Evaluation of 172 candidate polymorphisms for association with oligozoospermia or azoospermia in a large cohort of men of European

Hum. Reprod. Advance Access published April 8, 2010
Human Reproduction, Vol.00, No.0 pp. 1–15, 2010
doi:10.1093/humrep/deq081

human reproduction ORIGINAL ARTICLE *Andrology*

Evaluation of 172 candidate polymorphisms for association with oligozoospermia or azoospermia in a large cohort of men of European descent

Kenneth I. Aston[1], Csilla Krausz[2,3], Ilaria Laface[2], E. Ruiz-Castané[3], and Douglas T. Carrell[1,4,5,*]

[1]Andrology and IVF Laboratories, University of Utah School of Medicine, Salt Lake City, UT, USA [2]Andrology Unit, Department of Clinical Physiopathology, University of Florence, Florence, Italy [3]Andrology Service, Fundació Puigvert, Barcelona, Spain [4]Department of Surgery, Department of Obstetrics and Gynecology, University of Utah School of Medicine, Salt Lake City, UT, USA [5]Department of Physiology, University of Utah School of Medicine, Salt Lake City, UT, USA

*Correspondence address. E-mail: douglas.carrell@hsc.utah.edu

Submitted on January 26, 2010; resubmitted on March 2, 2010; accepted on March 9, 2010

BACKGROUND: In spite of tremendous efforts by a number of groups, the search for single nucleotide polymorphisms (SNPs) strongly associated with male factor infertility by means of gene re-sequencing studies has yielded few likely candidates. A recent pilot, genome-wide SNP association study (GWAS) identified a list of SNPs associated with oligozoospermia and azoospermia. This is an expanded follow-up study of the SNPs identified by the GWAS as well as other SNPs from previously published gene re-sequencing studies.

METHODS: On the basis of the pilot GWAS and SNPs with published associations with male infertility, 172 SNPs were genotyped in men with idiopathic azoospermia or oligozoospermia using the Illumina BeadXpress® platform.

RESULTS: Several SNPs were identified or confirmed to be significantly associated with oligozoospermia and/or azoospermia. More importantly, this follow-up study indicates that, at least in Caucasian men, no single common SNP accounts for a significant proportion of spermatogenic failure cases.

CONCLUSIONS: The associations reported in this study are promising, but much larger genome-wide studies will be necessary to confidently validate these SNPs and identify novel SNPs associated with male infertility.

Key words: SNP / male infertility / azoospermia / oligozoospermia / gene mutations

Introduction

Infertility affects approximately one in seven couples worldwide resulting in significant financial and emotional costs, and male factor infertility accounts for about half of all infertility cases. Despite its prevalence, infertility research has received much less attention than most common, complex diseases.

Common known genetic causes of male infertility include Klinefelter's syndrome and Y chromosome microdeletions, both resulting in severe oligozoospermia or azoospermia, and cystic fibrosis transmembrane conductance regulator mutations, which result in congenital bilateral absence of vas deferens leading to obstructive azoospermia (Nuti and Krausz, 2008). Taken together, known genetic causes of male infertility account for less than one-third of all male factor infertility cases leaving a large proportion of cases classified as idiopathic (Dohle *et al.*, 2002).

Due to the limited attention given to male infertility research and a concomitant lack of substantial research funding, the majority of research has focused on targeted gene re-sequencing studies in search of genetic risk factors. Although several single nucleotide polymorphisms (SNPs) associated with various classes of male infertility have been reported, for the majority, associations are weak and replication studies often fail to validate initial findings. To date only a single pilot genome-wide SNP association study (GWAS) has been performed for idiopathic male infertility (Aston and Carrell, 2009). In this pilot study from our laboratory, no single SNP reached the level

of genome-wide significance after adjusting for multiple comparisons (Bonferroni adjustment), however, we did find 20 SNPs with a raw *P*-value for association $<1 \times 10^{-5}$ and an estimated false discovery rate of <0.3 (Aston and Carrell, 2009).

