, Clarke D, Eastman C, Euskirchen G, Frietze S, Fu Y, Gertz J, Grubert F, Ouyang Z, Partridge EC, Patacsil D, Pauli F, Raha D, Ramirez L. Architecture of the human regulatory network derived from ENCODE data. *Nature* 2012;488:91–100.

SUPPLEMENTAL DATA

Recurrent X chromosome-linked deletions: discovery of new genetic factors in male infertility

Authors: D. Lo Giacco [1,2], C. Chianese [3], E. Ars [1,2], E.Ruiz-Castañé [2], G.Forti [3] and Csilla Krausz [2,3]

1 Molecular Biology Laboratory, Fundació Puigvert Universitat Autònoma de Barcelona, Barcelona, Instituto de Investigaciones Biomédicas Sant Pau (IIB-Sant Pau), Barcelona, Spain.

2 Andrology Service, Fundació Puigvert, Universitat Autònoma de Barcelona, Instituto de Investigaciones Biomédicas Sant Pau (IIB-Sant Pau), Barcelona, Spain.

3 Department of Experimental and Clinical Biomedical Sciences, University of Florence, Italy.

Corresponding author:

Csilla Krausz MD, PhD

Department of Experimental and Clinical Biomedical Sciences, University of Florence;

Viale Pieraccini 6- 50139 Florence, Italy. Phone: +39 055 4271415 – Fax: +39 055 4271413 ; email: c.krausz@dfc.unifi.it

CONTENT:

Supplementary Materials

Subjects

The patient group included 163 azoospermic (AZ), 325 severe oligozoospermic (SOZ; 0<Sperm Concentration (SC)≤5x10^6/ml), 138 mild oligozoospermic (MOZ; 5x10^6/ml<SC<20x10^6/ml) and one normozoospermic (SC ≥ 20x10^6/ml, but male partner of a couple with recurrent miscarriage). Infertile patients were selected on the basis of a comprehensive andrological examination including medical history, semen analysis, scrotal ultrasound, hormone analysis, karyotype and Y chromosome microdeletion screening. Patients with mono- or bilateral cryptorchidism, varicocele grades 2 and 3, obstructive azoospermia, recurrent infections, iatrogenic infertility, hypogonadotrophic hypogonadism, karyotype anomalies, Y chromosome microdeletions and patients with non-Italian or non-Spanish origin were excluded. Controls in the Spanish cohort were fertile normozoospermic men undergoing pre-vasectomy, whereas the Italian control cohort included 70% of normozoospermic volunteers not belonging to infertile couples and 30% normozoospermic men with proved fertility.

Supplementary Methods

Germline DNA was extracted from peripheral blood, germ cells or buccal swab in all participants with standard methods.

Detection and validation of CNV in the study population

First step screening: STSs +/- PCR was performed in a final volume of 10µl containing 70ng of genomic DNA, 400µM deoxynucleotides triphosphates, 3mM MgCl2, 10µM of specific primers, 50 U/ml of Taq DNA Polymerase (Promega PCR MASTER MIX 2X). Annealing temperatures were optimized between 58°C and 60°C, according to the primers' melting temperature. We used the following amplification conditions: initial denaturation at 95°C for 5 min; 35 cycles of denaturation at 95°C, annealing at 58°C and extension at 72°C for 30sec each and a final extension at 72°for 5min.

Validation steps: The second confirmatory step consisted in confirming suspected deletions by both lowering the annealing temperature (55°C) and performing additional PCRs with alternative primers (Supplentary Table 1).

Analysis of female DNA

Screening of CNV67 carriers' female relatives was performed through quantitative PCR (qPCR) using a pre-designed TaqMan® Copy Number Assay (Hs03323870_cn) (Applied Biosystems) located within the CNV67 minimum size and RNaseP as internal reference. All samples were analyzed in triplicate. The reaction mix component (final volume 20 ul) were: 1X TaqMan® Universal Master Mix, 1X TaqMan® Copy Number Assay, 1X TaqMan® Reference Copy Number Assay (RNaseP), 10 ng of genomic DNA.

CNV69 breakpoint definition

Amplification and sequencing of deletion junction

LR-PCR reaction (Advantage 2 PCR kit from Clonetech) was performed in a 25µl final volume, adding 200ng of genomic DNA, 10X Advantage 2 PCR Buffer, 10µM of each primer, 2.5mM deoxynucleotides triphosphates, 50X Advantage 2 Polymerase Mix. A negative template control was always used to exclude the hypothesis of PCR contaminations. We used the following LR-PCR conditions: initial denaturation at 95°C for 1min; 35 cycles of denaturation at 95°C for 30sec and annealing at 68°C for 8 min; final extension at 68°C x 10min. The sequence of primers is shown in Supplementary Table 2.

The LR-PCR product was then sequenced (using the same primers as for LR-PCR) by automatic sequencer using manufacturer's instructions (ABI PRISM 3130, Applied Biosystem, Foster City, CA, USA).

Classification CNV69 carriers according to deletion type

In order to classify CNV69 carriers according to deletion breakpoints (Type A and Type B), we performed a conventional +/- PCR screening using breakpoint-specific primers: 07.jun2.F (Forward: 5'- 3') and 07.jun2.R (Reverse: 5'-3') for deletion Type A; 08.jun2.F (Forward: 5'-3') and 08.jun2.R (Reverse: 5'-3') for deletion Type B. PCR condition conditions were: initial denaturation at 95°C for 5 min; 35 cycles of denaturation at 95°C, annealing at 58°C and extension at 72°C for 30sec each and a final extension at 72°for 5min.

Statistical Analysis

Statistical analysis was performed using SPSS software (version 17.0). We tested the significance of the frequency of CNVs between patients and controls using Fisher's exact test. We compared Total Sperm Count (TSC) and Total Motile sperm Count (TMC) between CNV carriers and non-carriers using the non-parametric Mann-Whitney's test. To correct for multiple testing, *p* values were adjusted according to Bonferroni-Holm procedure referring to

as significant the following corrected p values: i) p <0.017 for CNV67; ii) p <0.025 for CNV64; iii) p <0.05 for CNV69.

Supplementary Tables and Figure

Supplementary Table 1. Plus/minus PCR primers used for detection and confirmatory steps

CNV code	Primers for first step screening		Primers for validation	
64	del Xq27.3-F1	CGGCCAAAGTATTCTGGGTA	del Xq27.3-F2	ACCAACTGGTGTGGCTAAGG
			del Xq27.3-R2	CACGTGACTCTGACCAGCAT
	del Xq27.3-R1	CAGGGGAGGTGTGATAGCAT	del Xq27.3-F3	GGCACTGAAAAGATGGGAAA
			del Xq27.3-R3	GGTTTCACTGCTACCCTCCA
67	del Xq28-F1	GCCAGGCAGAAGAGTACCTG	del Xq28-F2	ACAGCATAGGCTCTGGAGGA
			del Xq28-R2	CCCTGAGCTAGGTGCTTCAC
	del Xq28-R1	CTGTGCTCCTGTTCCAGTGA	del Xq28-F3	CACTGGAACAGGAGCACAGA
			del Xq28-R3	TGTGGAGCAACAGGTGAGAG
69	del Xq28B2-F1	AGGGAAATGGAAACCAAAGG	del Xq28B2-F2	GCCAGAGGAAATGGCTTGTA
			del Xq28B2-R2	GATGGGTAGGTGGAGAACCA
	del Xq28B2-R1	TCTGAAGGCAAAGGATCACA	del Xq28B2-F3	TTACAGGTGTGTGCCACCAT
			del Xq28B2-R3	TGTCAACTGGTTGGAACAGG

Supplementary Table 2. Primers used for mapping of deletion breakpoints

CNV code	Primers		Product size (bp)	STS status (present/absent)
64	**CNV64_prox_F1**	GCTGAATTGTTGGCTGGATT	181	present
	CNV64_prox_R1	CAAAAGCACCACTCGTGAAA		
	CNV64_prox_F2	GATGTGCGTTTTGCAAATGT	180	present
	CNV64_prox_R2	CCTAATCTTGGGTTGGACTATGA		
	CNV64_Int_F1	TAGGTATCAGCAGGGCCAAC	225	absent
	CNV64_Int_R1	GCCATGAGGAGTTCTTCCAG		
	CNV64_Int_F2	CATGCCTTGTGGTCTGAGTG	205	absent
	CNV64_Int_R2	GCTTAATTTCCTGGGGCTCT		
67	**CNV67_prox_F1**	AGGGTATGTCCCAGCAACAG	665	present
	CNV67_prox_R1	TCCTTGGTTCCTCATTTTGC		
	CNV67_prox_F2	AGCAAGGACAGGAGCACACT	413	present
	CNV67_prox_R2	AAGGACAGCCTGAAAGGGTT		
	CNV67_Int_F1	GACACACCTGTCCATCCCTT	760	absent
	CNV67_Int_R1	TCACCACCGTTCATCTCAAA		
	CNV67_ Int _F2	CCAGGAATCCACTGATAAATGA	267	absent
	CNV67_ Int _R2	GCTCCCCATTCCCTCTTATC		
	CNV67_ Int _F3	GATTTGAAGATCGAGCTGGC	884	absent
	CNV67_ Int _R3	AAGTGAAGGACACCACAGGG		
	CNV67_dist_F1	CCATTTTCTACAGTGCCCGT	783	present
	CNV67_dist_R1	TTCTGCCCTTCATCCTCTGT		
	CNV67_dist_F2	CTCTGCCTAATTGCAAACCC	332	present
	CNV67_dist_R2	TTCTGCCCTTCATCCTCTGT		
	CNV67_dist_F3	TTGATGAAGCCATGGGCA	556	present
	CNV67_dist_R3	GCCCCTATTCCAGTGTGAAA		
	CNV67_dist_F4	AGCAGGAATGCCCCAGTAGT	559	present
	CNV67_dist_R4	CTTTTCCCTGGGCTTTGG		
69	**CNV69_prox_F1**	TATGGGACCGAGGTTTTCAG	155	present
	CNV69_prox_R1	GAGCCCCTAGGTGAAGGAAT		
	CNV69_prox_F2	TGTCACCCCAAACTGTTCA	349	present
	CNV69_prox_R2	GCTAAGGCTGGTTGCAGTGT		
	CNV69_Int_F1	CCTTCTCTTAGACCACAAGCC	277	absent
	CNV69_ Int _R1	CCACATTACCTGCAAACCAA		
	CNV69_ Int _F2	GAGAGATGGATGGATAGATGGA	156	absent
	CNV69_ Int _R2	CAACTGGACTATTTGGGAAGGA		
	CNV69_ Int _F3	TTTTGGTAGGGTTTGGGAAG	194	absent
	CNV69_ Int _R3	CATCTTCTTTACTGCAAACCCC		
	CNV69_dist_F1	TCAATGTCTATGCAAACTGTAGTGA	305	present
	CNV69_dist_R1	AATGTGAACCAAGCGAAAGG		
	CNV69_dist_F2	TGCTTGATCTGGGTTTGCTT	497	present
	CNV69_dist_R2	AGGAAGGTGGATCACCTGAG		
	CNV69_dist_F3	GCTCAGTTTCCACAGCCTCT	358	present
	CNV69_dist_R3	CTGAGGGAGGGAAACATTCA		
	Primers used for LR-PCR			
	LR2.F3	CGCCCAGGTTGTAGTGCAGTG	3500 (type A) 900 (type B)	
	LR1.R5	AAATTTCCACCCCAAGAAAC		
	Breakpoint-specific primers			
	07.JUN2.F	GTGCCAGTTTTCAAAGGGAA	700	
	07.JUN2.R	GCAGGGCTTAGAGGGAATTT		
	08.JUN2.F	GTCAGAGATTGGGATGGGAA	450	
	08.JUN2.R	AAGAGCAAAGCAAACCCAGA		

Supplementary Results

Supplementary Table 3. Impact of CNV64, CNV67, CNV69 on Total Sperm Count (SC) and Total Motile sperm Count (TMC).

	TSC Median (25 th-75 th)	**TMC** Median (25 th-75 th)	***p* value (TSC)**	***p* value (TMC)**
CNV64				
carrier	9.25 (0.27-173.75)	0.732 (0.0-73.42)	0.011	0.009
non-carrier	55.5 (3.9-253-81)	13.05 (0.10 -143.31)		
CNV67				
carrier	4.9 (0.51-31.14)	0.0 (0.0-7.12)	0.055	0.008
non-carrier	50.2 (3.0-252.0)	11.5 (0.08-139.17 9)		
CNV69				
carrier	8.75 (0.01- 160.0)	0.843(0.0-93.57)	0.051	0.036
non-carrier	50.4 (3.6-252.0)	11.65 (0.08-139.33)		

Supplementary Table 4. Phenotypic description of CNV64, CNV67 and CNV69 carriers

CNV code	Carrier ID	Nx10^6 sperm/ml	Mean testicular volume (ml)	FSH (U/L)	LH (U/L)	Testosterone (nmol/L)
CNV67	05-196	0	16	30.7	6.03	23.3
	11-041	7.3	17.5	N	N	N
	10-314	5	15	N	N	N
	MMP676	5	14	N	2.9	5.6
	MMP687	11	14	6.9	4.8	4.9
	MMP704	0.3	14	9.7	7.8	4.5
	A670	3.05	N	1.83	1.74	13.8
CNV64	08-108	2	15	3.6	3.6	5.03
	04-155	0.001	12	21.5	9.33	3.56
	05-062	0.028	11	6.9	9.4	5.14
	06-129	0	N	36.1	N	N
	07-341	0	15	15	N	7.66
	08-117	0.023	15	6.04	N	N
	08-284	0.0325	7	39.1	N	N
	08-532	0.019	20	13.3	2.5	N
	09-043	3	15	3.87	6.16	18.8
	09-068	16	22	1.73	16.9	N
	08-079	0	22	3.75	3.61	14.7
	09-465	19	22.5	N	N	N
	10-078	0.002	15	8.5	7.9	17
	10-087	0.44	20	2.34	N	N
	10-221	1	20	12.5	N	N
	09-162	13.3	14	N	N	N
	06-146	0.178	17.5	3.3	N	N
	07-411	0	15	N	N	N
	07-463	0	11	28.9	10.7	11.2
	09-345	0	22.5	3.94	N	N

Supplementary Table 4. Continue

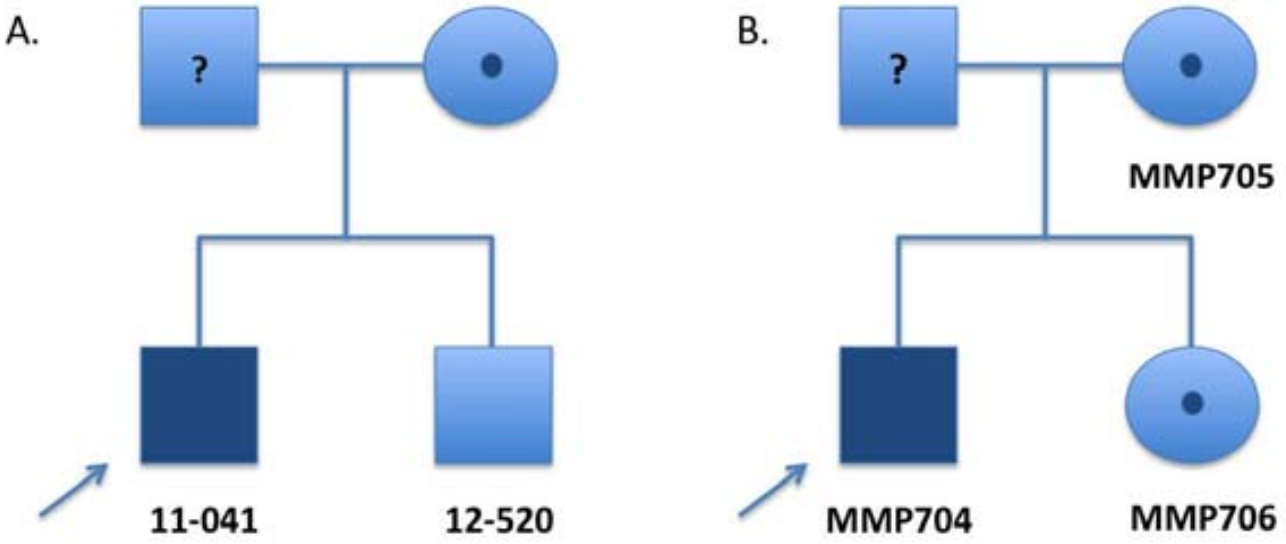
CNV code	Carrier ID	Nx10^6 sperm/ml	Mean testicular volume (ml)	FSH (U/L)	LH (U/L)	Testosterone (nmol/L)
CNV64	10-471	0.059	15	2.05	N	N
	12-232	0.07	17.5	13.3	4.9	12.1
	A224	7.00	N	1.92	2.63	16.34
	A241	0.00	12	5.70	27.60	3.26
	A414	0.60	15	12.0	4.00	4.88
	A567	9.40	10	11.2	4.6	6.70
	A616	0.01	21.5	7.2	2.00	5.42
	A828	6.3	13.5	12.05	7.4	3.22
	A850	6.5	16	N	N	N
	A1020	1.1	14.5	3.5	4.04	17.4
	MMP593	3	N	9.4	3.7	3.2
	MMP597	6	N	3.3	2.6	4.9
	A2	0.01	N	N	N	N
	M14	N	N	N	N	N
	M15	N	N	N	N	N
	A1260	1	27.5	4.8	4.24	25.5
CNV69	07-030	0	16	34	5.23	11.1
	07-170	0.02	17	5.8	N	N
	08-092	0	12	23	5.53	23.9
	08-498	0	15	17.6	N	N
	09-385	3.5	14	N	N	N
	05-205	0	15	13.2	N	14
	06-220	0.13	11	23	5.9	9.15
	10-316	6.15	22.5	2.55	N	N
	06-190	2	N	5.35	6.05	N
	A371	0.01	16.5	11.38	7.18	4.60
	A525	0	18	3.9	5.8	3.32
	A576	0	16	3.54	3.38	12.5