Although the pilot GWAS represented an important first step in the search for genetic variants associated with severe oligozoospermia and non-obstructive azoospermia on a genome-wide scale, the study evaluated a limited number of samples and the associations detected were marginal for a GWAS. In the initial study, we stressed the need for follow-up studies with increased sample size to evaluate the preliminary results. For the present study, we have evaluated a larger population of azoospermic and oligozoospermic men utilizing BeadXpress® (Illumina Inc., San Diego, CA, USA) technology as a cost effective medium-throughput assay for a targeted genotyping follow-up of the findings of the initial study. Additionally, we have simultaneously genotyped this population of men for SNPs identified in prior gene re-sequencing studies and target genes of spermatogenesis. We evaluated the data for the follow-up study separately, and after combining with the data from the initial GWAS.

Materials and Methods

Study samples

Details of patient and control selection for the pilot microarray study were published previously (Aston and Carrell, 2009). Caucasian individuals of central European descent, including normospermic controls ($n = 80$) with >20 million sperm/ml, severe oligozoospermic individuals ($n = 50$) with <5 million sperm/ml and non-obstructive azoospermic individuals ($n = 33$) with no detectable sperm in the ejaculate following evaluation of the centrifuged pellet and with no other known etiology, were included in the study.

Patient and control samples used in the follow-up targeted genotyping study included individuals of European decent-primarily of Mediterranean origin. Infertile patients included in the study were seeking a complete andrological diagnostic work-up for couple infertility. All infertile patients were defined as 'idiopathic' and selected on the basis of a comprehensive andrological examination including medical history and physical examination, semen analysis, scrotal ultrasound, hormone analysis, karyotype and Y chromosome microdeletion screening. Patients with mono or bilateral cryptorchidism, varicocele, previous testis trauma, obstructive azoospermia, recurrent infections, iatrogenic infertility, hypogonadotrophic hypogonadism, karyotype anomalies or Y chromosome microdeletions including gr/gr deletions were excluded.

In all, 158 normospermic controls (>20 million sperm/ml), 141 severe oligozoospermic individuals (<5 million sperm/ml) and 80 non-obstructive azoospermic individuals were included in the follow-up study. In addition, 63 moderately oligozoospermic individuals (5–10 million sperm/ml) were analyzed, and associations with all oligozoospermic men are reported separately.

DNA extraction

DNA was extracted from whole blood using the Gentra Puregene Blood Kit (Qiagen Inc., Valencia, CA, USA) according to manufacturer's recommendations. Briefly, red blood cells were lysed, and white cells pelleted followed by white cell lysis. Protein precipitation solution was added and lysis solution spun to remove proteins. DNA was then precipitated using 100% isopropanol, washed and re-suspended in Tris-EDTA. DNA concentration was assessed using the NanoDrop 1000 spectrophotometer (Thermo Scientific, Waltham, MA, USA) and subsequently adjusted to ~50 ng/μl. Purified DNA was stored at −80°C prior to use.

Follow-up targeted genotyping assay design

A total of 84 SNPs were evaluated as a follow-up to the initial GWAS, microarray pilot study. Three different groups of SNPs were selected for follow-up based on associations detected in the microarray phase of the study. Since the pilot study was limited in its power to detect by the relatively small sample size, we used a lower threshold *P*-value for selection of SNPs for the follow-up study. This was done to improve our chances of validating true associations in spite of the lack of power. SNPs with a *P*-value for association $<5 \times 10^{-5}$ ($n = 32$), SNPs located within 10 kb of genes with known fertility function based on phenotype in mouse gene knockout models (Bult *et al.*, 2008; Matzuk and Lamb, 2008) with a *P*-value for association $<5 \times 10^{-3}$ ($n = 20$), and non-synonymous SNPs with a *P*-value for association $<5 \times 10^{-3}$ ($n = 32$) were selected for follow-up with additional samples (Table I).

In addition to SNPs with significant associations based on the microarray pilot study, 21 SNPs associated with spermatogenic defects reported in literature (Table II) were evaluated as well as 67 non-synonymous SNPs located within genes important in spermatogenesis (Table I). SNPs targeted for genotyping based on previous publications were located in AHRR, BRCA2, ER beta, ERCC1, ESR1, FASLG FHL5, FKBP6, IL1B, KIT, MS, MTHFR, MTRR, PRM1, TSSK4, UBE2B and YBX2 (Table II). Non-synonymous SNPs in spermatogenesis genes included SNPs in AR, BRDT, CAMK4, CREM, DDX25, DMC1, FHL5, JMJD1A, KIF17, MSH4, MSH5, PRM1, PRM2, REC8, SMC1B, SPATA22, SPO11, STRA8, SYCP1, SYCP2, TEX11, TEX14, TEX15, TNP1, TNP2, USP26 and YBX2 (Table III). In all, the final BeadXpress assay targeted 172 SNPs (Table I).

Follow-up genotyping

Genotyping was performed on purified DNA samples using the Illumina BeadXpress® Assay (Illumina Inc., San Diego, CA, USA) according to manufacturer's recommendations. Briefly, DNA was purified from peripheral blood lymphocytes as described above. Approximately 500 ng of purified DNA at a concentration of 50 ng/μl was delivered to the University of Utah Genomics Core Facility (Salt Lake City, UT, USA) for genotyping where it was prepared by several activation and ligation steps followed by universal PCR and finally hybridization to SNP-specific genotyping beads. Following hybridization and wash steps, reaction plates were scanned, the raw data were normalized across samples and genotype calls were made using BeadStudio Software® (Illumina Inc).

Genotype and haplogroup association analysis

Associations were tested using GoldenHelix® SVS7 analysis software (Bozeman, MT, USA). Initial data cleanup was performed to remove poorly performing probes and samples from analysis. Probes with a call rate <0.95 and probes with significant deviations from Hardy–Weinberg equilibrium (HWE) within controls of $P < 0.001$ were removed prior to association analysis. Samples with a call rate of <0.85 were also removed from analysis.

SNP association analysis was performed applying allelic, additive, dominant and recessive models to the data. Allelic association was performed on autosomal probes, whereas all probes were analyzed using the other models. Primary comparisons made for association testing included azoospermic versus normospermic (AvN), severe oligozoospermic versus normospermic (OvN) and azoospermic and severe oligozoospermic combined versus normospermic (A+OvN). Only severe oligozoospermic

Table 1 Details of SNPs genotyped

Markers not analyzed	Design Category	SNP ID	Chromosome	Nearest gene	Coding status	Markers not analyzed	Design category	SNP ID	Chromosome	Nearest gene	Coding status
CR	M	rs2472493	9	ABCA1	NC		M	rs300503	12	NAV3	NC
	M	rs10491278	5	AFF4	NC	CR	M	rs1142530	19	NDUFS7	NS
	M	rs4705873	5	AFF4	NC		M	rs12913	16	NKD1	NC
	P	rs2292596	5	AHRR	NS		M	rs1400699	2	NR4A2	NC
MAF 0	S	rs9332968	X	AR	NS		M	rs11196389	10	NRAP	NS
MAF 0	S	rs9332970	X	AR	NS	CR	M	rs9825719	3	NSUN3	NC
CR	M	rs9814870	3	ARL6	NC		M	rs1399645	2	NXPH2	NC
	M	rs6763208	3	ARL6	NC		M	rs2063802	2	NXPH2	NC
	M	rs2290870	4	ATP8A1	NC		M	rs4954657	2	NXPH2	NC
	M	rs8100856	19	ATP8B3	NS		M	rs11204546	1	OR2W3	NS
	P	rs144848	13	BRCA2	NS		M	rs1491584	2	PAX8	NC
MAF 0	S	rs34674879	1	BRDT	NS		M	rs2863242	2	PAX8	NC
	S	rs10747493	1	BRDT	NS		M	rs4849179	2	PAX8	NC
	S	rs10783071	1	BRDT	NS		M	rs1047854	19	PCSK4	NC
	S	rs3088232	1	BRDT	NS		M	rs10841496	12	PDE3A	NC
	M	rs11577579	1	C1orf125	NS		M	rs4926219	19	PKN1	NS
	M	rs3746804	20	C20orf54	NS		M	rs1046116	12	PKP2	NS
MAF 0	S	rs35548075	5	CAMK4	NS		M	rs16981262	19	PLA2G4C	NC
CR	M	rs735295	12	CCDC77	NS		P	rs2301365	16	PRM1	NC
	M	rs11264963	1	CD1E	NC		S	rs35576928	16	PRM1	NS
	M	rs2072671	1	CDA	NS	CR	S	rs3177008	16	PRM2	NS
CR	M	rs3130981	6	CDSN	NS		M	rs4484160	3	PROK2	NC
	M	rs11707608	3	CNTN3	NC		M	rs4450678	20	PTPRT	NC
	M	rs2976084	3	CNTN3	NC	CR	S	rs35425516	14	REC8	NS
	M	rs1545125	7	COBL	NC	MAF 0	S	rs34075659	14	REC8	NS
MAF 0	S	rs34227693	10	CREM	NS		M	rs2291219	8	RPESP	NS
MAF 0	S	rs561704	11	DDX25	NS		M	rs6531	6	RXRB	Syn
MAF 0	S	rs2227914	22	DMC1	NS		M	rs6068020	20	SALL4	NC
	M	rs529208	9	DOCK8	NS		M	rs11435	5	SH3RF2	NS
	M	rs6444952	3	EIF5A2	NC		M	rs6476866	9	SLC1A1	NC
	P	rs1256049	14	ER beta	Syn	CR	M	rs4884207	13	SLITRK1	NC
	P	rs3212986	19	ERCC1	NC	CR	S	rs9614653	22	SMC1B	NS
	P	ESR1325C>G	6	ESR1	Syn	MAF 0	S	rs12157922	22	SMC1B	NS