Supplementary Table 4. Continue

CNV code	Carrier ID	Nx10^6 sperm/ml	Mean testicular volume (ml)	FSH (U/L)	LH (U/L)	Testosterone (nmol/L)
CNV69	A609	2.4	15	7.96	8.45	15.4
	MMP52	7	N	2.40	3.60	4.80
	MMP511	3	N	10.50	12.30	2.90
	MMP604	7	N	9.8	4.3	3.4
	MMP687	11	14	6.9	4.8	4.9
	A955	0.5	13.5	N	N	N
	A1293	0.375	22.5	6.33	3.86	11.9
	A947	7.5	21	N	N	N
	A628	84*	N	N	N	N
	A1222	0	12.5	24.2	10.1	11.8

N= referred as in the normal range according to medical history

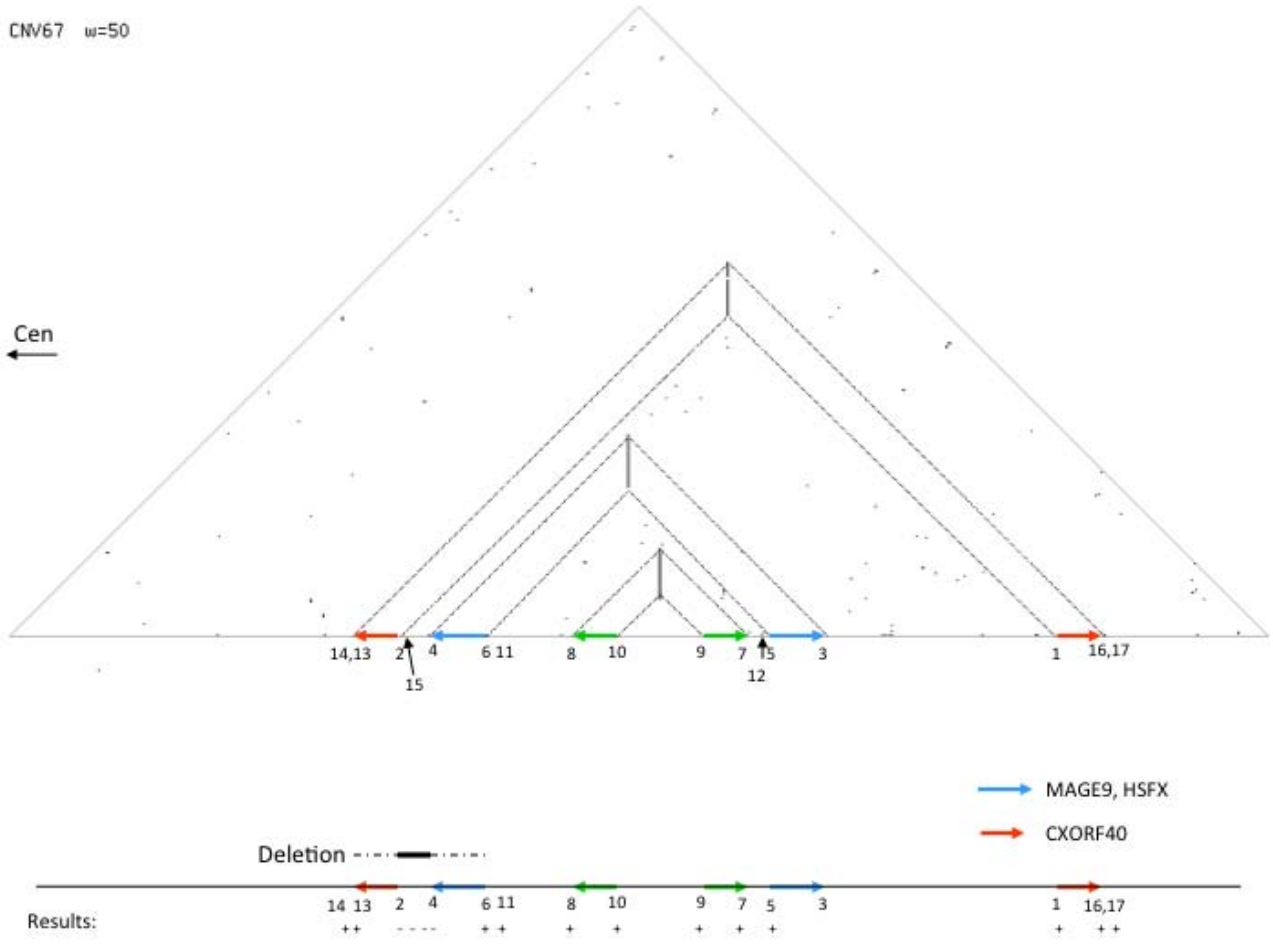
Supplementary Figure 1. Family Pedigree



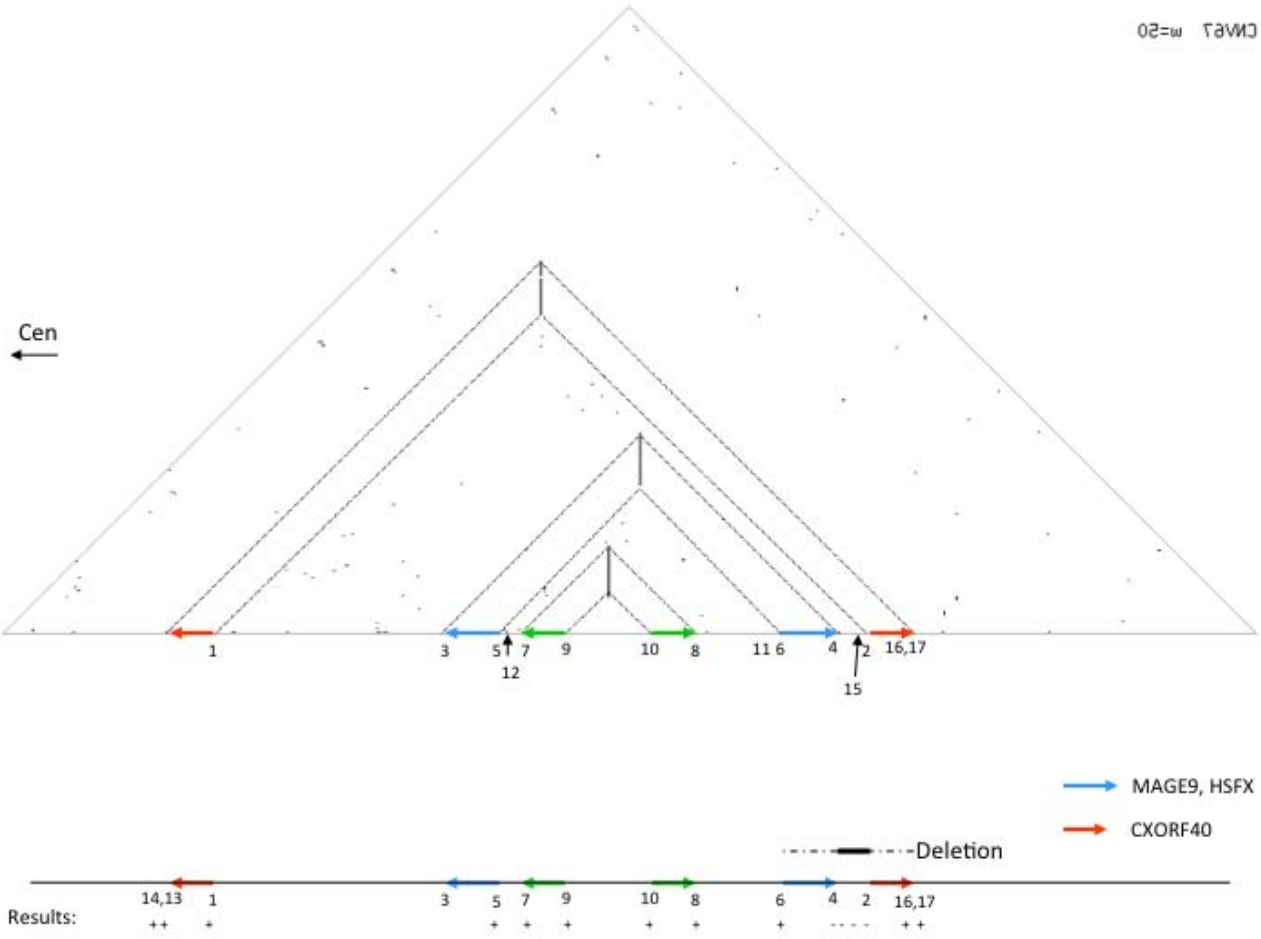
CNV67 analysis in carriers' relatives. A) Patient 11-041's family pedigree: the proband's mother carried the deletion, whereas his brother did not and had a normozoospermic semen phenotype. B) Patient MMP704's family pedigree: both the proband's mother and sister carried the deletion. No premature ovarian failure or anovulation in the mother and the sister were reported.

Supplementary Figure 2. Scenarios for CNV67 flanking region

A.



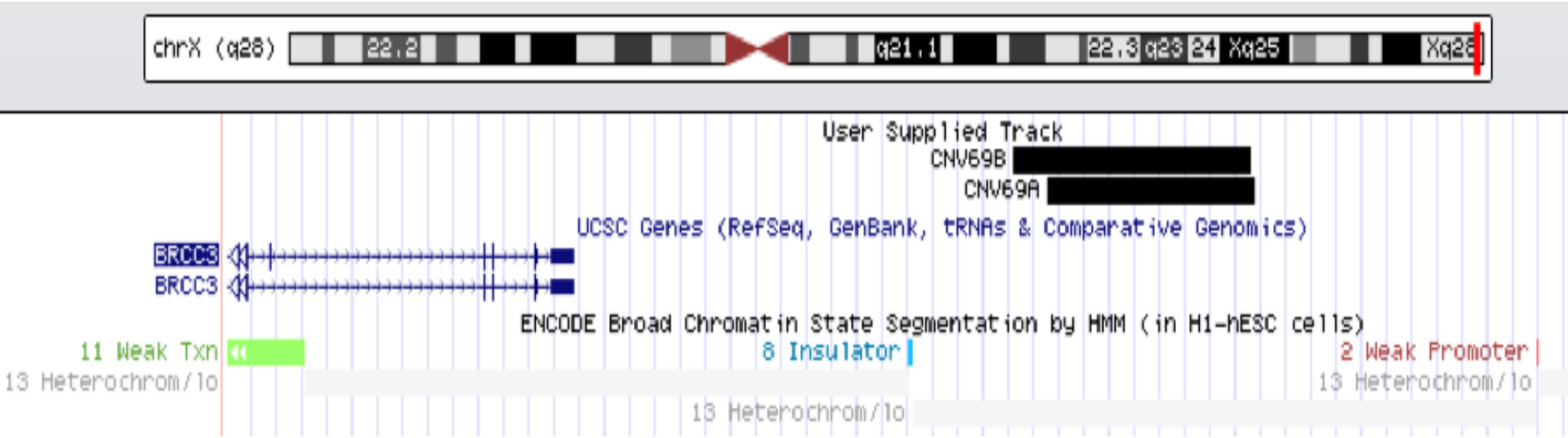
B.



Possible configurations of CNV67 flanking region. A highly repetitive region rich in ampliconic sequences, which might undergo inversions, surrounds CNV67. The triangular dot plot highlights these sequence similarities: A) Arrows of the same color represent homologous amplicons. CNV67 (Deletion) location is denoted: the thick black line represents the deletion ascertained by chromosome walking, whereas the potential extension is depicted by the dotted line. The centromere (cen) and the genes included within the amplicons are also indicated. Each number corresponds to a couple of primers used for the chromosome walking and listed in Supplementary Table 2: *14= prox_F+R1; 13= prox_F+R2; 2= int_F+R1; 15= int_F+R2; 4= int_F+R3; 6= dist_F+R1; 11= dist_F+R2; 8= dist_F+R3; 10= dist_F+R4; 7= dist_F+R5; 12= dist_F+R6; 5= dist_F+R7; 3= dist_F+R8; 1= dist_F+R9; 16= dist_F+R10; 7= dist_F+R11.* B) Representation of a possible inversion event for which the entire segment between red amplicons may be inverted, thus changing the location of CNV67 from the reference proximal red amplicon to the distal red amplicon.

Supplementary Figure 3. UCSC representation of the two types of deletion identified for CNV69. The CNVs localization in respect to the *BRCC3* gene and the insulator is shown. A number of regulatory elements located within the flanking region are also indicated. Being the larger, type B deletion (7.8 Kb) maps closer to an upstream insulator compared to type A.

Supplementary Figure 3. UCSC representation of the two types of CNV69



6. DISCUSSION

A considerable amount of research in the field of male infertility has focused on identifying possible underlying genetic factors. Nevertheless, our understanding of the genetics of male infertility is still limited. The present thesis addressed this topic by exploring the role of specific genetic variants in the etiology of two forms of disturbed male reproductive fitness: cryptorchidism and idiopathic spermatogenic failure.

6.1 The complex etiology of non-syndromic cryptorchidism: a multifactorial polygenic disesase

Cryptorchidism is a well-characterized risk factor for impaired sperm production and testicular cancer in adulthood. This common congenital birth defect may lead to (especially bilateral forms) or probably reflect (based on the Testicular Dysgenesis Syndrome theory) a primary testicular dysfunction/dysgenesis.

While our understanding of the physiology of human testis descent has significantly increased during the last decades, the etiology of the most common non-syndromic forms of cryptorchidism remains elusive.

Animal models have substantially contributed to our understanding of the mechanisms regulating testicular descent and suggested a number of candidate genes for human cryptorchidism (Table 3) even though, in most of the cases findings from animal models cannot be translated to human.

Table 3. Animal models with cryptorchidism. From *Foresta et al 2008*

Animal models	Testicular position	References
Insl3 $^{-/-}$ (INSL3 deficiency)	Intrabdominal	Nef & Parada 1999; Zimmermann et al. 1999
Csrp (Great$^{-/-}$, Rxfp2$^{-/-}$) (INSL3 receptor deficiency)	Intrabdominal	Overbeek et al. 2001
Hoxa10$^{-/-}$ (HOXA 10 deficiency)	Intrabdominal	Satokata et al. 1995
Hoxa11$^{-/-}$(HOXA 11 deficiency)	Intrabdominal	Hsieh-Li et al. 1995
P450arom^{+} transgenic (aromatase overexpression)	Intrabdominal	Li et al. 2001
Hpg (GnRH deficiency)	Inguinoscrotal	Charlton et al. 2001
Gnrhr$^{-/-}$ (GnRH receptor deficiency)	Inguinoscrotal	Pask et al. 2005
Gnrh-promoter driven SV40-T mutants (LH receptor/FSH receptor deficiency)	Inguinoscrotal	Radovick et al. 1991
LuRKO (LHR deficiency)	Inguinoscrotal	Zhang et al. 2001
Tfm (androgen receptor deficiency)	Inguinoscrotal	Hutson et al. 1986
Pygmy insertional mutants[HMGI protein(s) insertional inactivation]	Inguinoscrotal	Zhou et al.1995
Desrt -/- (ARID class transcription factor deficiency)	Inguinoscrotal	Lahoud et al. 2001
TS-rat (CGRP receptor down-regulation)	Inguinoscrotal	Ikadai et al.1988

INSL3 and *RXFP2* are among the strongest candidate genes for human cryptorchidism, since INSL3/RXFP2 hormone-receptor system has been demonstrated to play a pivotal role also in human testis descent. Nevertheless, none of the known sequence variants in these two genes turned to be useful for diagnostic purposes so far.

In this regard, it is worth noting that in contrast to mouse models in which cryptorchidism is exclusively associated to homozygous genotype, all *INSL3* and *RXFP2* genetic variants detected in humans so far, are heterozygous and were also found in subjects with normal testis descent. This is also the case of the T222P variant, the so far most extensively studied and at the same time, the most controversial genetic variant of *RXFP2*. A clear-cut cause-effect relation between T222P and non-syndromic cryptorchidism was originally proposed based on *in vitro* studies, suggesting a potential functional effect, and on the reported presence of this variant, always in heterozygosity, exclusively in men with history of testicular maldescent (Bogatcheva et al. 2007). However, our subsequent multicenter study called into question these early conclusions reporting this variant also in non cryptorchid men (Nuti et al 2008). In this thesis we add further evidence, based on two Mediterranean populations, against the consideration of T222P as a pathogenic mutation causing cryptorchidism (Ars et al. 2011). Indeed, in both populations studied, Spanish and Italian, the genetic variant resulted to be present in both patient and control groups, and in the latter with more than 1% prevalence. Gene haplotyping based on 5 exonic and intronic polymorphism provided evidence for a founder effect, and thus would support previous observations about the geographic restriction of this polymorphism to the Mediterranean area (Ferlin et al. 2003b; El Houate et al. 2008; Nuti et al. 2008). In the Spanish population no statistical difference was found in T222P frequency between cases and controls (*p* value=0.8; OR=0.89; 95% CI=0.2-3.4) indicating a lack of pathogenic effect in this population. Although in Italy, it seems to confer a mild risk to cryptorchidism (*p* value=0.031; OR=3.17; 95% CI=1.07-9.34), its screening for diagnostic purposes is not advised because of the relatively high frequency of control carriers.

Given the key role of INSL3/RXFP2 system in controlling the transabdominal phase of testis descent, genetic variants in the respective genes remain the strongest candidate genetic factors in respect to monogenic causes of non-syndromic cryptorchidism. For instance, to date only heterozygous T222P carriers have been identified among over 1900 cryptorchid patients screened, thus remaining the question whether the T222P may exert a pathogenic effect in homozygosity or in compound heterozygosity with another *RXFP2* pathogenic mutation. An interesting missense substitution in *RXFP2*, the D294G, has been recently discovered in a mouse model (Harris et al. 2010). Data on other SNPs in the human *RXFP2* is relatively scarce due to the fact that the full analysis of the entire coding sequence has been performed only in about 160 patients so far (Gorlov et al. 2002).
Another newly proposed candidate gene discovered in an animal model is the *PTGDS* gene (Philibert et al. 2013). Interestingly, unilateral cryptorchidism phenotype was observed in 16% and 24% of heterozygous (*Ptgds*$^{+/-}$) and homozygous (*Ptgds*$^{-/-}$) mice deficient for lipocalin-type prostaglandin D_2 (PGD_2) synthase, respectively. This gene is particularly promising since it is the first example of a heterozygous disruption associated with cryptorchid phenotype in mouse. No clinically relevant mutations have been found so far in cryptorchid patients, notwithstanding it is still premature to draw conclusions about the role of this gene in testis descent, since only 29 cryptorchid patients have been screened so far.