Continued

Table 1 *Continued*

Markers not analyzed	Design Category	SNP ID	Chromosome	Nearest gene	Coding status	Markers not analyzed	Design category	SNP ID	Chromosome	Nearest gene	Coding status
	P	rs2234693	6	ESR1	NC	MAF 0	S	rs16993928	22	SMC1B	NS
	M	rs7342883	17	EVPL	NS		S	rs5764698	22	SMC1B	NS
	P	rs763110	1	FASL	NC		M	rs7224496	17	SMYD4	NS
CR	P	rs2273621	6	FHL5	NC	MAF 0	S	rs12602543	17	SPATA22	NS
	P	rs9398152	6	FHL5	NC		S	rs1488690	17	SPATA22	NS
	S	rs35157931	6	FHL5	NS		S	rs2291604	17	SPATA22	NS
	S	rs9373985	6	FHL5	NS	MAF 0	S	rs28368082	20	SPO11	NS
MAF 0	P	rs3750075	7	FKBP6	NC	MAF 0	S	rs3736832	20	SPO11	NS
	M	rs11114486	12	FLJ90579	NS		M	rs7746261	6	SSR1	NS
	M	rs2083098	2	FMNL2	NC	MAF 0	S	rs12669242	7	STRA8	NS
	M	rs2032278	18	GALR1	NC	MAF 0	S	rs1053812	1	SYCP1	NS
tr9>CR	M	rs4343755	4	GNPDA2	NC	MAF 0	S	rs12563933	1	SYCP1	NS
	M	rs11718848	3	GOLGA4	NS	CR	S	rs3736763	20	SYCP2	NS
	M	rs983034	1	GPR177	NS	MAF 0	S	rs1359836	20	SYCP2	NS
	M	rs10492239	12	GPR19	NC	MAF 0	S	rs6128714	20	SYCP2	NS
	M	rs12213993	6	HSF2	NC		S	rs6071006	20	SYCP2	NS
	M	rs3778348	6	HSF2	NC		M	rs10246939	7	TAS2R38	NS
	M	rs9375131	6	HSF2	NC		M	rs1726866	7	TAS2R38	NS
	P	rs1143634	2	IL1B	Syn		M	rs2010834	18	TCEB3B	NS
	M	rs2059807	19	INSR	NC		M	rs4255473	10	TCF7L2	NC
CR	M	rs7146310	14	IPO4	NS		S	rs4844247	X	TEX11	NS
MAF 0	S	rs11677451	2	JMJD1A	NS		S	rs6525433	X	TEX11	NS
	S	rs2030259	2	JMJD1A	NS	CR	S	rs35551271	17	TEX14	NS
	S	rs34605051	2	JMJD1A	NS	MAF 0	S	rs34960869	17	TEX14	NS
	S	rs2296225	1	KIF17	NS	MAF 0	S	rs35081269	17	TEX14	NS
	S	rs522496	1	KIF17	NS	MAF 0	S	rs35195402	17	TEX14	NS
	S	rs631357	1	KIF17	NS	MAF 0	S	rs35927726	17	TEX14	NS
	P	rs3819392	4	KIT	NC	MAF 0	S	rs7220834	17	TEX14	NS
	M	rs268917	19	KLK4	NC		S	rs389389	17	TEX14	NS
	M	rs981684	17	KRT25C	NS	CR	S	rs323347	8	TEX15	NS
	M	rs6755901	2	LHCGR	NC	MAF 0	S	rs9297162	8	TEX15	NS
	M	rs5911500	X	LOC203413	NC		S	rs323343	8	TEX15	NS
	M	rs7889596	X	LOC203413	NC		S	rs323344	8	TEX15	NS
	M	rs4541736	6	LRFN2	NC		S	rs323345	8	TEX15	NS
	M	rs215702	7	LSM5	NC		S	rs323346	8	TEX15	NS