As stated above, mutation screening in the major candidate genes have been performed only in a limited number of patients, therefore the only conclusion we can draw at this point is that mutations in these genes are not a frequent cause of non-syndromic cryptorchidism. A more

complete picture of the genes involved in the disease will be obtained in the future with high throughput technologies. Especially next generation sequencing (NGS) makes now feasible either whole exome sequencing or the simultaneous analysis of panels of genes in a cost efficient way. NGS represent a powerful tool for the exploration of the complex genetics of non-syndromic cryptorchidism and will help to identify new genes involved in its etiopathogenesis.

A part from monogenic causes, a proportion of cryoptorchid cases may derive from the action of environmental factors on a predisposing genetic background. This hypothesis is supported by the identification of chemicals with endocrine disrupting properties referred to as Environmental Endocrine Disruptors (EEDs). Multiple evidence based on animal models suggests that such substances may have antiandrogenic or estrogenic activity and may therefore cause disorders of genital development during fetal life (Olesen et al. 2007). For instance, some of these chemicals have been reported to directly inhibit *Insl3* expression in rodents, suggesting a direct link between exposure to estrogenic compounds and cryptorchidism (Emmen et al. 2000; Nef et al. 2000). These observations led Skakkebaek to the formulation of the TDS theory which postulates a common etiology for cryptorchidism, hypospadias, testicular cancer, and spermatogenic impairment. According to this theory, these pathological conditions may reflect an underlying gonadal disgenesis resulting from the combined action of genetic and environmental factors, including endocrine disrupting chemicals. The so called "estrogen hypothesis" explained the increase in the prevalence of cryptorchidism (and other signs of TDS), observed during the last few decades in some populations, as related to an increased environmental estrogen exposure of developing gonads combined with a genetic predisposition (Sharpe and Skakkebaek 1993).
Genetic susceptibility to the deleterious effects of EEDs is supposed to be mostly determined by variants in genes coding for steroid receptors (*AR* and *ESR1* thought to be involved in the interaction with EEDs) and in genes coding for enzymes involved in detoxification and metabolism of certain EEDs, as cytochrome P450, family 1, subfamily A, polypeptide 1 (*CYP1A1*) and hydroxysteroid (17-beta) dehydrogenase 4 (*HSD17B4).* Given the intra-uterine origin of TDS, it has been proposed that not only the individual genetic background but also the maternal one may modulate the susceptibility to the deleterious effect of EEDs on testis descent. Increasing evidence exists on the potential involvement of maternal exposure and genetic makeup. For instance, a significant association between persistent levels of pesticides with endocrine disrupting properties in breast milk and cryptorchidism has been reported in Danish and Finnish populations (Damgaard et al. 2006). Furthermore polymorphisms in detoxifying enzymes such as *CYP1A1* gene in the maternal genome have been reported to be associated with hypospadias in Japanase population and their involvement in cryptorchidism is also plausible in the light of the TDS theory (Kurahashi et al. 2005).
The majority of the studies exploring genetic susceptibility to the effect of EEDs have focused on candidate polymorphisms of steroid receptor genes in the affected subject. In this context, studies on the role of *AR* variants, as genetic modifiers of EEDs effects, have been performed principally on two polymorphic sites, the CAG and GGN repeats, the role of which in the etiology of non-syndromic cryptorchidism is still debated. Concerning the *ESR1* polymorphisms, the so called AGATA haplotype and its tag SNP, SNP12, have been investigated in relation to cryptorchidism in different ethnic groups (Yoshida et al. 2005; Galan et al. 2007; Wang et al. 2008; Tang et al. 2011) with contradictory results. SNP12 has been reported to be a significant risk factor for cryptorchidim in Japanese and Chinese populations and it was associated with more severe cryptorchid phenotype in a multiracial American population, whereas it has been described as a protective factor in a study

population from Italy. Data on this SNP in the Spanish population is restricted to severe spermatogenic disturbances and no association was found (Galan et al. 2007).

On the other hand, several other *ESR1* polymorphisms have been extensively analyzed in relation to male infertility/spermatogenic impairment or hypospadia (Kukuvitis et al. 2002; Guarducci et al. 2006; Galan et al. 2008). Among them the (TA)n VNTR within the promoter region of *ESR1* has been considered as the most promising because its position in the promoter implies a potential regulatory function on gene transcriptional activity. Furthermore the "high number of TA repeats" genotype (the so called genotype A) has been significantly associated with two (potentially) estrogen-dependent processes: post menopausal bone matrix mineralization and spermatogenesis. More specifically, genotype A has been found to significantly correlate with higher bone mineral density in post menopausal women and with less spermatogenic efficiency in normozoospermic men. In the light of these observations and given the relation between cryptorchidism and low sperm count in the context of TDS, we explored for the first time the role of (TA)n polymorphism, and in particular of genotype A, in the etiology of cryptorchidism (Lo Giacco et al. 2011). Two independent Mediterranean study populations (Italian and Spanish) were analyzed but no association between specific (TA)n genotypes and non-syndromic cryptorchidism was found neither in the Spanish (*p* value=0.91; OR=0.984 95% CI=0.729-1,329) nor in the Italian (*p* value=0.41; OR=1.146; 95% CI=0.828-1.585) populations. These findings therefore excluded an association between the TA repeat genotype A and cryptorchidism but further investigations are needed to establish whether under specific environmental conditions, such as exposure to high doses of EEDs, this polymorphism may contribute to the disease. In fact, future combined studies in which both the level of EED exposure and *ESR1* genotype will be analyzed, are likely to provide important insights into the role of gene-environmental interaction in this disease. Moreover, given the intra-uterine origin of the disease, future studies should also investigate both the affected individual and his mother's genetic makeup.
In conclusion, despite the efforts, no clear-cut causing factor for non-syndromic cryptorchidism has been identified so far and, thus, no genetic test is still available for diagnostic purposes. The availability of a genetic test for cryptorchidism would be useful in a clinical setting because it may guide clinicians in taking decisions about treatment and thus allow optimizing the management of patients. Surgical (orchidopexy) and hormonal therapy are the two options for prevention of the deleterious effects of cryptorchidism on testicular function in adulthood. In particular, surgical correction is generally recommended within the first year of life even though the possibility of a spontaneous testis descent brings sometimes to delay the intervention at one year and a half. Ideally, next generation sequencing will provide a more complete picture of the genes involved in cryptorchidism and we expect that genetic variants with clear genotype-phenotype correlation will also be identified. In this way genetic screening may become a useful tool to optimize surgical timing, by allowing to discern between patients candidate for an early surgical correction of cryptorchidism and individuals with higher chance of spontaneous testis descent, who would not undergo to unnecessary surgery during the first months of life. Finally, considering the potential involvement of the endocrine disruptors in the etiology of TDS, the identification of a genetic predisposition to the deleterious effect of a given environmental factor, would allow to use a preventive strategy based on the control of lifestyle risk factors such as mother/son exposure.

6.2 Sex chromosome linked CNVs

Observational and experimental evidence suggests that an increased genomic instability may be the underlying factor of some cases of male infertility. Two epidemiological studies,

reported increased incidence of morbidity (including cancer, urogenital disorders and an overall reduced general health) and decreased life expectancy, among men with spermatogenic disturbance (Jensen et al. 2009; Salonia et al. 2009), suggesting that spermatogenic defects might be just one aspect of a systemic problem, possibly associated with a less stable genome. The increased sperm aneuploidy rate (Martin 2008; Zhou et al. 2011; Harton and Tempest 2012), the higher frequency of chromosomal anomalies compared to the general population and the higher level of sperm DNA fragmentation reported in a proportion of infertile men (Schulte et al. 2010), can be interpreted as an expression of increased genomic instability. Furthermore, growing evidence exists supporting the involvement of CNVs, duplications or deletions of genomic DNA, in spermatogenic defects (Jorgez et al. 2011; Tüttelmann et al. 2011; Krausz et al. 2012; Stouffs et al. 2012; Lopes et al. 2013).

The Y chromosome displays an inherent genetic instability related to the accumulation of a high proportion of segmental duplications which provide the structural basis for the generation of CNVs (Skaletsky et al. 2003; Jobling M.A 2008). In particular, the most distal *AZF* region of the Y chromosome, the *AZFc* region, with its structural intricacy only paralleled by the major histocompatibility complex, represents a major generator of large-scale variation in the human genome.

The Y chromosome is enriched in genes with exclusive expression in the testis and thus it is not surprising that the most clinically relevant CNVs in male infertility are Y chromosome-linked.

Among them, *AZF* deletions are the most frequent known genetic cause of severe spermatogenic failure and their molecular screening is one of the few genetic tests available in the diagnostic work-up of male infertility. Furthermore, several recurrent partial deletions (named *gr/gr*, *b2/b3* and *b1/b3*) and duplications (*b2/b4* duplication) have been reported in the *AZFc* region, in addition to the classical *AZFc* deletion. Among them only the *gr/gr* deletion resulted to be clinically relevant. The clinical significance of this recurrent deletion has been object of a long-lasting debate. Notwithstanding, the 4 meta-analysis published so far on this topic indicate that *gr/gr* deletion represents a significant risk factor for impaired sperm production with an average 2 fold increased risk for oligozoospermia (Tüttelmann et al. 2007; Visser et al. 2009; Navarro-Costa et al. 2010; Stouffs et al. 2011).

We addressed the topic of Y-linked CNVs in male infertility presenting the 8-year experience of Fundació Puigvert clinic in testing infertile men for Y microdeletions and partial *AZFc* rearrangements (Lo Giacco *et al* 2013 *EJHG,* under review). A detailed description of the genetic makeup, such as karyotype and classical *AZF* deletions, is provided for 806 mainly Spanish consecutive infertile men referring to the genetic laboratory of the clinic.

Concerning *AZF* classical deletions, our study adds further data to the existing literature (Oliva et al. 1998; Martínez et al. 2000) on the prevalence of Y microdeletions in a fertility clinic in Spain, and provides interesting insights on a few still debated issues concerning the routine Y microdeletions screening. One of them is represented by the sperm concentration cut-off value for the routine testing which is currently considered as <5 millions/ml. In our survey only 1 subject out of 21 carrying "genuine" *AZF* deletions, had a sperm concentration between 1-2x10^6/ml which is in accordance with the literature showing only 2.0% of all the *AZF* deletion carriers presenting a sperm concentration of more than 2x10^6/ml. This finding suggests that Y microdeletion screening could be eventually restricted to infertile men with sperm concentration ≤2x10^6/ml. Another debated issue concerns the genotype/phenotype correlation of *AZF* deletions, especially in terms of their predictive value for sperm retrieval at TESE. *AZFa* deletion is a rare event, representing less than 5% of the reported deletions, thus a small number of cases have been reported so far in the literature and all invariably

associated with a SCO histology. Notwithstanding, the largest review published so far on this topic, reports the presence of spermatids in the testes in 2/26 *AZFa* deleted men (Kleiman et al. 2012). The single *AZFa* deleted patient reported in the present thesis showed a complete bilateral SCOS, further supporting that the chance of finding sperm in testis is virtually zero. Thus, TESE should not be advised in these patients. In our survey, azoospermic *AZFc* deleted men showed a 9.1% sperm retrieval rate. This value lies below the lower limit of the range of sperm recovery rates reported in the literature for *AZFc* deleted patients (14.3%-80.0%) (Maurer et al. 2001; Peterlin et al. 2002; Hopps et al. 2003; Choi et al. 2004; Ferlin et al. 2007; Simoni et al. 2008; Stahl et al. 2010). The high proportion of pure SCOS cases among our *AZFc* deletion carriers represents a possible explanation for such a low retrieval rate. However, this is most likely related to technical issues such as low amount of testicular sample retrieved (single biopsy from each testis) and the procedure used (classical TESE); indeed, laboratories which performed microTESE with multiple sampling reported higher sperm recovery rate. Therefore, microTESE which allows better outcomes (60%-80% recovery in the literature) should be regarded as the best option for sperm retrieval in *AZF* deletion carriers.
There is evidence both in favor and against a hypothetical progressive deterioration of semen quality over time in *AZFc* deleted men. Accordingly, the relevance of sperm cryopreservation as an option to counteract this hypothetical phenomenon and hence to preserve the already scarce reproductive potential of these patients is also debated. In this regard, our experience is based on a single *AZFc* deletion carrier with severe oligozoospermia that developed into criptozoospermia. This finding indicates that a progressive decline in spermatogenic activity in *AZFc* deleted patients may occur. However, further longitudinal studies are needed to distinguish between physiological oscillations and real impairment of sperm parameters over time in order to provide a clear indication for preventive cryopreservation.

While is well known that Y microdeletion frequency varies according to the semen phenotype; it is unclear whether Y background might predispose to the formation of deletions and contribute to the observed "inter-population" variation in the deletion frequency even within similar semen categories. In this thesis, we report a number of observations supporting a possible Y-background effect on deletion frequencies (Lo Giacco et al. 2013).
The prevalence of Y microdeletions in our study (3.3%) is in line with the overall data presented in the literature (3.5%, according to a recent meta-analysis) (Hofherr et al. 2011). The higher deletion frequency observed in the Spanish compared to the German population (the lowest ever frequency in the literature) (Simoni et al. 2008) and the almost identical figures observed between Spanish and Italian (displaying a more similar Y background) support a possible Y-background effect on deletion frequencies. The significantly higher deletion frequency observed in non-Spanish compared to Spanish idiopathic infertile men is plausibly due to the different phenotypic composition of the two cohorts, in terms of a higher proportion of azoospermic men (44.4% in the non-Spanish versus 29.3% in the Spanish group). In fact, azoospermic men are more frequently affected by *AZF* deletion than oligozoospermic men. Nevertheless, Y background might also represent a contributory factor influencing deletion frequencies. Y hgr analysis showed that Spanish *AZF* deletion carriers all belonged to the P,Q,R lineages. In the non-Spanish cohort, consistently with the high proportion of North African patients (53.2%) included, *AZF* deletions were mostly found on hgr E, which recent surveys suggested as more prone to Y microdeletions (Imken et al. 2007; Eloualid et al. 2012). Moreover, the deletion frequency reported in idiopathic azoospermic and severe oligozoospermic men (9.09% and 5.5%) in Moroccan population (Eloualid et al. 2012) is consistent with our findings in non-Spanish idiopathic azoospermic and oligozoospermic men (10.7% and 6.9%, respectively).

As for partial *AZFc* rearrangements, our aims were to i) corroborate the clinical relevance of *gr/gr* deletion in the Spanish population; ii) study the relationship of Y hgrs with the phenotypic expression of *gr/gr* deletions; iii) provide information about the role of *b2/b3* rare deletions and partial duplications in spermatogenesis.
A detailed molecular characterization of the *AZFc* region was performed in a carefully matched case/control study setting. Spanish *gr/gr* deletion carriers displayed an increased probability (OR=2.8; 95% CI=1.017-8.007; P<0.05) of impaired spermatogenesis compared to non-carriers. Overall, this data together with a previous pilot study (de Llanos et al. 2005) further confirms the *gr/gr* deletion as a significant risk factor in the Spanish population (OR=4.8; 95% CI=1.863-12.623; P<0.001), providing additional support of its clinical relevance in Caucasians, consistently with the meta-analyses published so far (Tüttelmann et al. 2007; Visser et al. 2009; Navarro-Costa et al. 2010; Stouffs et al. 2011). The clinical implication of this finding in the Spanish population reinforces the idea that the *gr/gr* deletion screening should gain more consideration when dealing with infertile couples. This issue is of particular importance considering that, in some populations, partial deletions were shown to favour the occurrence of complete deletions (Zhang et al. 2007; Lu et al. 2009).
The majority of *gr/gr* deletion carriers belonged to the P,Q,R branches in both patients and controls, supporting that the phenotypic variability of *gr/gr* deletion is independent of Y-chromosomal background in Europeans (Krausz et al. 2009).
Conversely, Y background seems to modulate a potential deleterious effect of *b2/b3* on spermatogenesis (this deletion has been found exclusively in patients), as suggested by the clustering of *b2/b3* deletion carriers in C,F,G,H,I Y hgrs which are frequent in Moroccan population (even though these lineages are quite represented also in the Spanish population), for which *b2/b3* deletion has been recently reported in association with male infertility (Imken et al. 2007; Eloualid et al. 2012).
In this thesis we performed the first screening for partial *AZFc* duplication in the Spanish population. The findings reported, consistently with those regarding the Italian population (Giachini et al. 2008), confirmed that both partial and complete *AZFc* duplications do not represent any risk for spermatogenic failure in the Caucasian population. The results of *DAZ* gene dosage, performed independently from the type of partial rearrangements (*gr/gr*, *b2/b3,* partial or complete *AZFc* duplications), mirror the findings from the case-control study, showing that a reduced *DAZ* copy number (2 copies) is associated with a significant reduction in semen quality in terms of sperm concentration and total motile sperm count, whereas an increase (6 or 8 copies) in respect to the most common *DAZ* gene dosage (4 copies) does not affect significantly semen quality.

Overall, the literature and our study clearly show that men affected by spermatogenic disturbances have a higher frequency of Y chromosome-linked deletions. However, by performing an X-chromosome specific high resolution a-CGH, we have also demonstrated that the X chromosome of infertile men also displays an increased frequency of deletions (Krausz et al. 2012). In this study, a total of 96 idiopathic infertile subjects with different grade of spermatogenic impairment and 103 normozoospermic controls were analyzed and 73 CNVs (29 losses and 44 gains) were identified, the majority of which (75.3%) were novel. The analysis of the CNV distribution in patient and control groups, suggested that gains were less likely to affect spermatogenesis since 63% of them were found also in the normozoospermic group. On the contrary, deletions were less frequent in controls (only 38% were present also in normozoospermic group) indicating that abnormal sperm phenotype is more likely to be associated with deletions. Interestingly, the mean amount of DNA loss was found to be significantly higher in patients than controls and accordingly, the significantly higher number of CNVs/person observed among patients, was dependent on the over-representation of

losses in this group. Furthermore, a significantly lower sperm concentration and total sperm count was found in patient with >1 X-linked CNV with respect to those with ≤1 X-linked CNV.
Overall these findings support the hypothesis that infertile patients are more prone to the loss of genetic material and that such deletion burden is potentially linked to spermatogenic failure. These conclusions are supported by two other studies evaluating the involvement of CNVs in male infertility on a genome-wide scale (Tüttelmann et al. 2011; Lopes et al. 2013). More specifically, Tuttelmann and colleagues (2011) reported a significant over-representation of sex-chromosomal CNVs in azoospermic men with SCO histology and a significant negative correlation between sperm count and number of deletions at whole genome level, among normozoospermic men. More recently, Lopes ad colleagues (2013) reported a genome-wide CNV burden in azoospermic and oligozoospermic men replicating our finding of an X-linked CNV burden in men with spermatogenic failure (Krausz et al. 2012).