CR	M	rs12920268	16	MAF	NC		M	rs3777654	6	TINAG	NC
	M	rs3105782	3	MASP1	NC		M	rs11800462	1	TNFRSF25	NS
	M	rs3795958	2	MGC16372	NS	MAF 0	S	TNP188GA	2	TNP1	NS
HWE	M	rs9610624	22	MGC35206	NS	MAF 0	S	TNP217AG	16	TNP2	NS
CR	M	rs10848911	12	MGC4266	NC		S	rs34904070	16	TNP2	NS
	M	rs11024970	11	MRGPRX2	NS		S	TNP2391CT	16	TNP2	NS
	P	rs1805087	1	MS	NS	CR	M	rs2283817	22	TPST2	NC
CR	S	rs5745325	1	MSH4	NS	MAF 0	P	TSSK4987+108G>A	14	TSSK4	NC
MAF 0	S	rs5745329	1	MSH4	NS		P	rs17167484	5	UBE2B	NC
	S	rs5745549	1	MSH4	NS		P	rs3777373	5	UBE2B	NC
CR	S	rs28381359	6	MSH5	NS		M	rs3736563	3	UBP1	NS
	S	rs1802127	6	MSH5	NS	MAF 0	S	rs4582709	X	USP26	NS
	S	rs28381349	6	MSH5	NS		S	rs35397110	X	USP26	NS
	S	rs28399977	6	MSH5	NS	CR	P	YBX2intr6	17	YBX2	NC
	P	rs1801133	1	MTHFR	NS	CR	S	rs391768	17	YBX2	NS
	M	rs12537	22	MTMR3	NC		P	YBX2exon8	17	YBX2	Syn
	P	rs1801394	5	MTRR	NS	CR	M	rs2704838	X	ZFX	NC
	M	rs3826942	19	MUM1	NS		M	rs12912744	15	ZNF690	NS

Markers not analyzed: associations were not tested on some markers due to genotyping call rate <0.95 (CR), deviation from HWE ($P < 0.001$) or a lack of the minor alleles in all samples (MAF 0). Design category: selection of SNPs for study was based on previous microarray results (M), SNPs with previously published associations with infertility (P) or non-synonymous SNPs in spermatogenesis genes (S). Coding Status: SNPs analyzed were categorized as non-coding (NC), non-synonymous (NS) or synonymous (Syn).

Table II SNPs selected for study based on previously published reports of association with male infertility

SNP	Gene	Coding status	Cases	Controls	Ethnicity	Results	Reference
rs2292596	AHRR	NS	112 azoo/oligos	212 normo controls	Estonian	GG 31.5% cases, 18% contr $P = 0.006$	Merisalu *et al.* (2007)
rs144848	BRCA2	NS	240 azoo/sev oligo	250 proven fertiles	Western Chinese	C 23.5% cases, 17.6% contr	Zhoucun *et al.* (2006)
rs1256049	ER beta	S	106 sev oligo, 86 test. cancer pts, 51 hypospadias, 23 cryptorchid	186>5 million/ml	Caucasian	AG 13.2% cases, 4.3% contr	Aschim *et al.* (2005)
rs3212986	ERCC1	NC	202 azoo	187 fert contr	Chinese	A 31.7% cases, 21.7% contr $P = 0.002$	Ji *et al.* (2008)
ESR1325C>G	ESR1	S	31 azoo	46 fert conrt	Japanese	CG+GG 90% cases, 66% contr $P < 0.01$	Suzuki *et al.* (2002)
rs2234693	ESR1	NC	104 azoo/sev oligo	95 unselected contr	Spanish	TT 40.4% cases, 31.6% contr $P = 0.006$	Galan *et al.* (2005)
rs763110	FASL	NC	161 azoo/23 sev oligo	236 fert contr	Chinese	TT 10.3% cases, 3.8% contr $P = 0.008$	Wang *et al.* (2009)
rs2273621	FHL5	NC	47 azoo/49 sev oligo	69 fert contr	Caucasian	A 76% cases 62% $P = 0.0072$	Christensen *et al.* (2006)
rs9398152	FHL5	NC	47 azoo/49 sev oligo	69 fert contr	Caucasian	T 92% cases, 81% contr $P = .0023$	Christensen *et al.* (2006)
rs3750075	FKBP6	NC	323 azoo/sev oligo	205 fert contr	Western Chinese	A 1.7% cases, 5.4% contr $P = 0.0007$	Zhang *et al.* (2007)
rs1143634	IL1B	S	435 non-normo cases (incl 95 OAT)	127 normo contr	Caucasian	TT 7% of all cases, 1% contr, 14% OAT $P = 0.001$ OAT versus contr	Bentz *et al.* (2007)
rs3819392	KIT	NC	167 azoo/sev oligo	465 unselected contr	Spanish	A 30.1% cases, 37.9% contr $P = 0.01$	Galan *et al.* (2006)
rs1805087	MS	NS	174 azoo, 186 OAT	325 fert contr	Korean	GG 3.1% cases, 1.2% contr $P = 0.09$ GG sig increased in azoos w/MTHFR 677 CC $P = 0.018$	Lee *et al.* (2006)
rs1801133	MTHFR	NS	174 azoo, 186 OAT	325 fert contr	Korean	TT 17.8% cases, 12.6% contr $P = 0.048$	Lee *et al.* (2006)
rs1801133	MTHFR	NS	179 OAT	200 fert contr	Indian	TT 11.7% cases, 15% contr $P = 0.15$	Dhillon *et al.* (2007)
rs1801133	MTHFR	NS	151 azoo/sev oligo	200 fert contr	Indian	TT 4% cases, 0 contr $P = 0.004$	Singh *et al.* (2005)
rs1801133	MTHFR	NS	228 azoo, 127 sev oligo	252 fert, normo contr	Chinese	TT 21% azoos, 13.4% oligos, 11.5% contr $P = 0.023$ cases versus contr, $P = 0.004$ azoos versus contr	A *et al.* (2007)
rs1801133	MTHFR	NS	286 azoo, 59 sev OAT, 26 OAT, 2 oligo	396 fert contr	Korean	TT 16.9% cases, 12.9% contr $P = 0.019$	Park *et al.* (2005)
rs1801133	MTHFR	NS	93 infert	105 fert contr	Italian	TT 20.4% cases, 27.6% contr not sig.	Stuppia *et al.* (2003)
rs1801133	MTHFR	NS	77 subfertile	113 fert contr	Caucasian	TT 9.1% cases, 13.3% contr not sig.	Ebisch *et al.* (2003)
rs1801133	MTHFR	NS	255 infert	200 contr	Caucasian	TT 18.8% cases, 9.5% contr $P = .008$	Bezold *et al.* (2001)
rs1801394	MTRR	NS	174 azoo, 186 OAT	327 fert contr	Korean	GG 12.8% cases, 8.9% contr $P = 0.047$	Lee *et al.* (2006)

rs2301365	PRM1	NC	65 infert w/ <9% normal forms, 155 w/ >9%	101 contr	Caucasian	AA 13.8% <9%, 4.5% >9%, 3.0% contr $P < .005 < 9\%$ versus contr	Gazquez *et al.* (2008)
TSSK4987+108G>A	TSSK4	NC	219 azoo, 153 sev oligo	220 normo fertile contr	Western Chinese	GA 5.5% cases, 15% contr $P < 0.001$	Su *et al.* (2008)
rs17167484	UBE2B	NC	350 azoo, 75 OA, 105 OAT	300 fert contr	Indian	G 6.2% cases, 2.0% contr $P = 0.002$	Suryavathi *et al.* (2008)
rs3777373	UBE2B	NC	350 azoo, 75 OA, 105 OAT	300 fert contr	Indian	G 23% cases, 5.3% contr $P < .001$	Suryavathi *et al.* (2008)
YBX2exon8	YBX2	S	47 azoo, 49 sev oligo, 96 abn prm	96 fert contr	Caucasian	AA 100% contr, 43.3% azoo, oligo, 40.0% abn prm $P = 0.0001$	Hammoud *et al.* (2009)
YBX2intr6	YBX2	NC	47 azoo, 49 sev oligo, 96 abn prm	96 fert contr	Caucasian	CT 32.8% contr, 36.8% azoo, oligo, 1.8% abn prm $P = 0.0001$ abn prm versus contr	Hammoud *et al.* (2009)