Another interesting finding of the X-chormosome a-CGH study included in the present thesis, concerns the identification of a number of patient-enriched and patient-specific deletions. The potential clinical implication of the latter was further evaluated through the screening of a large group of patients and controls which revealed that >90% of these deletions were private or rare (frequency <1%). Again, these findings are in line with the previous whole genome a-CGH study by Tuttelmann et al, in which among 27 patient-specific CNVs only one recurrent CNV (a duplication) was found. Similarly in the paper by Stouff and colleagues (2012) among 10 patient-specific autosomal CNVs only two were recurrent. Furthermore, enrichment in large rare CNVs was mostly drove the CNV burden reported by Lopes and colleagues (2013) in men with spermatogenic failure. The role of rare CNVs has already been established for other multifactorial diseases (Manolio et al. 2009; Pinto et al. 2010) and since mutations causing spermatogenic disturbance are unlikely transmitted to the offspring, it is plausible that *de novo* rare mutations play a prominent role in primary testicular failure.
Through bioinformatic analysis of functional genomic regions mapping inside or nearby the above mentioned patient-specific deletions, we found that X-linked CTA genes were recurrently affected, suggesting that CTA gene dosage variation may play a role in the CNV-related spermatogenic failure (Krausz et al. 2012; Lopes et al. 2013). Human CTA gene families are largely unexplored and no clinical data are available. Nevertheless their germ-cell restricted expression suggests a potential involvement in spermatogenesis. X-linked CTA genes are the most interesting in this context. Clues regarding functionality for many of these CTAs point to a role in cell cycle regulation or transcriptional control. Furthermore, these genes, mostly multi-copy, have been demonstrated to be under a strong diversifying pressure and are amongst the fastest-evolving genes in the human genome (Stevenson et al. 2007). This observation has been supported by a recent survey which based on a more complete and accurate human X chromosome reference sequence provided a reliable comparison of human and mouse X chromosomes (Mueller et al. 2013). This analysis revealed that most human X-linked ampliconic and multi-copy genes does not display orthologs in mouse having been acquired independently by the human X chromosome since the human and mouse lineages began to diverge from a common ancestor. Interestingly, these ampliconic genes resulted to be predominantly expressed in testicular germ cells suggesting a specialization of these genes for sperm production. Another clue supporting the potential importance of X-linked multi-copy gene in spermatogenesis in contrast to single-copy genes derives from a study revealing that multi-copy genes show a post-meiotic expression, thus suggesting that the multi-copy status may be a strategy to escape the MSCI.

As stated above, the majority of the 29 X-linked losses detected were "rare". However, 3 recurrent deletions (frequency >1%) caught our attention for their exclusive (CNV67) or prevalent (CNV64 and CNV69) presence in patients. The analysis of these deletions was object of the last part of the present thesis (Lo Giacco et al. 2013 submitted to J. Med. Genet.). An in-depth analysis was performed including: i) the screening of a large series of strictly idiopathic patients (n=627) and normozoospermic controls (n=628) from two Mediterranean populations (Spanish and Italian); ii) the molecular characterization of deletions; iii) the exploration for functional elements in the region of interest and iv) genotype-phenotype correlation analysis. All three deletions are mapping to the long arm of the X chromosome in q27.3 (CNV64) and q28 (CNV67 and CNV69). The smallest deletion was CNV64 removing at least 3.923 Kb, followed by CNV67, the minimum size of which was estimated to be 11.664 Kb, and CNV69 removing between 16.06-18.53 Kb. The alignment of the flanking sequences indicates that none of deletions originates from homologous recombination, thus, the most likely mechanism is NHEJ.

CNV64 and CNV69 both resulted significantly associated with impaired spermatogenesis as supported by the increased probability of spermatogenic failure reported in carriers (OR=1.9 and 2.2, for CNV64 and CNV69, respectively) and the lower sperm concentration /total motile sperm number observed in carriers compared to non-carriers. Interestingly, the molecular characterization of CNV69 deletion, revealed that at least two subtypes of CNV69 exist, named type A and type B. Only type B deletion was significantly more represented in patients than controls, thus suggesting that this deletion pattern may account for the potential deleterious effect of CNV69 on sperm production. This may be related to the closer position of type B deletion to an upstream insulator compared to type A. Importantly, through the breakpoint definition, a simple PCR-based assay has been developed for type B deletion which will allow its screening to be easily performed in other independent series of cases and controls. Despite the observed significant association between the above deletions and the infertile phenotype, no genes have been identified inside the maximum size of the deletions. However, since according to the ENCODE project, a biochemical function can be assigned to a very high percentage of non-coding DNA (80%), we explored the regions around each CNV for the presence of functional elements. A number of regulatory elements, including weak and strong enhancers, insulators and weak promoters have been found to be potentially affected because of their proximity to the deletion. It is also important to note, that large deletions may affect gene transcription by changing 3D structure of chromatin leading to downstream effects on the regulation of protein coding regions.

The most interesting finding of our study is related to the CNV67 deletion which was exclusively found in patients at a frequency of 1,1% (p<0.01).
Although the highly repetitive nature of the genomic region involved, and the incomplete assembly of the currently available reference sequence of the human X chromosome, prevented the fine mapping of deletion breakpoints, chromosome walking allowed to better define the deletion and to hypothesize the involvement of the proximal copy of the CTA gene *MAGEA9* and/or of its regulatory elements. Furthermore, considering that large-scale CNVs may also affect gene activity through a positional effect, also the distal copy of *MAGEA9* (Melanoma antigen family A, 9) may be affected by CNV67. In the light of the evidence suggesting a potential involvement of CTA genes in spermatogenesis and of CTA gene dosage variation in spermatogenic impairment, it is plausible that the loss of one *MAGEA9* copy would affect spermatogenesis explaining the phenotype observed in CNV67 carriers.

Since CNV67 may also affect regulatory elements of another independently acquired X-linked multi-copy gene, *HSFX1/2* with testis-specific expression, the alteration of the expression of this gene may also account for the pathological semen phenotype observed in CNV67 carriers. Pedigree analyses of two CNV67 carriers indicated that CNV67 deletion is maternally inherited, thus not affecting female fertility in heterozygous state. This is in accordance with the lack of expression of both the *MAGEA9* and *HSFX1/2* in the ovary. Patient 11-041's family is especially informative since the pathological semen phenotype of the carrier (11-041) versus his normozoospermic non-carrier brother is a strong indicator for a pathogenic effect of the deletion on spermatogenesis.

In conclusion for the first time, we provide evidence for a significant association between recurrent X-linked deletions and impaired sperm production. Further studies in other ethnic-geographic groups are needed to confirm the association of CNV64 and CNV69 type B with spermatogenic failure. Due to the specific association of CNV67 with spermatogenic impairment a parallelism with *AZF* deletions of the Y chromosome is tempting. This CNV merits further investigation in order to disentangle the structural complexity of this region, thus providing a feasible substrate for fine molecular characterization, and to further evaluate its putative diagnostic value.

7. CONCLUSIONS

7.1 Genetics of non-syndromic cryptrochidism

1. In the Spanish population the heterozygous T222P variant of the *RXFP2* gene can be regarded to as a polymorphism with no pathogenic effect, whereas in the Italian population it seems to confer a mild risk to cryptorchidism. Data on both Mediterranean populations add further evidence against the previous description of T222P as a pathogenic mutation causing cryptorchidism.
2. A founder effect of the T222P variant seems to be the most plausible explanation for its polymorphic frequency in restricted geographic areas such as Italy and Spain.
3. The $(TA)_n$ polymorphism of the *ESR1* promoter is not associated with non-syndromic cryptrochidism neither in the Spanish nor in the Italian population.

7.2 Sex-linked CNVs in spermatogenic impairment

1. The indication for routine Y-chromosome microdeletion screening may be limited to subjects with sperm concentration ≤2 $x10^6$/ml.
2. In azoospermic *AZFc* deletion carriers, classical TESE is associated with a low sperm recovery rate (9.1%), therefore microTESE should be regarded as the best option for sperm retrieval in these patients.
3. Y chromosome background could partially account for the differences observed for Y microdeletion frequency among different populations.
4. *gr/gr* deletion is a significant risk factor for impaired spermatogenesis in the Spanish population.
5. Both partial and complete *AZFc* duplications do not represent any risk for spermatogenic failure in the Spanish population.
6. A reduced *DAZ* gene copy number is associated with a significant decrease in sperm count and motility, whereas an increase of the most common *DAZ* gene dosage does not affect sperm parameters.
7. A significantly increased X-linked deletion burden observed in infertile men indicates its involvement in the etiology of spermatogenic impairment.
8. X-linked CNVs in infertile men mostly involve CTA family genes, indicating that their dosage variation could be related to spermatogenic failure.
9. Three deletions found in this study to be recurrent on the X chromosome (named CNV64, CNV67, CNV69) are significantly associated with spermatogenic failure. Interestingly, CNV67 is specific to oligo/azoospermic men with a frequency of 1,1%, resembling the *AZF* regions on the Y chromosome.

8. REFERENCES

Acerini CL, Miles HL, Dunger DB, Ong KK, Hughes IA. 2009. The descriptive epidemiology of congenital and acquired cryptorchidism in a UK infant cohort. *Arch Dis Child.* 94: 868–72.

Adham IM, Agoulnik AI. 2004. Insulin-like 3 signalling in testicular descent. *Int J Androl* 27: 257–65.

Adham IM, Burkhardt E, Benahmed M, Engel W. 1993. Cloning of a cDNA for a novel insulin-like peptide of the testicular Leydig cells. *J Biol Chem* 268: 26668–72.

Adham IM, Steding G, Thamm T, Büllesbach EE, Schwabe C, Paprotta I, Engel W. 2002. The overexpression of the insl3 in female mice causes descent of the ovaries. *Mol Endocrinol* 16: 244–52.

Aitman TJ, Dong R, Vyse TJ, Norsworthy PJ, Johnson MD, Smith J, Mangion J, Roberton-Lowe C, Marshall AJ, Petretto E, Hodges MD, Bhangal G, et al. 2006. Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. *Nature* 439: 851–5.

Akimoto C, Kitagawa H, Matsumoto T, Kato S. 2008. Spermatogenesis-specific association of SMCY and MSH5. *Genes Cells.* 13: 623–33.

Aksglaede L, Christiansen P, Sørensen K, Boas M, Linneberg A, Main KM, Andersson A-M, Skakkebaek NE, Juul A. 2011a. Serum concentrations of Anti-Müllerian Hormone (AMH) in 95 patients with Klinefelter syndrome with or without cryptorchidism. *Acta Paediatr.* 100: 839–45.

Aksglaede L, Juul a. 2013. Testicular function and fertility in men with Klinefelter syndrome: a review. *Eur J Endocrinol* 168: R67–76.

Aksglaede L, Skakkebaek NE, Almstrup K, Juul A. 2011b. Clinical and biological parameters in 166 boys, adolescents and adults with nonmosaic Klinefelter syndrome: a Copenhagen experience. *Acta paediatr.* 100: 793–806.

Aksglaede L, Skakkebaek NE, Juul A. 2008. Abnormal sex chromosome constitution and longitudinal growth: serum levels of insulin-like growth factor (IGF)-I, IGF binding protein-3, luteinizing hormone, and testosterone in 109 males with 47,XXY, 47,XYY, or sex-determining region of the Y chromosome . *J Clin Endocrinol Metab.* 93: 169–76.

Aksglaede L, Wikström AM, Rajpert-De Meyts E, Dunkel L, Skakkebaek NE, Juul A. 2006. Natural history of seminiferous tubule degeneration in Klinefelter syndrome. *Hum Reprod Update* 12: 39–48.

Almeida LG, Sakabe NJ, deOliveira AR, Silva MCC, Mundstein AS, Cohen T, Chen Y-T, Chua R, Gurung S, Gnjatic S, Jungbluth AA, Caballero OL, et al. 2009. CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens. *Nucleic Acids Res.* 37: D816–9.

Anand-Ivell R, Ivell R, Driscoll D, Manson J. 2008. Insulin-like factor 3 levels in amniotic fluid of human male fetuses. *Hum Reprod.* 23: 1180–6.

Anand-Ivell R, Wohlgemuth J, Haren MT, Hope PJ, Hatzinikolas G, Wittert G, Ivell R. 2006. Peripheral INSL3 concentrations decline with age in a large population of Australian men. *Int J Androl* 29: 618–26.

Andrés O, Kellermann T, López-Giráldez F, Rozas J, Domingo-Roura X, Bosch M. 2008. RPS4Y gene family evolution in primates. *BMC Evol Biol.* 8: 142.

Ars E, Giacco D Lo, Bassas L, Nuti F, Rajmil O, Ruíz P, Garat JM, Ruiz-Castané E, Krausz C. 2011. Further insights into the role of T222P variant of RXFP2 in non-syndromic cryptorchidism in two Mediterranean populations. *Int J Androl* 34: 333–8.

Aschim EL, Nordenskjöld A, Giwercman A, Lundin KB, Ruhayel Y, Haugen TB, Grotmol T, Giwercman YL. 2004. Linkage between cryptorchidism, hypospadias, and GGN repeat length in the androgen receptor gene. *J Clin Endocrinol Metab.* 89: 5105–9.

Aston KI, Carrell DT. Genome-wide study of single-nucleotide polymorphisms associated with azoospermia and severe oligozoospermia. *J Androl.* 30: 711–25.

Aston KI, Krausz C, Laface I, Ruiz-Castané E, Carrell DT. 2010. Evaluation of 172 candidate polymorphisms for association with oligozoospermia or azoospermia in a large cohort of men of European descent. *Hum Reprod.* 25: 1383–97.

Baarends WM, Laan R van der, Grootegoed JA. 2000. Specific aspects of the ubiquitin system in spermatogenesis. *J Endocrinol Invest.* 23: 597–604.

Bacolla A, Jaworski A, Larson JE, Jakupciak JP, Chuzhanova N, Abeysinghe SS, O'Connell CD, Cooper DN, Wells RD. 2004. Breakpoints of gross deletions coincide with non-B DNA conformations. *Proc Natl Acad Sci U S A.* 101: 14162–7.

Bassaganyas L, Riveira-Muñoz E, García-Aragonés M, González JR, Cáceres M, Armengol L, Estivill X. 2013. Worldwide population distribution of the common LCE3C-LCE3B deletion associated with psoriasis and other autoimmune disorders. *BMC genomics* 14: 261.

Bay K, Hartung S, Ivell R, Schumacher M, Jürgensen D, Jorgensen N, Holm M, Skakkebaek NE, Andersson A-M. 2005. Insulin-like factor 3 serum levels in 135 normal men and 85 men with testicular disorders: relationship to the luteinizing hormone-testosterone axis. *J Clin Endocrinol Metab.* 90: 3410–8.

Bay K, Main KM, Toppari J, Skakkebæk NE. 2011. Testicular descent: INSL3, testosterone, genes and the intrauterine milieu. *Nat Rev Urol.* 8: 187–96.

Bay K, Matthiesson KL, McLachlan RI, Andersson A-M. 2006. The effects of gonadotropin suppression and selective replacement on insulin-like factor 3 secretion in normal adult men. *J Clin Endocrinol Metab.* 91: 1108–11.

Bay K, Virtanen HE, Hartung S, Ivell R, Main KM, Skakkebaek NE, Andersson A-M, Toppari J. 2007. Insulin-like factor 3 levels in cord blood and serum from children: effects of age, postnatal hypothalamic-pituitary-gonadal axis activation, and cryptorchidism. *J Clin Endocrinol Metab.* 92: 4020–7.

Becherini L, Gennari L, Masi L, Mansani R, Massart F, Morelli A, Falchetti A, Gonnelli S, Fiorelli G, Tanini A, Brandi ML. 2000. Evidence of a linkage disequilibrium between polymorphisms in the human estrogen receptor alpha gene and their relationship to bone mass variation in postmenopausal Italian women. *Hum mol genet* 9: 2043–50.

Bellott DW, Skaletsky H, Pyntikova T, Mardis ER, Graves T, Kremitzki C, Brown LG, Rozen S, Warren WC, Wilson RK, Page DC. 2010. Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. *Nature* 466: 612–6.

Benko S, Gordon CT, Mallet D, Sreenivasan R, Thauvin-Robinet C, Brendehaug A, Thomas S, Bruland O, David M, Nicolino M, Labalme A, Sanlaville D, Callier P, Malan V, Huet F, Molven A, Dijoud F, Munnich A, Faivre L, Amiel J, Harley V, Houge G, Morel Y, Lyonnet S. 2011. Disruption of a long distance regulatory region upstream of SOX9 in isolated disorders of sex development. *J Med Genet*. 48:825-30

Bentvelsen FM, McPhaul MJ, Wilson JD, George FW. 1994. The androgen receptor of the urogenital tract of the fetal rat is regulated by androgen. *Mol Cell Endocrinol.* 105: 21–6.

Bergen AW, Pratt M, Mehlman PT, Goldman D. 1998. Evolution of RPS4Y. *Mol Biol Evol.* 15:1412–9.

Bogatcheva N V, Ferlin A, Feng S, Truong A, Gianesello L, Foresta C, Agoulnik AI. 2007. T222P mutation of the insulin-like 3 hormone receptor LGR8 is associated with testicular maldescent and hinders receptor expression on the cell surface membrane. *Am J Physiol Endocrinol Metab.* 292: E138–44.

Bogatcheva N V, Truong A, Feng S, Engel W, Adham IM, Agoulnik AI. 2003. GREAT/LGR8 is the only receptor for insulin-like 3 peptide. *Mol Endocrinol.* 17: 2639–46.

Bojesen A, Juul S, Gravholt CH. 2003. Prenatal and postnatal prevalence of Klinefelter syndrome: a national registry study. *J Clin Endocrinol Metab.* 88: 622–6.

Bosch E, Jobling MA. 2003. Duplications of the AZFa region of the human Y chromosome are mediated by homologous recombination between HERVs and are compatible with male fertility. *Hum Mol Genet.* 12: 341–7.

Brandell RA, Mielnik A, Liotta D, Ye Z, Veeck LL, Palermo GD, Schlegel PN. 1998. AZFb deletions predict the absence of spermatozoa with testicular sperm extraction: preliminary report of a prognostic genetic test. *Hum Reprod.* 13: 2812–5.

Brown GM, Leversha M, Hulten M, Ferguson-Smith MA, Affara NA, Furlong RA. 1998. Genetic analysis of meiotic recombination in humans by use of sperm typing: reduced recombination within a heterozygous paracentric inversion of chromosome 9q32-q34.3. *Am J Hum Genet.* 62:1484–92.

Bruder CEG, Piotrowski A, Gijsbers AACJ, Andersson R, Menzel U, Sandgren J, Tell D Von, Erickson S, Sta TD De, Poplawski A, Crowley M, Crasto C, et al. 2008. Phenotypically Concordant and Discordant Monozygotic Twins Display Different DNA Copy-Number-Variation Profiles. *Am J Hum Genet.* 82: 763–771.

Caron C, Pivot-Pajot C, Grunsven LA van, Col E, Lestrat C, Rousseaux S, Khochbin S. 2003. Cdyl: a new transcriptional co-repressor. *EMBO Rep.* 4: 877–82.

Castagnoli A, Maestri I, Bernardi F, Senno L Del. 1987. PvuII RFLP inside the human estrogen receptor gene. *Nucleic Acids Res.* 15: 866.

Castro-Nallar E, Bacallao K, Parada-Bustamante A, Lardone MC, López P V, Madariaga M, Valdevenito R, Piottante A, Ebensperger M, Castro A. 2010. Androgen receptor gene CAG and GGN repeat polymorphisms in Chilean men with primary severe spermatogenic failure. *J Androl.* 31: 552–9.

Cederroth CR, Schaad O, Descombes P, Chambon P, Vassalli J-D, Nef S. 2007. Estrogen receptor alpha is a major contributor to estrogen-mediated fetal testis dysgenesis and cryptorchidism. *Endocrinology* 148: 5507–19.

Chai NN, Salido EC, Yen PH. 1997. Multiple functional copies of the RBM gene family, a spermatogenesis candidate on the human Y chromosome. *Genomics* 45: 355–61.

Chai NN, Zhou H, Hernandez J, Najmabadi H, Bhasin S, Yen PH. 1998. Structure and organization of the RBMY genes on the human Y chromosome: transposition and amplification of an ancestral autosomal hnRNPG gene. *Genomics* 49: 283–9.

Charlton HM, Halpin DM, Iddon C, Rosie R, Levy G, McDowell IF, Megson A, Morris JF, Bramwell A, Speight A, Ward BJ, Broadhead J, Davey-Smith G, Fink G 1983 The effects of daily administration of single and multiple injections of gonadotropin-releasing hormone on pituitary and gonadal function in the hypogonadal (hpg) mouse. *Endocrinology* 113:535–44

Chevalier D, Cauffiez C, Bernard C, Lo-Guidice JM, Allorge D, Fazio F, Ferrari N, Libersa C, Lhermitte M, D'Halluin JC, Broly F. 2001. Characterization of new mutations in the coding sequence and 5'-untranslated region of the human prostacylcin synthase gene (CYP8A1). *Hum Genet.* 108: 148–55.

Choi JM, Chung P, Veeck L, Mielnik A, Palermo GD, Schlegel PN. 2004. AZF microdeletions of the Y chromosome and in vitro fertilization outcome. *Fertil Steril.* 81: 337–41.

Coleman RT, Taylor JE, Shine JJ, Frossard PM. 1988. Human estrogen receptor (ESR) gene locus: PssI dimorphism. *Nucleic Acids Res.* 16: 7208.

Conrad DF, Andrews TD, Carter NP, Hurles ME, Pritchard JK. 2006. A high-resolution survey of deletion polymorphism in the human genome. *Nat Genet.* 38: 75–81.

Cook EH, Scherer SW. 2008. Copy-number variations associated with neuropsychiatric conditions. *Nature* 455: 919–23.

Cortes D, Thorup JM, Visfeldt J. 2001. Cryptorchidism: aspects of fertility and neoplasms. A study including data of 1,335 consecutive boys who underwent testicular biopsy simultaneously with surgery for cryptorchidism. *Horm Res.* 55: 21–7.

Cram D, Lynch M, O'Bryan MK, Salvado C, McLachlan RI, Kretser DM de. 2004. Genetic screening of infertile men. *Reprod Fertil Dev.* 16: 573–80.

Cruger DG, Agerholm I, Byriel L, Fedder J, Bruun-Petersen G. 2003. Genetic analysis of males from intracytoplasmic sperm injection couples. *Clin Genet.*64: 198–203.

Damgaard IN, Skakkebæk NE, Toppari J, Virtanen HE, Shen H, Schramm K-W, Petersen JH, Jensen TK, Main KM. 2006. Persistent Pesticides in Human Breast Milk and Cryptorchidism. Environmental Health Perspectives. *Environ Health Perspect.* 114: 1133–1138.

Davis-Dao C, Koh CJ, Hardy BE, Chang A, Kim SS, Filippo R De, Hwang A, Pike MC, Carroll JD, Coetzee GA, Vandenberg D, Siegmund K, et al. 2012. Shorter androgen receptor CAG repeat lengths associated with cryptorchidism risk among Hispanic white boys. *J Clin Endocrinol Metab.* 97: E393–9.

Dear PH. 2009. Copy-number variation: the end of the human genome?*Trends Biotechnol.*27:448–54.

Denney RM, Koch H, Craig IW. 1999. Association between monoamine oxidase A activity in human male skin fibroblasts and genotype of the MAOA promoter-associated variable number tandem repeat. *Hum Genet.* 105: 542–51.

Ditton HJ, Zimmer J, Kamp C, Rajpert-De Meyts E, Vogt PH. 2004. The AZFa gene DBY (DDX3Y) is widely transcribed but the protein is limited to the male germ cells by translation control. *Hum Mol Genet.* 13: 2333–41.

Donaldson KM, Tong SY, Washburn T, Lubahn DB, Eddy EM, Hutson JM, Korach KS. 1996. Morphometric study of the gubernaculum in male estrogen receptor mutant mice. *J Androl.*17:91–5.

Dorus S, Gilbert SL, Forster ML, Barndt RJ, Lahn BT. 2003. The CDY-related gene family: coordinated evolution in copy number, expression profile and protein sequence. *Hum Mol Genet.* 12: 1643–50.

Dromard M, Bompard G, Glondu-Lassis M, Puech C, Chalbos D, Freiss G. 2007. The putative tumor suppressor gene PTPN13/PTPL1 induces apoptosis through insulin receptor substrate-1 dephosphorylation. *Cancer Res* 67*:* 6806–13.

Duan Z, Duan Y, Lamendola DE, Yusuf RZ, Naeem R, Penson RT, Seiden M V. 2003. Overexpression of MAGE/GAGE genes in paclitaxel/doxorubicin-resistant human cancer cell lines. *Clin Cancer Res.* 9: 2778–85.

Ehrmann I, Dalgliesh C, Tsaousi A, Paronetto MP, Heinrich B, Kist R, Cairns P, Li W, Mueller C, Jackson M, Peters H, Nayernia K, et al. 2008. Haploinsufficiency of the germ cell-specific nuclear RNA binding protein hnRNP G-T prevents functional spermatogenesis in the mouse. *Hum Mol Genet.* 17: 2803–18.

Elliott DJ. 2004. The role of potential splicing factors including RBMY, RBMX, hnRNPG-T and STAR proteins in spermatogenesis. *Int J Androl.* 27: 328–34.

Elliott DJ, Bourgeois CF, Klink A, Stévenin J, Cooke HJ. 2000. A mammalian germ cell-specific RNA-binding protein interacts with ubiquitously expressed proteins involved in splice site selection. *Proc Natl Acad Sci U S A.* 97: 5717–22.

Elliott DJ, Millar MR, Oghene K, Ross A, Kiesewetter F, Pryor J, McIntyre M, Hargreave TB, Saunders PT, Vogt PH, Chandley AC, Cooke H. 1997. Expression of RBM in the nuclei of human germ cells is dependent on a critical region of the Y chromosome long arm. *Proc Natl Acad Sci U S A.*94: 3848–53.

Eloualid A, Rhaissi H, Reguig A, Bounaceur S, Houate B El, Abidi O, Charif M, Louanjli N, Chadli E, Barakat A, Bashamboo A, McElreavey K, et al. 2012. Association of spermatogenic failure with the b2/b3 partial AZFc deletion. *PloS one* 7: e34902.

Emmen JM, McLuskey A, Adham IM, Engel W, Verhoef-Post M, Themmen AP, Grootegoed JA, Brinkmann AO. 2000. Involvement of insulin-like factor 3 (Insl3) in diethylstilbestrol-induced cryptorchidism. *Endocrinology* 141: 846–9.

Estop AM, Munné S, Cieply KM, Vandermark KK, Lamb AN, Fisch H. 1998. Meiotic products of a Klinefelter 47,XXY male as determined by sperm fluorescence in-situ hybridization analysis. *Hum Reprod.* 13: 124–7.

Fanciulli M, Norsworthy PJ, Petretto E, Dong R, Harper L, Kamesh L, Heward JM, Gough SCL, Smith A de, Blakemore AIF, Froguel P, Owen CJ, et al. 2007. FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. *Nat Genet*. 39: 721–3.

Fanciulli M, Petretto E, Aitman TJ. 2010. Gene copy number variation and common human disease. *Clin Genet.* 77: 201–13.

Feng S, Cortessis VK, Hwang A, Hardy B, Koh CJ, Bogatcheva N V, Agoulnik AI. 2004. Mutation analysis of INSL3 and GREAT/LGR8 genes in familial cryptorchidism. *Urology* 64: 1032–6.

Ferlin A, Arredi B, Speltra E, Cazzadore C, Selice R, Garolla A, Lenzi A, Foresta C. 2007. Molecular and clinical characterization of Y chromosome microdeletions in infertile men: a 10-year experience in Italy. *J Clin Endocrinol Metab.* 92: 762–70.

Ferlin A, Arredi B, Zuccarello D, Garolla A, Selice R, Foresta C. 2006a. Paracrine and endocrine roles of insulin-like factor 3. *J Endocrinol Invest.* 29: 657–64.

Ferlin A, Garolla A, Bettella A, Bartoloni L, Vinanzi C, Roverato A, Foresta C. 2005. Androgen receptor gene CAG and GGC repeat lengths in cryptorchidism. *Eur J Endocrinol.* 152: 419–25.

Ferlin A, Garolla A, Rigon F, Rasi Caldogno L, Lenzi A, Foresta C. 2006b. Changes in serum insulin-like factor 3 during normal male puberty. *J Clin Endocrinol Metab.* 91: 3426–31.

Ferlin A, Moro E, Rossi A, Dallapiccola B, Foresta C. 2003a. The human Y chromosome's azoospermia factor b (AZFb) region: sequence, structure, and deletion analysis in infertile men. *J Med Genet.* 40: 18–24.

Ferlin A, Simonato M, Bartoloni L, Rizzo G, Bettella A, Dottorini T, Dallapiccola B, Foresta C. 2003b. The INSL3-LGR8/GREAT ligand-receptor pair in human cryptorchidism. *J Clin Endocrinol Metab.* 88: 4273–9.

Ferlin A, Vinanzi C, Garolla A, Selice R, Zuccarello D, Cazzadore C, Foresta C. 2006c. Male infertility and androgen receptor gene mutations: clinical features and identification of seven novel mutations. *Clin Endocrinol.* 65: 606–10.

Ferlin A, Zuccarello D, Zuccarello B, Chirico MR, Zanon GF, Foresta C. 2008. Genetic alterations associated with cryptorchidism. *JAMA.* 300: 2271–6.

Fernandes S, Paracchini S, Meyer LH, Floridia G, Tyler-Smith C, Vogt PH. 2004. A large AZFc deletion removes DAZ3/DAZ4 and nearby genes from men in Y haplogroup N. *Am J Hum Genet.* 74: 180–7.

Fischle W, Franz H, Jacobs SA, Allis CD, Khorasanizadeh S. 2008. Specificity of the chromodomain Y chromosome family of chromodomains for lysine-methylated ARK(S/T) motifs. *J Biol Chem.* 283: 19626–35.

Foresta C, Bettella A, Vinanzi C, Dabrilli P, Meriggiola MC, Garolla A, Ferlin A. 2004. A novel circulating hormone of testis origin in humans. *J Clin Endocrinol Metab.* 89: 5952–8.

Foresta C, Ferlin A, Moro E. 2000. Deletion and expression analysis of AZFa genes on the human Y chromosome revealed a major role for DBY in male infertility. *Hum mol genet*. 9: 1161–9.

Foresta C, Galeazzi C, Bettella A, Marin P, Rossato M, Garolla A, Ferlin A. 1999. Analysis of meiosis in intratesticular germ cells from subjects affected by classic Klinefelter's syndrome. *J Clin Endocrinol Metab.*84: 3807–10.

Foresta C, Galeazzi C, Bettella A, Stella M, Scandellari C. 1998. High incidence of sperm sex chromosomes aneuploidies in two patients with Klinefelter's syndrome. *J Clin Endocrinol Metab.* 83: 203–5.

Foresta C, Zuccarello D, Garolla A, Ferlin A. 2008. Role of hormones, genes, and environment in human cryptorchidism. *Endocr Rev.*29: 560–80.

Forti G, Krausz C. 1998. Clinical review 100: Evaluation and treatment of the infertile couple. *J Clin Endocrinol Metab.* 83: 4177–88.

Fratta E, Coral S, Covre A, Parisi G, Colizzi F, Danielli R, Nicolay HJM, Sigalotti L, Maio M. 2011. The biology of cancer testis antigens: putative function, regulation and therapeutic potential. Molecular oncology. *Mol Oncol.* 5: 164–82.

Freeman JL, Perry GH, Feuk L, Redon R, McCarroll S a, Altshuler DM, Aburatani H, Jones KW, Tyler-Smith C, Hurles ME, Carter NP, Scherer SW, et al. 2006. Copy number variation: new insights in genome diversity. *Genome Res.*16: 949–61.

Fullerton G, Hamilton M, Maheshwari a. 2010. Should non-mosaic Klinefelter syndrome men be labelled as infertile in 2009? *Hum Reprod.*25: 588–97.

Galan JJ, Buch B, Cruz N, Segura A, Moron FJ, Bassas L, Martinez-Pineiro L, Real LM, Ruiz A. 2005. Multilocus analyses of estrogen-related genes reveal involvement of the ESR1 gene in male infertility and the polygenic nature of the pathology. *Fertil Steril.* 84: 910–8.

Galan JJ, Buch B, Pedrinaci S, Jimenez-Gamiz P, Gonzalez A, Serrano-Rios M, Salinas A, Rivero M del C, Real LM, Royo JL, Ruiz A. 2008. Identification of a 2244 base pair interstitial deletion within the human ESR1 gene in the Spanish population. *J Med Genet.* 45: 420–4.

Galan JJ, Guarducci E, Nuti F, Gonzalez A, Ruiz M, Ruiz A, Krausz C. 2007. Molecular analysis of estrogen receptor alpha gene AGATA haplotype and SNP12 in European populations: potential protective effect for cryptorchidism and lack of association with male infertility. *Hum Reprod.* 22: 444–9.

Gennari L, Merlotti D, Paola V De, Calabrò A, Becherini L, Martini G, Nuti R. 2005. Estrogen receptor gene polymorphisms and the genetics of osteoporosis: a HuGE review. *Am J Epidemiol.* 161: 307–20.

Giachini C, Laface I, Guarducci E, Balercia G, Forti G, Krausz C. 2008. Partial AZFc deletions and duplications: clinical correlates in the Italian population. *Hum Genet.* 124: 399–410.

Giwercman A, Rylander L, Lundberg Giwercman Y. 2007. Influence of endocrine disruptors on human male fertility. *Reprod Biomed Online.* 15: 633–42.

Glynn SA, Gammell P, Heenan M, O'Connor R, Liang Y, Keenan J, Clynes M. 2004. A new superinvasive in vitro phenotype induced by selection of human breast carcinoma cells with the chemotherapeutic drugs paclitaxel and doxorubicin. *Br J Cancer.* 91: 1800–7.

Gorlov IP, Kamat A, Bogatcheva N V, Jones E, Lamb DJ, Truong A, Bishop CE, McElreavey K, Agoulnik AI. 2002. Mutations of the GREAT gene cause cryptorchidism. *Hum Mol Genet.* 11: 2309–18.

Gu W, Zhang F, Lupski JR. 2008. Mechanisms for human genomic rearrangements. *PathoGenetics* 1: 4.

Guarducci E, Nuti F, Becherini L, Rotondi M, Balercia G, Forti G, Krausz C. 2006. Estrogen receptor alpha promoter polymorphism: stronger estrogen action is coupled with lower sperm count. *Hum Reprod.* 21: 994–1001.

Guerrier D, Tran D, Vanderwinden JM, Hideux S, Outryve L Van, Legeai L, Bouchard M, Vliet G Van, Laet MH De, Picard JY. 1989. The persistent Müllerian duct syndrome: a molecular approach. *J Clin Endocrinol Metab.* 68: 46–52.

Guttenbach M, Michelmann HW, Hinney B, Engel W, Schmid M. 1997. Segregation of sex chromosomes into sperm nuclei in a man with 47,XXY Klinefelter's karyotype: a FISH analysis. *Hum Genet.* 99: 474–7.

Habermann B, Mi HF, Edelmann A, Bohring C, Bäckert IT, Kiesewetter F, Aumüller G, Vogt PH. 1998. DAZ (Deleted in AZoospermia) genes encode proteins located in human late spermatids and in sperm tails. *Hum Reprod.*13: 363–9.