Coding Status: SNPs analyzed were categorized as non-coding (NC), non-synonymous (NS) or synonymous (S). Azoo, azoospermia; sev oligo, severe oligozoospermia; abn prm, abnormal protamine; OAT, oligoasthenoteratozoospermia; infert, infertile; normo, normospermic; pt, patient; contr, control; sig., significant.

Table III Non-synonymous SNPs selected for study based on gene function

Gene	SNP	Function
AR	rs9332968	Transcription factor for androgen-responsive genes
	rs9332970	
BRDT	rs34674879	Transcriptional regulation during spermatogenesis
	rs10747493	
	rs10783071	
	rs3088232	
CAMK4	rs35548075	Transcriptional regulation during spermatogenesis
CREM	rs34227693	Role in spermatogenesis and spermatid maturation
DDX25	rs561704	mRNA export and translation during spermatid development
DMC1	rs2227914	Meiotic recombination
FHL5	rs35157931	Activates CREM
	rs9373985	
JMJD1A	rs11677451	Histone demethylase, regulates chromatin packaging during spermatogenesis
	rs2030259	
	rs34605051	
KIF17	rs2296225	Regulates subcellular distribution of FHL5
	rs522496	
	rs631357	
MSH4	rs5745325	Meiotic recombination
	rs5745329	
	rs5745549	
MSH5	rs28381359	DNA mismatch repair, meiotic recombination
	rs1802127	
	rs28381349	
	rs28399977	
PRM1	rs35576928	Sperm chromatin compaction
PRM2	rs3177008	Sperm chromatin compaction
REC8	rs35425516	Homologous chromosome and sister chromatid separation
	rs34075659	
SMC1B	rs9614653	Synapsis and meiotic recombination
	rs12157922	
	rs16993928	
	rs5764698	
SPATA22	rs12602543	Testes development
	rs1488690	
	rs2291604	
SPO11	rs28368082	DNA double stand breaks during meiosis
	rs3736832	
STRA8	rs12669242	Initiation of meiosis
SYCP1	rs1053812	Component of synaptonemal complex

Continued

Table III *Continued*

Gene	SNP	Function
	rs12563933	
SYCP2	rs3736763	Component of synaptonemal complex
	rs1359836	
	rs6128714	
	rs6071006	
TEX11	rs4844247	Regulates meiotic crossovers
	rs6525433	
TEX14	rs35551271	Formation of germ cell intercellular bridges
	rs34960869	
	rs35081269	
	rs35195402	
	rs35927726	
	rs7220834	
	rs389389	
TEX15	rs323347	Chromosomal synapsis
	rs9297162	
	rs323343	
	rs323344	
	rs323345	
	rs323346	
TNP1	TNP188GA	Intermediate protein in sperm chromatin compaction
TNP2	TNP217AG	Intermediate protein in sperm chromatin compaction
	rs34904070	
	TNP2391CT	
USP26	rs4582709	Ubiquitin-dependent proteolytic pathway in testis
	rs35397110	
YBX2	rs391768	Stabilization of testis mRNAs

samples were included in the initial analysis to match the phenotypes included in the pilot GWAS.

In addition to the primary comparisons, an additional aim of the study was to evaluate the question of whether azoospermia or severe oligozoospermia result from SNPs specific to each individual infertility phenotype, or whether SNPs common to both of these groups result in differing severities of spermatogenic defects. In order to address this question, oligozoospermic and normospermic data combined served as the control group and were compared with azoospermic individuals (A versus O+N), and conversely azoospermic and normospermic groups were combined and compared with oligozoospermic individuals (O versus A+N). A *P*-value of <0.05 was considered significant.