Hadziselimovic F. 2006. Early successful orchidopexy does not prevent from developing azoospermia. *Int Braz J Urol.* 32: 570–3.

Hammer MF ZS. 2002. The human Y chromosome haplogroup tree: Nomenclature and Phylogeography of Its Major Divisions. *Annu Rev Anthropol* 31: 303–321.

Harris RM, Finlayson C, Weiss J, Fisher L, Hurley L, Barrett T, Emge D, Bathgate R a D, Agoulnik AI, Jameson JL. 2010. A missense mutation in LRR8 of RXFP2 is associated with cryptorchidism. *Mamm Genome.* 21: 442–9.

Harton GL, Tempest HG. 2012. Chromosomal disorders and male infertility. *Asian J Androl.* 14: 32–9.

Hennebicq S, Pelletier R, Bergues U, Rousseaux S. 2001. Risk of trisomy 21 in offspring of patients with Klinefelter's syndrome. *Lancet* 357: 2104–5.

Herrinton LJ, Zhao W, Husson G. 2003. Management of cryptorchism and risk of testicular cancer. *Am J Epidemiol.* 157: 602–5.

Hsieh-Li HM, Witte DP, Weinstein M, Branford W, Li H, Small K, Potter SS 1995 Hoxa 11 structure, extensive antisense transcription, and function in male and female fertility. *Development* 121: 1373–1385

Hochstenbach R, Hackstein JH. 2000. The comparative genetics of human spermatogenesis: clues from flies and other model organisms. *Results Probl Cell Differ.* 28: 271–98.

Hofherr SE, Wiktor AE, Kipp BR, Dawson DB, Dyke DL Van. 2011. Clinical diagnostic testing for the cytogenetic and molecular causes of male infertility: the Mayo Clinic experience. *J Assist Reprod Genet.* 28: 1091–8.

Hopps C V, Mielnik A, Goldstein M, Palermo GD, Rosenwaks Z, Schlegel PN. 2003. Detection of sperm in men with Y chromosome microdeletions of the AZFa, AZFb and AZFc regions. *Hum Reprod.* 18: 1660–5.

Houate B El, Rouba H, Imken L, Sibai H, Chafik A, Boulouiz R, Chadli E, Hassar M, McElreavey K, Barakat A. 2008. No association between T222P/LGR8 mutation and cryptorchidism in the Moroccan population. *Horm Res.*70: 236–9.

Houate B El, Rouba H, Sibai H, Barakat A, Chafik A, Chadli EB, Imken L, Bogatcheva N V, Feng S, Agoulnik AI, McElreavey K. 2007. Novel mutations involving the INSL3 gene associated with cryptorchidism. *J Urol.* 177: 1947–51.

Hrabovszky Z, Hutson JM. 1999. Capsaicin restores gubernacular contractility in TS rats. *J Pediatr Surg.* 34: 1769–72.

Hsu SY, Nakabayashi K, Nishi S, Kumagai J, Kudo M, Sherwood OD, Hsueh AJW. 2002. Activation of orphan receptors by the hormone relaxin. *Science.* 295: 671–4.

Hu Z, Xia Y, Guo X, Dai J, Li HH, Hu H, Jiang Y, Lu F, Wu Y, Yang X, Yao B, Lu C, et al. 2012. A genome-wide association study in Chinese men identifies three risk loci for non-obstructive azoospermia. *Nat. genet.* 44: 183–6.

Huang WJ, Lin Y-W, Hsiao K-N, Eilber KS, Salido EC, Yen PH. 2008. Restricted expression of the human DAZ protein in premeiotic germ cells. *Hum Reprod.* 23: 1280–9.

Hurles ME, Dermitzakis ET, Tyler-Smith C. 2008. The functional impact of structural variation in humans. *Trends Genet.*24: 238–45.

Hurst LD. 2001. Evolutionary genomics. Sex and the X. *Nature* 411: 149–50.

Hutson JM 1986 Testicular feminization: a model for testicular descent in mice and men. *J Pediatr Surg* 21:195–8

Hutson JM, Hasthorpe S, Heyns CF. 1997. Anatomical and functional aspects of testicular descent and cryptorchidism. *Endocr Rev.* 18: 259–80.

Iafrate a J, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. 2004. Detection of large-scale variation in the human genome. *Nat. genet.* 36: 949–51.

Ikadai H, Ajisawa C, Taya K, Imamichi T 1988 Suprainguinal ectopic scrota of TS inbred rats. *J Reprod Fertil* 84:701–7

Imken L, Houate B El, Chafik A, Nahili H, Boulouiz R, Abidi O, Chadli E, Louanjli N, Elfath A, Hassar M, McElreavey K, Barakat A, et al. 2007. AZF microdeletions and partial deletions of AZFc region on the Y chromosome in Moroccan men. *Asian J Androl.* 9: 674–8.

Ivell R, Hartung S. 2003. The molecular basis of cryptorchidism. *Mol Hum Reprod.* 9: 175–81.

Jackson JA, Fink GR. 1981. Gene conversion between duplicated genetic elements in yeast. *Nature* 292: 306–11.

Jensen TK, Jacobsen R, Christensen K, Nielsen NC, Bostofte E. 2009. Good semen quality and life expectancy: a cohort study of 43,277 men. *Am J Epidemiol.* 170: 559–65.

Jobling M a, Tyler-Smith C. 2003. The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet.* 4: 598–612.

Jobling M a, Williams G a, Schiebel G a, Pandya G a, McElreavey G a, Salas G a, Rappold G a, Affara N a, Tyler-Smith C. 1998. A selective difference between human Y-chromosomal DNA haplotypes. *Curr Biol*.8: 1391–4.

Jobling M.A. 2008. Copy number variation on the human Y chromosome. *Cytogenet Genome Res* 123: 253–262.

Johnson MD. 1998. Genetic risks of intracytoplasmic sperm injection in the treatment of male infertility: recommendations for genetic counseling and screening. *Fertil Steril.* 70: 397–411.

Jorgez CJ, Weedin JW, Sahin A, Tannour-Louet M, Han S, Bournat JC, Mielnik A, Cheung SW, Nangia AK, Schlegel PN, Lipshultz LI, Lamb DJ. 2011. Aberrations in pseudoautosomal regions (PARs) found in infertile men with Y-chromosome microdeletions. *J Clin Endocrinol Metab.* 96: E674–9.

Kaleva M, Toppari J. 2003. Genetics and hormones in testicular descent. *Hormones* 2: 211–6.

Kamp C, Huellen K, Fernandes S, Sousa M, Schlegel PN, Mielnik A, Kleiman S, Yavetz H, Krause W, Küpker W, Johannisson R, Schulze W, et al. 2001. High deletion frequency of the complete AZFa sequence in men with Sertoli-cell-only syndrome. *Mol Hum Reprod.* 7: 987–94.

Karafet TM, Mendez FL, Meilerman MB, Underhill P a, Zegura SL, Hammer MF. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res.*18: 830–8.

Keymolen K, Staessen C, Verpoest W, Michiels A, Bonduelle M, Haentjens P, Vanderelst J, Liebaers I. 2009. A proposal for reproductive counselling in carriers of Robertsonian translocations: 10 years of experience with preimplantation genetic diagnosis. *Hum Reprod.* 24: 2365–71.

Kichine E, Rozé V, Cristofaro J Di, Taulier D, Navarro A, Streichemberger E, Decarpentrie F, Metzler-Guillemain C, Lévy N, Chiaroni J, Paquis-Flucklinger V, Fellmann F, et al. 2012. HSFY genes and the P4 palindrome in the AZFb interval of the human Y chromosome are not required for spermatocyte maturation *Hum Reprod.* 27: 615–24.

Kim B, Lee Y, Kim Y, Lee KH, Chun S, Rhee K, Seo JT, Kim SW, Paick J-S. 2009. Polymorphic expression of DAZ proteins in the human testis. *Hum Reprod.* 24: 1507–15.

Kim J, Daniel J, Espejo A, Lake A, Krishna M, Xia L, Zhang Y, Bedford MT. 2006. Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO Rep.* 7: 397–403.

Kim PM, Lam HYK, Urban AE, Korbel JO, Affourtit J, Grubert F, Chen X, Weissman S, Snyder M, Gerstein MB. 2008. Analysis of copy number variants and segmental duplications in the human genome: Evidence for a change in the process of formation in recent evolutionary history. *Genome Res.*18: 1865–74.

Kinoshita K, Shinka T, Sato Y, Kurahashi H, Kowa H, Chen G, Umeno M, Toida K, Kiyokage E, Nakano T, Ito S, Nakahori Y. 2006. Expression analysis of a mouse orthologue of HSFY, a candidate for the azoospermic factor on the human Y chromosome. *J Med Invest.* I 53: 117–22.

Kleiman SE, Almog R, Yogev L, Hauser R, Lehavi O, Paz G, Yavetz H, Botchan A. 2012. Screening for partial AZFa microdeletions in the Y chromosome of infertile men: is it of clinical relevance? *Fertil Steril.* 98: 43–7.

Kleiman SE, Bar-Shira Maymon B, Hauser R, Botchan A, Paz G, Yavetz H, Yogev L. 2008. Histone H4 acetylation and AZFc involvement in germ cells of specimens of impaired spermatogenesis. *Fertil Steril.* 89: 1728–36.

Kosova G, Scott NM, Niederberger C, Prins GS, Ober C. 2012. Genome-wide association study identifies candidate genes for male fertility traits in humans. *Am J Hum Genet.* 90: 950–61.

Krausz C. 2011. Male infertility: pathogenesis and clinical diagnosis. *Best practice & research. Clin Endocrinol Metab.* 25: 271–85.

Krausz C. 2012. An encore for the repeats: new insights into an old genetic variant. *J Clin Endocrinol Metab.* 97: 764–7.

Krausz C, Chianese C, Giachini C, Guarducci E, Laface I, Forti G. 2011. The Y chromosome-linked copy number variations and male fertility. *J Endocrinol Invest.* 34: 376–82.

Krausz C, Degl'Innocenti S. 2006. Y chromosome and male infertility: update, 2006. *Front Biosci.* 11: 3049–61.

Krausz C, Degl'Innocenti S, Nuti F, Morelli A, Felici F, Sansone M, Varriale G, Forti G. 2006. Natural transmission of USP9Y gene mutations: a new perspective on the role of AZFa genes in male fertility. Hum Mol genet 15: 2673–81.

Krausz C, Giachini C, Giacco D Lo, Daguin F, Chianese C, Ars E, Ruiz-Castane E, Forti G, Rossi E. 2012. High resolution X chromosome-specific array-CGH detects new CNVs in infertile males. PloS one 7: e44887.

Krausz C, Giachini C, Xue Y, O'Bryan MK, Gromoll J, Rajpert-de Meyts E, Oliva R, Aknin-Seifer I, Erdei E, Jorgensen N, Simoni M, Ballescà JL, et al. 2009. Phenotypic variation within European carriers of the Y-chromosomal gr/gr deletion is independent of Y-chromosomal background. *J Med Genet.* 46: 21–31.

Krausz C, Quintana-Murci L, Barbaux S, Siffroi JP, Rouba H, Delafontaine D, Souleyreau-Therville N, Arvis G, Antoine JM, Erdei E, Taar JP, Tar A, et al. 1999. A high frequency of Y chromosome deletions in males with nonidiopathic infertility. *J Clin Endocrinol Metab.* 84: 3606–12.

Krausz C, Quintana-Murci L, Fellous M, Siffroi JP, McElreavey K. 2000a. Absence of mutations involving the INSL3 gene in human idiopathic cryptorchidism. *Mol Hum Reprod.* 6: 298–302.

Krausz C, Quintana-Murci L, McElreavey K. 2000b. Prognostic value of Y deletion analysis: what is the clinical prognostic value of Y chromosome microdeletion analysis? *Hum Reprod.* 15: 1431–4.

Kubota Y, Nef S, Farmer PJ, Temelcos C, Parada LF, Hutson JM. 2001. Leydig insulin-like hormone, gubernacular development and testicular descent. *J Urol.*165: 1673–5.

Kubota Y, Temelcos C, Bathgate RAD, Smith KJ, Scott D, Zhao C, Hutson JM. 2002. The role of insulin 3, testosterone, Müllerian inhibiting substance and relaxin in rat gubernacular growth. *Mol Hum Reprod.* 8: 900–5.

Kukuvitis A, Georgiou I, Bouba I, Tsirka A, Giannouli CH, Yapijakis C, Tarlatzis B, Bontis J, Lolis D, Sofikitis N, Papadimas J. 2002. Association of oestrogen receptor alpha polymorphisms and androgen receptor CAG trinucleotide repeats with male infertility: a study in 109 Greek infertile men. *Int J Androl.* 25: 149–52.

Kurahashi N, Sata F, Kasai S, Shibata T, Moriya K, Yamada H, Kakizaki H, Minakami H, Nonomura K, Kishi R. 2005. Maternal genetic polymorphisms in CYP1A1, GSTM1 and GSTT1 and the risk of hypospadias. *Mol Hum Reprod.* 11: 93–8.

Kuroda-Kawaguchi T, Skaletsky H, Brown LG, Minx PJ, Cordum HS, Waterston RH, Wilson RK, Silber S, Oates R, Rozen S, Page DC. 2001. The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nat Genet.* 29: 279–86.

Lahn BT, Page DC. 1997. Functional coherence of the human Y chromosome. *Science* 278: 675–80.

Lahn BT, Page DC. 1999. Retroposition of autosomal mRNA yielded testis-specific gene family on human Y chromosome. *Nat Genet.*21: 429–33.

Lahn BT, Tang ZL, Zhou J, Barndt RJ, Parvinen M, Allis CD, Page DC. 2002. Previously uncharacterized histone acetyltransferases implicated in mammalian spermatogenesis. *Proc Natl Acad Sci U S A.* 99: 8707–12.

Lahoud MH, Ristevski S, Venter DJ, Jermiin LS, Bertoncello I, Zavarsek S, Hasthorpe S, Drago J, de Kretser D, Hertzog PJ, Kola I 2001 Gene targeting of Desrt, a novel ARID class DNA-binding protein, causes growth retardation and abnormal development of reproductive organs. *Genome Res* 11:1327–34

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409: 860–921.

Lanfranco F, Kamischke A, Zitzmann M, Nieschlag E. Klinefelter's syndrome. *Lancet* 364: 273–83.

Lee C, Scherer SW. 2010. The clinical context of copy number variation in the human genome. *Expert Rev Mol Med.* 12: e8.

Lee JA, Carvalho CMB, Lupski JR. 2007a. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* 131: 1235–47.

Lee KH, Song GJ, Kang IS, Kim SW, Paick J-S, Chung CH, Rhee K. 2003. Ubiquitin-specific protease activity of USP9Y, a male infertility gene on the Y chromosome. *Reprod Fertil Dev.* 15: 129–33.

Lee MG, Norman J, Shilatifard A, Shiekhattar R. 2007b. Physical and functional association of a trimethyl H3K4 demethylase and Ring6a/MBLR, a polycomb-like protein. *Cell* 128: 877–87.

Lercher MJ, Urrutia AO, Hurst LD. 2003. Evidence that the human X chromosome is enriched for male-specific but not female-specific genes. *Mol Biol Evol.* 20: 1113–6.

Levron J, Aviram-Goldring A, Madgar I, Raviv G, Barkai G, Dor J. 2000. Sperm chromosome analysis and outcome of IVF in patients with non-mosaic Klinefelter's syndrome. *Fertil Steril.* 74: 925–9.

Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, MacDonald JR, et al. 2007. The diploid genome sequence of an individual human. *PLoS Biol.* 5: e254.

Li X, Nokkala E, Yan W, Streng T, Saarinen N, Wa¨rri A, Huhtaniemi I, Santti R, Ma¨kela¨ S, Poutanen M 2001 Altered structure and function of reproductive organs in transgenic male mice overexpressing human aromatase. *Endocrinology* 142:2435–42

Lieber MR. 2008. The mechanism of human nonhomologous DNA end joining. *J Biol Chem.*283:1–5.

Lin Y, Page DC. 2005. Dazl deficiency leads to embryonic arrest of germ cell development in XY C57BL/6 mice. *Dev Biol.* 288: 309–16.

Lin Y-W, Hsu LC-L, Kuo P-L, Huang WJ, Chiang H-S, Yeh S-D, Hsu T-Y, Yu Y-H, Hsiao K-N, Cantor RM, Yen PH. 2007. Partial duplication at AZFc on the Y chromosome is a risk factor for impaired spermatogenesis in Han Chinese in Taiwan. *Hum Mutat.* 28: 486–94.

Llanos M de, Ballescà JL, Gázquez C, Margarit E, Oliva R. 2005. High frequency of gr/gr chromosome Y deletions in consecutive oligospermic ICSI candidates. *Hum Reprod.* 20: 216–20.

Lo Giacco D, Ars E, Bassas L, Galan JJ, Rajmil O, Ruíz P, Caffaratti J, Guarducci E, Ruiz-Castané E, Krausz C. 2011. ESR1 promoter polymorphism is not associated with nonsyndromic cryptorchidism. *Fertil Steril.* 95: 369–71, 371.e1–2.

Longman D, Johnstone IL, Cáceres JF. 2000. Functional characterization of SR and SR-related genes in Caenorhabditis elegans. *EMBO J.* 19: 1625–37.

Lopes AM, Aston KI, Thompson E, Carvalho F, Gonçalves J, Huang N, Matthiesen R, Noordam MJ, Quintela I, Ramu A, Seabra C, Wilfert AB, et al. 2013. Human spermatogenic failure purges deleterious mutation load from the autosomes and both sex chromosomes, including the gene DMRT1. *PLoS Genet.* 9: e1003349.

Lu C, Zhang F, Yang H, Xu M, Du G, Wu W, An Y, Qin Y, Ji G, Han X, Gu A, Xia Y, et al. 2011. Additional genomic duplications in AZFc underlie the b2/b3 deletion-associated risk of spermatogenic impairment in Han Chinese population. *Hum Mol Genet.* 20: 4411–21.

Lu C, Zhang J, Li Y, Xia Y, Zhang F, Wu B, Wu W, Ji G, Gu A, Wang S, Jin L, Wang X. 2009. The b2/b3 subdeletion shows higher risk of spermatogenic failure and higher frequency of complete AZFc deletion than the gr/gr subdeletion in a Chinese population. *Hum Mol Genet.* 18: 1122–30.

Luddi A, Margollicci M, Gambera L, Serafini F, Cioni M, Leo V De, Balestri P, Piomboni P. 2009. Spermatogenesis in a man with complete deletion of USP9Y. *N Engl J Med.* 360: 881–5.

Maiburg M, Repping S, Giltay J. 2012. The genetic origin of Klinefelter syndrome and its effect on spermatogenesis. *Fertil Steril.*98: 253–60.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, et al. 2009. Finding the missing heritability of complex diseases. *Nature* 461: 747–53.

Martin RH. 2008. Meiotic errors in human oogenesis and spermatogenesis. *Reprod Biomed Online.* 16: 523–531.

Martínez MC, Bernabé MJ, Gómez E, Ballesteros A, Landeras J, Glover G, Gíl-Salom M, Remohí J, Pellicer A. 2000. Screening for AZF deletion in a large series of severely impaired spermatogenesis patients. *J Androl.* 21: 651–5.

Matzuk MM, Lamb DJ. 2008. The biology of infertility: research advances and clinical challenges. *Nat Med.* 14: 1197–213.

Maurer B, Gromoll J, Simoni M, Nieschlag E. 2001. Prevalence of Y chromosome microdeletions in infertile men who consulted a tertiary care medical centre: the Münster experience. *Andrologia* 33: 27–33.

McLachlan JA. 2001. Environmental signaling: what embryos and evolution teach us about endocrine disrupting chemicals. *Endocr Rev.* 22: 319–41.

McVey M, Lee SE. 2008. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet.* 24: 529–38.

Mohandas TK, Speed RM, Passage MB, Yen PH, Chandley AC, Shapiro LJ. 1992. Role of the pseudoautosomal region in sex-chromosome pairing during male meiosis: meiotic studies in a man with a deletion of distal Xp. *Am J Hum Genet.* 51: 526–33.

Morel F, Bernicot I, Herry A, Bris MJ Le, Amice V, Braekeleer M De. 2003. An increased incidence of autosomal aneuploidies in spermatozoa from a patient with Klinefelter's syndrome. *Fertil Steril.* 79 Suppl 3: 1644–6.

Mueller JL, Mahadevaiah SK, Park PJ, Warburton PE, Page DC, Turner JMA. 2008. The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. *Nat Genet*. 40: 794–9.

Mueller JL, Skaletsky H, Brown LG, Zaghlul S, Rock S, Graves T, Auger K, Warren WC, Wilson RK, Page DC. 2013. Independent specialization of the human and mouse X chromosomes for the male germ line. *Nat Genet*. 1–7.

Mulligan P, Westbrook TF, Ottinger M, Pavlova N, Chang B, Macia E, Shi Y-J, Barretina J, Liu J, Howley PM, Elledge SJ, Shi Y. 2008. CDYL bridges REST and histone methyltransferases for gene repression and suppression of cellular transformation. *Mol Cell.* 32: 718–26.

Mumm S, Molini B, Terrell J, Srivastava A, Schlessinger D. 1997. Evolutionary features of the 4-Mb Xq21.3 XY homology region revealed by a map at 60-kb resolution. *Genome Res.* 7: 307–14.

Najmabadi H, Chai N, Kapali A, Subbarao MN, Bhasin D, Woodhouse E, Yen P, Bhasin S. 1996. Genomic structure of a Y-specific ribonucleic acid binding motif-containing gene: a putative candidate for a subset of male infertility. *J Clin Endocrinol Metab.* 81: 2159–64.

Naranjo S, Voesenek K, de la Calle-Mustienes E, Robert-Moreno A, Kokotas H, Grigoriadou M, Economides J, Van Camp G, Hilgert N, Moreno F, Alsina B, Petersen MB, Kremer H, Gómez-Skarmeta JL. 2010 Multiple enhancers located in a 1-Mb region upstream of POU3F4 promote expression during inner ear development and may be required for hearing.*Hum Genet*. 2010 128:411-9.

Navarro-Costa P. 2012. Sex, rebellion and decadence: the scandalous evolutionary history of the human Y chromosome. *Biochim Biophs Acta* 1822: 1851–63.

Navarro-Costa P, Goncalves J, Plancha CE. 2010. The AZFc region of the Y chromosome: at the crossroads between genetic diversity and male infertility. *Hum Reprod Update*.16: 525–542.

Nef S, Parada LF. 1999. Cryptorchidism in mice mutant for Insl3. *Nature Genet.* 22: 295–9.

Nef S, Parada LF. 2000. Hormones in male sexual development. *Genes Dev.*14: 3075–86.

Nef S, Shipman T, Parada LF. 2000. A molecular basis for estrogen-induced cryptorchidism. *Dev Biol.* 224: 354–61.

Nguyen D-Q, Webber C, Ponting CP. 2006. Bias of selection on human copy-number variants. *PLoS Genet.* 2: e20.

Noma T, Kanai Y, Kanai-Azuma M, Ishii M, Fujisawa M, Kurohmaru M, Kawakami H, Wood SA, Hayashi Y. 2002. Stage- and sex-dependent expressions of Usp9x, an X-linked mouse ortholog of Drosophila Fat facets, during gonadal development and oogenesis in mice. *Mech Dev.* Suppl : S91–5.

Nomura T, Kanzaki T. 1977. Induction of urogenital anomalies and some tumors in the progeny of mice receiving diethylstilbestrol during pregnancy. *Cancer Res.* 37: 1099–104.

Noordam MJ, Westerveld GH, Hovingh SE, Daalen SKM van, Korver CM, Veen F van der, Pelt AMM van, Repping S. 2011. Gene copy number reduction in the azoospermia factor c (AZFc) region and its effect on total motile sperm count. *Hum Mol Genet.* 20: 2457–63.

Nuti F, Krausz C. 2008. Gene polymorphisms/mutations relevant to abnormal spermatogenesis. *Reprod Biomed Online.* 16: 504–513.

Nuti F, Marinari E, Erdei E, El-Hamshari M, Echavarria MG, Ars E, Balercia G, Merksz M, Giachini C, Shaeer KZM, Forti G, Ruiz-Castané E, et al. 2008. The leucine-rich repeat-containing G protein-coupled receptor 8 gene T222P mutation does not cause cryptorchidism. *J Clin Endocrinol Metab.* 93: 1072–6.

O'Donnell L, Robertson KM, Jones ME, Simpson ER. 2001. Estrogen and spermatogenesis. *Endocr Rev.* 22: 289–318.

Olesen IA, Sonne SB, Hoei-Hansen CE, Rajpert-De Meyts E, Skakkebaek NE. 2007. Environment, testicular dysgenesis and carcinoma in situ testis. Best practice & research. *Best Pract Res Clin Endocrinol Metab.*21: 462–78.

Oliva R, Margarit E, Ballescá JL, Carrió A, Sánchez A, Milà M, Jiménez L, Alvarez-Vijande JR, Ballesta F. 1998. Prevalence of Y chromosome microdeletions in oligospermic and azoospermic candidates for intracytoplasmic sperm injection. *Fertil Steril*. 70: 506–10.

Overbeek PA, Gorlov IP, Sutherland RW, Houston JB, Harrison WR, Boettger-Tong HL, Bishop CE, Agoulnik AI 2001 A transgenic insertion causing cryptorchidism in mice. *Genesis* 30:26–35Page DC, Harper ME, Love J, Botstein D. Occurrence of a transposition from the X-chromosome long arm to the Y-chromosome short arm during human evolution. *Nature* 311: 119–23.

Paracchini S, Stuppia L, Gatta V, Palka G, Moro E, Foresta C, Mengua L, Oliva R, Ballescà JL, Kremer JA, Golde RJ van, Tuerlings JH, et al. 2000. Y-chromosomal DNA haplotypes in infertile European males carrying Y-microdeletions. *J Endocrinol Invest*. 23: 671–6.

Park J-H, Kong G-H, Lee S-W. 2002. hMAGE-A1 overexpression reduces TNF-alpha cytotoxicity in ME-180 cells. *Mol Cells* 14: 122–9.

Pask AJ, Kanasaki H, Kaiser UB, Conn PM, Janovick JA, Stockton DW, Hess DL, Justice MJ, Behringer RR 2005 A novel mouse model of hypogonadotrophic hypogonadism: N-ethyl-N-nitrosourea- induced gonadotropin-releasing hormone receptor gene mutation. *Mol Endocrinol* 19:972–81

Pawelczak KS, Turchi JJ. 2008. A mechanism for DNA-PK activation requiring unique contributions from each strand of a DNA terminus and implications for microhomology-mediated nonhomologous DNA end joining. *Nucleic Acids Res.* 36: 4022–31.

Perry GH, Ben-Dor A, Tsalenko A, Sampas N, Rodriguez-Revenga L, Tran CW, Scheffer A, Steinfeld I, Tsang P, Yamada NA, Park HS, Kim J-I, et al. 2008. The fine-scale and complex architecture of human copy-number variation. *Am J Hum Genet*. 82: 685–95.

Perry GH, Tito RY, Verrelli BC. 2007. The evolutionary history of human and chimpanzee Y-chromosome gene loss. *Mol Biol Evol.* 24: 853–9.

Peterlin B, Kunej T, Sinkovec J, Gligorievska N, Zorn B. 2002. Screening for Y chromosome microdeletions in 226 Slovenian subfertile men. *Hum Reprod.* 17: 17–24.

Philibert P, Boizet-Bonhoure B, Bashamboo A, Paris F, Aritake K, Urade Y, Leger J, Sultan C, Poulat F. 2013. Unilateral cryptorchidism in mice mutant for Ptgds. *Hum Mutat.* 34: 278–82.

Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, Conroy J, Magalhaes TR, Correia C, Abrahams BS, Almeida J, Bacchelli E, et al. 2010. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 466: 368–72.

Piotrowski A, Bruder CEG, Andersson R, Diaz de Ståhl T, Menzel U, Sandgren J, Poplawski A, Tell D von, Crasto C, Bogdan A, Bartoszewski R, Bebok Z, et al. 2008. Somatic mosaicism for copy number variation in differentiated human tissues. *Hum Mutat.*29: 1118–24.

Radovick S, Wray S, Lee E, Nicols DK, Nakayama Y, Weintraub BD, Westphal H, Cutler Jr GB, Wondisford FE 1991 Migratory arrest of gonadotropin-releasing hormone neurons in transgenic mice. *Proc Natl Acad Sci USA* 88:3402–6

Raicu F, Popa L, Apostol P, Cimponeriu D, Dan L, Ilinca E, Dracea LL, Marinescu B, Gavrila L. 2003. Screening for microdeletions in human Y chromosome--AZF candidate genes and male infertility. *J Cell Mol Med.* 7: 43–8.

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, et al. 2006. Global variation in copy number in the human genome. *Nature* 444: 444–54.

Reijo R, Lee TY, Salo P, Alagappan R, Brown LG, Rosenberg M, Rozen S, Jaffe T, Straus D, Hovatta O. 1995. Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. *Nat Genet.* 10: 383–93.

Reijo RA, Dorfman DM, Slee R, Renshaw AA, Loughlin KR, Cooke H, Page DC. 2000. DAZ family proteins exist throughout male germ cell development and transit from nucleus to cytoplasm at meiosis in humans and mice. *Biol Reprod.* 63: 1490–6.

Repping S, Daalen SKM van, Brown LG, Korver CM, Lange J, Marszalek JD, Pyntikova T, Veen F van der, Skaletsky H, Page DC, Rozen S. 2006. High mutation rates have driven extensive structural polymorphism among human Y chromosomes. *Nat Genet.*38: 463–7.

Repping S, Daalen SKM van, Korver CM, Brown LG, Marszalek JD, Gianotten J, Oates RD, Silber S, Veen F van der, Page DC, Rozen S. 2004. A family of human Y chromosomes has dispersed throughout northern Eurasia despite a 1.8-Mb deletion in the azoospermia factor c region. *Genomics* 83: 1046–52.

Repping S, Skaletsky H, Brown L, Daalen SKM van, Korver CM, Pyntikova T, Kuroda-Kawaguchi T, Vries JWA de, Oates RD, Silber S, Veen F van der, Page DC, et al. 2003. Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. *Nat Genet.* 35: 247–51.

Reynolds N, Cooke HJ. 2005. Role of the DAZ genes in male fertility. *Reprod Biomed Online* 10: 72–80.

Rodríguez-Santiago B, Brunet A, Sobrino B, Serra-Juhé C, Flores R, Armengol L, Vilella E, Gabau E, Guitart M, Guillamat R, Martorell L, Valero J, et al. 2010. Association of common copy number variants at the glutathione S-transferase genes and rare novel genomic changes with schizophrenia. *Mol Psychiatry* 15: 1023–33.

Roh J, Virtanen H, Kumagai J, Sudo S, Kaleva M, Toppari J, Hsueh AJW. Lack of LGR8 gene mutation in Finnish patients with a family history of cryptorchidism. *Reprod Biomed Online* 7: 400–6.

Roll-Mecak A, Shin BS, Dever TE, Burley SK. 2001. Engaging the ribosome: universal IFs of translation. *Trends Biochem Sci.* 26: 705–9.

Rosner A, Rinkevich B. 2007. The DDX3 subfamily of the DEAD box helicases: divergent roles as unveiled by studying different organisms and in vitro assays. *Curr Med Chem.* 14: 2517–25.

Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, Muzny D, Platzer M, Howell GR, Burrows C, Bird CP, Frankish A, Lovell FL, Howe KL, Ashurst JL, Fulton RS, Sudbrak R, Wen G, Jones MC, Hurles ME, Andrews TD, Scott CE, Searle S, Ramser J, Whittaker A, Deadman R, Carter NP, Hunt SE, Chen R, Cree A, Gunaratne P, Havlak P, Hodgson A, Metzker ML, Richards S, Scott G, Steffen D, Sodergren E, Wheeler DA, Worley KC, Ainscough R, Ambrose KD, Ansari-Lari MA, Aradhya S, Ashwell RI, Babbage AK, Bagguley CL, Ballabio A, Banerjee R, Barker GE, Barlow KF, Barrett IP, Bates KN, Beare DM, Beasley H, Beasley O, Beck A, Bethel G, Blechschmidt K, Brady N, Bray-Allen S, Bridgeman AM, Brown AJ, Brown MJ, Bonnin D, Bruford EA, Buhay C, Burch P, Burford D, Burgess J, Burrill W, Burton J, Bye JM, Carder C, Carrel L, Chako J, Chapman JC, Chavez D, Chen E, Chen G, Chen Y, Chen Z, Chinault C, Ciccodicola A, Clark SY, Clarke G, Clee CM, Clegg S, Clerc-Blankenburg K, Clifford K, Cobley V, Cole CG, Conquer JS, Corby N, Connor RE, David R, Davies J, Davis C, Davis J, Delgado O, Deshazo D, et al. 2005. The DNA sequence of the human X chromosome. *Nature* 434: 325-37

Rozen S, Skaletsky H, Marszalek JD, Minx PJ, Cordum HS, Waterston RH, Wilson RK, Page DC. 2003. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* 423: 873–6.

Rozen SG, Marszalek JD, Irenze K, Skaletsky H, Brown LG, Oates RD, Silber SJ, Ardlie K, Page DC. 2012. AZFc deletions and spermatogenic failure: a population-based survey of 20,000 Y chromosomes. *Am J Hum Genet.* 91: 890–6.

Sabol SZ, Hu S, Hamer D. 1998. A functional polymorphism in the monoamine oxidase A gene promoter. *Hum Genet.* 103: 273–9.

Salonia A, Matloob R, Gallina A, Abdollah F, Saccà A, Briganti A, Suardi N, Colombo R, Rocchini L, Guazzoni G, Rigatti P, Montorsi F. 2009. Are infertile men less healthy than fertile men? Results of a prospective case-control survey. *Eur Urol.* 56: 1025–31.

Santos C, Rodriguez-Revenga L, Madrigal I, Badenas C, Pineda M, Milà M. 2006. A novel mutation in JARID1C gene associated with mental retardation. *Eur J Hum Genet.* 14: 583–6.

Sato Y, Yoshida K, Shinka T, Nozawa S, Nakahori Y, Iwamoto T. 2006. Altered expression pattern of heat shock transcription factor, Y chromosome (HSFY) may be related to altered differentiation of spermatogenic cells in testes with deteriorated spermatogenesis. *Fertil Steril.* 86: 612–8.

Satokata I, Benson G, Maas R 1995 Sexually dimorphic sterility phenotypes in Hoxa 10-deficient mice. *Nature* 374:460–463

Saus E, Brunet A, Armengol L, Alonso P, Crespo JM, Fernández-Aranda F, Guitart M, Martín-Santos R, Menchón JM, Navinés R, Soria V, Torrens M, et al. 2010. Comprehensive copy number variant (CNV) analysis of neuronal pathways genes in psychiatric disorders identifies rare variants within patients. *J Psychiatr Res.* 44: 971–8.

Saxena R, Brown LG, Hawkins T, Alagappan RK, Skaletsky H, Reeve MP, Reijo R, Rozen S, Dinulos MB, Disteche CM, Page DC. 1996. The DAZ gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned. *Nat Genet.* 14: 292–9.

Saxena R, Vries JW de, Repping S, Alagappan RK, Skaletsky H, Brown LG, Ma P, Chen E, Hoovers JM, Page DC. 2000. Four DAZ genes in two clusters found in the AZFc region of the human Y chromosome. *Genomics* 67: 256–67.

Schempp W, Binkele A, Arnemann J, Gläser B, Ma K, Taylor K, Toder R, Wolfe J, Zeitler S, Chandley AC. 1995. Comparative mapping of YRRM- and TSPY-related cosmids in man and hominoid apes. *Chromosome Res.*3: 227–34.

Schulte RT, Ohl D a, Sigman M, Smith GD. 2010. Sperm DNA damage in male infertility: etiologies, assays, and outcomes. *J Assit Reprod Genet.* 27: 3–12.

Schwartz A, Chan DC, Brown LG, Alagappan R, Pettay D, Disteche C, McGillivray B, la Chapelle A de, Page DC. 1998. Reconstructing hominid Y evolution: X-homologous block, created by X-Y transposition, was disrupted by Yp inversion through LINE-LINE recombination. *Hum Mol Genet.* 7: 1–11.

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Månér S, Massa H, Walker M, Chi M, Navin N, Lucito R, et al. 2004. Large-scale copy number polymorphism in the human genome. *Science* 305: 525–8.

Sekiguchi T, Iida H, Fukumura J, Nishimoto T. 2004. Human DDX3Y, the Y-encoded isoform of RNA helicase DDX3, rescues a hamster temperature-sensitive ET24 mutant cell line with a DDX3X mutation. *Exp Cell Res.* 300: 213–22.

Senno L del, Aguiari GL, Piva R. 1992. Dinucleotide repeat polymorphism in the human estrogen receptor (ESR) gene. *Hum Mol Genet.* 1: 354.

Shamoo Y, Abdul-Manan N, Williams KR. 1995. Multiple RNA binding domains (RBDs) just don't add up. *Nucleic Acids Res.* 23: 725–8.

Sharpe RM. 2003. The "oestrogen hypothesis"- where do we stand now? *Int J Androl.* 26: 2–15.

Sharpe RM, Skakkebaek NE. 1993. Are oestrogens involved in falling sperm counts and disorders of the male reproductive tract? *Lancet* 341: 1392–5.

Shaw CJ, Lupski JR. 2004. Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. *Hum Mol Genet.* 13 Spec No: R57–64.

Shi Q, Martin RH. 2001. Aneuploidy in human spermatozoa: FISH analysis in men with constitutional chromosomal abnormalities, and in infertile men. *Reproduction* 121: 655–66.

Shi Y, Sawada J, Sui G, Affar EB, Whetstine JR, Lan F, Ogawa H, Luke MP-S, Nakatani Y, Shi Y. 2003. Coordinated histone modifications mediated by a CtBP co-repressor complex. *Nature* 422: 735–8.

Shinka T, Sato Y, Chen G, Naroda T, Kinoshita K, Unemi Y, Tsuji K, Toida K, Iwamoto T, Nakahori Y. 2004. Molecular characterization of heat shock-like factor encoded on the human Y chromosome, and implications for male infertility. *Biol Reprod.* 71: 297–306.

Simoni M, Tüttelmann F, Gromoll J, Nieschlag E. 2008. Clinical consequences of microdeletions of the Y chromosome: the extended Münster experience. *Reprod Biomed Online* 16: 289–303.

Sin H-S, Koh E, Shigehara K, Sugimoto K, Maeda Y, Yoshida A, Kyono K, Namiki M. 2010. Features of constitutive gr/gr deletion in a Japanese population. *Hum Reprod.* 25: 2396–403.

Skakkebaek NE, Rajpert-De Meyts E, Main KM. 2001. Testicular dysgenesis syndrome: an increasingly common developmental disorder with environmental aspects. *Hum Reprod.* 16: 972–8.

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T, Chinwalla A, Delehaunty A, et al. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423: 825–37.

Skrisovska L, Bourgeois CF, Stefl R, Grellscheid S-N, Kister L, Wenter P, Elliott DJ, Stevenin J, Allain FH-T. 2007. The testis-specific human protein RBMY recognizes RNA through a novel mode of interaction. *EMBO Rep.* 8: 372–9.

Smith EP, Boyd J, Frank GR, Takahashi H, Cohen RM, Specker B, Williams TC, Lubahn DB, Korach KS. 1994. Estrogen resistance caused by a mutation in the estrogen-receptor gene in a man. *N Engl J Med.* 331: 1056–61.

Song R, Ro S, Michaels JD, Park C, McCarrey JR, Yan W. 2009. Many X-linked microRNAs escape meiotic sex chromosome inactivation. *Nat Genet.* 41: 488–93.

Spada AR La, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH. 1991. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* 352: 77–9.

Staessen C, Tournaye H, Assche E Van, Michiels A, Landuyt L Van, Devroey P, Liebaers I, Steirteghem A Van. 2003. PGD in 47,XXY Klinefelter's syndrome patients. *Hum Reprod Update* 9: 319–30.

Stahl PJ, Masson P, Mielnik A, Marean MB, Schlegel PN, Paduch DA. 2010. A decade of experience emphasizes that testing for Y microdeletions is essential in American men with azoospermia and severe oligozoospermia. *Fertil Steril.* 94: 1753–6.

Stankiewicz P, Lupski JR. 2010. Structural variation in the human genome and its role in disease. *Annu Rev Med.* 61: 437–55.

Stevenson BJ, Iseli C, Panji S, Zahn-Zabal M, Hide W, Old LJ, Simpson AJ, Jongeneel CV. 2007. Rapid evolution of cancer/testis genes on the X chromosome. *BMC genomics* 8: 129.

Stouffs K, Lissens W, Landuyt L Van, Tournaye H, Steirteghem A Van, Liebaers I. 2001. Characterization of the genomic organization, localization and expression of four PRY genes (PRY1, PRY2, PRY3 and PRY4). *Mol Hum Reprod.* 7: 603–10.

Stouffs K, Lissens W, Tournaye H, Haentjens P. 2011. What about gr/gr deletions and male infertility? Systematic review and meta-analysis. *Hum Reprod Update* 17: 197–209.

Stouffs K, Lissens W, Verheyen G, Landuyt L Van, Goossens A, Tournaye H, Steirteghem A Van, Liebaers I. 2004. Expression pattern of the Y-linked PRY gene suggests a function in apoptosis but not in spermatogenesis. *Mol Hum Reprod.*10: 15–21.

Stouffs K, Tournaye H, Liebaers I, Lissens W. 2009. Male infertility and the involvement of the X chromosome. *Hum Reprod Update* 15: 623–37.

Stouffs K, Vandermaelen D, Massart A, Menten B, Vergult S, Tournaye H, Lissens W. 2012. Array comparative genomic hybridization in male infertility. *Hum Reprod.* 27: 921–9.

Stuppia L, Gatta V, Fogh I, Gaspari AR, Morizio E, Mingarelli R, Santo M Di, Pizzuti A, Calabrese G, Palka G. 2001. Genomic organization, physical mapping, and involvement in Yq microdeletions of the VCY2 (BPY 2) gene. *Genomics* 72: 153–7.

Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW. 1983. The double-strand-break repair model for recombination. *Cell* 33: 25–35.

Tahiliani M, Mei P, Fang R, Leonor T, Rutenberg M, Shimizu F, Li J, Rao A, Shi Y. 2007. The histone H3K4 demethylase SMCX links REST target genes to X-linked mental retardation. *Nature* 447: 601–5.

Tajul-Arifin K, Teasdale R, Ravasi T, Hume DA, Mattick JS. 2003. Identification and analysis of chromodomain-containing proteins encoded in the mouse transcriptome. *Genome Res.* 13: 1416–29.

Tang K-F, Zheng J-Z, Xing J-P. 2011. Molecular analysis of SNP12 in estrogen receptor α gene in hypospadiac or cryptorchid patients from Northwestern China. *Urol Int.* 87: 359–62.

Tessari A, Salata E, Ferlin A, Bartoloni L, Slongo ML, Foresta C. 2004. Characterization of HSFY, a novel AZFb gene on the Y chromosome with a possible role in human spermatogenesis. *Mol Hum Reprod.* 10: 253–8.

Thomas NS, Hassold TJ. 2003. Aberrant recombination and the origin of Klinefelter syndrome. *Hum Reprod Update* 9: 309–17.

Tiepolo L, Zuffardi O. 1976. Localization of factors controlling spermatogenesis in the nonfluorescent portion of the human Y chromosome long arm. *Hum Genet.* 34: 119–24.

Toppari J, Larsen JC, Christiansen P, Giwercman A, Grandjean P, Guillette LJ, Jégou B, Jensen TK, Jouannet P, Keiding N, Leffers H, McLachlan JA, et al. 1996. Male reproductive health and environmental xenoestrogens. *Environ Health Perspect.* 104 Suppl : 741–803.

Tse JYM, Wong EYM, Cheung ANY, O WS, Tam PC, Yeung WSB. 2003. Specific expression of VCY2 in human male germ cells and its involvement in the pathogenesis of male infertility. *Biol Reprod.* 69: 746–51.

Turner DJ, Miretti M, Rajan D, Fiegler H, Carter NP, Blayney ML, Beck S, Hurles ME. 2008. Germline rates of de novo meiotic deletions and duplications causing several genomic disorders. *Nat Genet.* 40: 90–5.

Tüttelmann F, Rajpert-De Meyts E, Nieschlag E, Simoni M, Meyts ER-D. 2007. Gene polymorphisms and male infertility – a meta-analysis and literature review. *Reprod Biomed Online* 15: 643–658.

Tüttelmann F, Simoni M, Kliesch S, Ledig S, Dworniczak B, Wieacker P, Röpke A. 2011. Copy number variants in patients with severe oligozoospermia and Sertoli-cell-only syndrome. *PloS one* 6: e19426.

Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D, Olson M V, Eichler EE. 2005. Fine-scale structural variation of the human genome. *Nat Genet.* 37: 727–32.

Tyler-Smith C. 2008. An evolutionary perspective on Y-chromosomal variation and male infertility. *Int J Androl.* 31: 376–82.

Tyler-Smith C, Krausz C. 2009. The will-o'-the-wisp of genetics--hunting for the azoospermia factor gene. *N Engl J Med.* 360: 925–7.

Underhill PA, Kivisild T. 2007. Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annu Rev Genet.* 41: 539–64.

Venables JP, Elliott DJ, Makarova O V, Makarov EM, Cooke HJ, Eperon IC. 2000. RBMY, a probable human spermatogenesis factor, and other hnRNP G proteins interact with Tra2beta and affect splicing. *Hum Mol genet.* 9: 685–94.

Venables JP, Vernet C, Chew SL, Elliott DJ, Cowmeadow RB, Wu J, Cooke HJ, Artzt K, Eperon IC. 1999. T-STAR/ETOILE: a novel relative of SAM68 that interacts with an RNA-binding protein implicated in spermatogenesis. *Hum Mol genet.* 8: 959–69.

Vincent M-C, Daudin M, De MP, Massat G, Mieusset R, Pontonnier F, Calvas P, Bujan L, Bourrouillout G. 2002. Cytogenetic investigations of infertile men with low sperm counts: a 25-year experience. *J Androl.* 23: 18–22; discussion 44–5.

Virtanen HE, Bjerknes R, Cortes D, Jørgensen N, Rajpert-De Meyts E, Thorsson A V, Thorup J, Main KM. 2007. Cryptorchidism: classification, prevalence and long-term consequences. *Acta Paediatr* 96: 611–616.

Visser L, Westerveld GH, Korver CM, Daalen SKM van, Hovingh SE, Rozen S, Veen F van der, Repping S. 2009. Y chromosome gr/gr deletions are a risk factor for low semen quality. *Hum Reprod.* 24: 2667–73.

Vogt PH, Falcao CL, Hanstein R, Zimmer J. 2008. The AZF proteins. *Int J Androl.* 31: 383–94.

Wang PJ, McCarrey JR, Yang F, Page DC. 2001. An abundance of X-linked genes expressed in spermatogonia. *Nat Genet.* 27: 422–6.

Wang PJ, Page DC, McCarrey JR. 2005. Differential expression of sex-linked and autosomal germ-cell-specific genes during spermatogenesis in the mouse. Human molecular genetics 14: 2911–8.

Wang Y, Barthold J, Figueroa E, González R, Noh PH, Wang M, Manson J. 2008. Analysis of five single nucleotide polymorphisms in the ESR1 gene in cryptorchidism. *Birth defects Res A Clin Mol Teratol.* 82: 482–5.

Watanabe M, Yoshida R, Ueoka K, Aoki K, Sasagawa I, Hasegawa T, Sueoka K, Kamatani N, Yoshimura Y, Ogata T. 2007. Haplotype analysis of the estrogen receptor 1 gene in male genital and reproductive abnormalities. *Hum Reprod.* 22: 1279–84.

Weterings E, Gent DC van. 2004. The mechanism of non-homologous end-joining: a synopsis of synapsis. *DNA Repair* 3: 1425–35.

WHO. 2010. WHO laboratory manual for the examination and processing of human semen - 5th edition.

Wikström AM, Bay K, Hero M, Andersson A-M, Dunkel L. 2006. Serum insulin-like factor 3 levels during puberty in healthy boys and boys with Klinefelter syndrome. *J Clin Endocrinol Metab.* 91: 4705–8.

Willcocks LC, Lyons PA, Clatworthy MR, Robinson JI, Yang W, Newland SA, Plagnol V, McGovern NN, Condliffe AM, Chilvers ER, Adu D, Jolly EC, et al. 2008. Copy number of FCGR3B, which is associated with systemic lupus erythematosus, correlates with protein expression and immune complex uptake. *J Exp Med.* 205: 1573–82.

Wong EYM, Tse JYM, Yao K-M, Lui VCH, Tam P-C, Yeung WSB. 2004. Identification and characterization of human VCY2-interacting protein: VCY2IP-1, a microtubule-associated protein-like protein. *Biol Reprod.* 70: 775–84.

Wong EYM, Tse JYM, Yao K-M, Tam P-C, Yeung WSB. 2002. VCY2 protein interacts with the HECT domain of ubiquitin-protein ligase E3A. *Biochem Biophys Res Commun.* 296: 1104–11.

Wu B, Lu NX, Xia YK, Gu AH, Lu CC, Wang W, Song L, Wang SL, Shen HB, Wang XR. 2007. A frequent Y chromosome b2/b3 subdeletion shows strong association with male infertility in Han-Chinese population. *Hum Reprod.* 22: 1107–13.

Wu H, Min J, Antoshenko T, Plotnikov AN. 2009. Crystal structures of human CDY proteins reveal a crotonase-like fold. *Proteins* 76: 1054–61.

Xu EY, Moore FL, Pera RA. 2001. A gene family required for human germ cell development evolved from an ancient meiotic gene conserved in metazoans. *Proc Natl Acad Sci U S A.* 98: 7414–9.

Yamazawa K, Wada Y, Sasagawa I, Aoki K, Ueoka K, Ogata T. 2007. Mutation and polymorphism analyses of INSL3 and LGR8/GREAT in 62 Japanese patients with cryptorchidism. *Horm Res.* 67: 73–6.

Yang Y, Ma M, Li L, Su D, Chen P, Ma Y, Liu Y, Tao D, Lin L, Zhang S. 2010. Differential effect of specific gr/gr deletion subtypes on spermatogenesis in the Chinese Han population. *Int J Androl.* 33: 745–54.

Yang Y, Ma M, Li L, Zhang W, Chen P, Ma Y, Liu Y, Tao D, Lin L, Zhang S. 2008a. Y chromosome haplogroups may confer susceptibility to partial AZFc deletions and deletion effect on spermatogenesis impairment. Human reproduction (Oxford, England) 23: 2167–72.

Yang Y, Ma M, Li L, Zhang W, Xiao C, Li S, Ma Y, Tao D, Liu Y, Lin L, Zhang S. 2008b. Evidence for the association of Y-chromosome haplogroups with susceptibility to spermatogenic failure in a Chinese Han population. *J Med Genet.* 45: 210–5.

Yoshida R, Fukami M, Sasagawa I, Hasegawa T, Kamatani N, Ogata T. 2005. Association of cryptorchidism with a specific haplotype of the estrogen receptor alpha gene: implication for the susceptibility to estrogenic environmental endocrine disruptors. *J Clin Endocrinol Metab.* 90: 4716–21.

Yu VWC, Gauthier C, St-Arnaud R. 2006. Inhibition of ATF4 transcriptional activity by FIAT/gamma-taxilin modulates bone mass accrual. *Ann N Y Acad Sci.* 1068: 131–42.

Zegers-Hochschild F, Adamson GD, Mouzon J de, Ishihara O, Mansour R, Nygren K, Sullivan E, Vanderpoel S. 2009. International Committee for Monitoring Assisted Reproductive Technology (ICMART) and the World Health Organization (WHO) revised glossary of ART terminology, 2009. *Fertil Steril*. 92: 1520–4.

Zeng M, Sun H, Chen S, Wang X, Yang Y, Liu Y, Tao D, Yang Z, Zhang S, Ma Y. 2008. Identification of target messenger RNA substrates for mouse RBMY. *Mol Hum Reprod.* 14: 331–6.

Zhang F, Lu C, Li Z, Xie P, Xia Y, Zhu X, Wu B, Cai X, Wang X, Qian J, Wang X, Jin L. 2007. Partial deletions are associated with an increased risk of complete deletion in AZFc: a new insight into the role of partial AZFc deletions in male infertility. *J Med Genet.* 44: 437–44.

Zhang FP, Poutanen M, Wilbertz J, Huhtaniemi I 2001 Normal prenatal but arrested postnatal sexual development of luteinizing hormone receptor knockout (LuRKO) mice. *Mol Endocrinol* 15: 172–183

Zhao H, Xu J, Zhang H, Sun J, Sun Y, Wang Z, Liu J, Ding Q, Lu S, Shi R, You L, Qin Y, et al. 2012. A genome-wide association study reveals that variants within the HLA region are associated with risk for nonobstructive azoospermia. *Am J Hum Genet.* 90: 900–6.

Zheng K, Yang F, Wang PJ. 2010. Regulation of male fertility by X-linked genes. *J Androl.* 31: 79–85.

Zhou X, Benson KF, Ashar HR, Chada K 1995 Mutation responsible for the mouse pygmy phenotype in the developmentally regulated factor HMGI-C. *Nature* 376:771–4

Zhou D, Xia Y, Li Y, Song L, Hu F, Lu C, Zhou Z, Sha J, Wang X. 2011. Higher proportion of haploid round spermatids and spermatogenic disomy rate in relation to idiopathic male infertility. *Urology* 77: 77–82.

Zimmermann S, Steding G, Emmen JM, Brinkmann AO, Nayernia K, Holstein AF, Engel W, Adham IM. 1999. Targeted disruption of the Insl3 gene causes bilateral cryptorchidism. *Mol Endocrinol.* 13: 681–91.Zuppan PJ, Hall JM, Ponglikitmongkol M et al. 1989. Polymorphisms at the estrogen receptor (ESR) locus and linkage relationships on chromosome 6q. *Cytogenet Cell Genet* 51: 1116.