In addition to a univariate analysis to evaluate the association between single SNPs and spermatogenic defects, haplogroup association was also performed to evaluate whether a specific haplogroup was more powerful in predicting a phenotype than a single SNP. For this analysis, a moving window of 100 kb was used to define haplogroups. The expected-maximization (EM) algorithm with 50 iterations and a convergence tolerance of 0.0001 was employed for haplotype estimation (Excoffier and Slatkin, 1995). A *P*-value of <0.05 for association was considered significant.

As haplogroup association assumes linkage between SNPs, additional analysis was performed to determine whether azoospermic and oligozoospermic samples carried more total independent risk SNPs compared with controls based on associations detected by univariate analysis. Differences in total numbers of risk SNPs between groups were evaluated using the un-paired Student's *t*-test.

Oligozoospermic sperm count and genotype analysis

For SNPs found to be associated specifically with oligozoospermia, further analysis was performed to determine whether an association between the 'risk allele' and sperm concentration existed. To perform this analysis, samples were grouped based on genotype for each associated SNP, and sperm concentrations were compared between groups.

Results

Sample quality

There were 158 normospermic controls, 63 moderately oligozoospermic samples, 141 severe oligozoospermic samples and 80 non-obstructed azoospermic samples analyzed by BeadXpress® assay. Poorly performing samples with genotyping call rates of <0.85 ($n=2$ severe oligozoospermic samples) were removed from association testing leaving 158 controls, 63 moderately oligozoospermic samples, 139 severe oligozoospermic samples and 80 azoospermic samples. Five samples were run in duplicate across plates, and average SNP concordance between samples was 99% across all markers.

Genotype associations for microarray follow-up SNPs

Association data from the pilot microarray study were recently published (Aston and Carrell, 2009). A total of 84 markers were designed to target SNPs as a follow-up to results obtained following a pilot genome-wide association study. Probes with a call rate <0.95 ($n=13$), and probes with significant deviations from HWE within controls of $P<0.001$ ($n=1$), were removed prior to association analysis, leaving 70 markers for analysis.

Of the 70 SNPs evaluated for association in the follow-up study based on associations detected in the pilot study, four SNPs displayed improved associations upon combining genotypes from the two projects (Table IV).

Genotype associations for published and spermatogenesis gene SNPs

Additionally, 67 non-synonymous SNPs in genes important in spermatogenesis, and 21 SNPs with published associations with spermatogenic defects for a total of 88 SNPs were genotyped. Because 11 markers had a call rate of <0.95, they were removed, leaving 77 markers included in the analysis.

Of the 77 SNPs evaluated for association based on previously published associations or SNP function, 31 displayed a minor allele frequency of 0 in our study population, so those SNPs were not included in association analysis. The remaining 46 SNPs were tested for association with azoospermia or oligozoospermia, and five SNPs were found to have significant associations at $P<0.05$ (Table IV).

Table IV SNPs with significant associations with azoospermia, severe oligozoospermia or all infertiles

Design category	Test	SNP ID	Chromosome	Gene	Microarray P	Targeted genotyping P A/SO/N samples	Microarray + targeted genotyping combined P A/SO/N samples	Minor Allele	Major Allele	Cases				Controls			
										% D	% DD	% Dd	% dd	% D	% DD	% Dd	% dd
P	SOvN dom	rs763110	1	FASLG	NA	2.81×10^{-3}	NA	A	G	0.50	0.19	0.60	0.20	0.36	0.12	0.49	0.39
S	AvN rec	rs34605051	2	JMJD1A	NA	3.23×10^{-3}	NA	G	A	0.18	0.08	0.21	0.71	0.16	0.01	0.30	0.69
M	SOvN add	rs5911500	X	LOC203413	4.77×10^{-5}	4.21×10^{-3}	8.32×10^{-7}	T	C	0.08	NA	NA	NA	0.27	NA	NA	NA
S	A+SOvN all	rs323344	8	TEX15	NA	8.22×10^{-3}	NA	C	A	0.08	0.01	0.14	0.84	0.15	0.04	0.22	0.75
S	A+SOvN all	rs323345	8	TEX15	NA	1.21×10^{-2}	NA	G	A	0.09	0.01	0.15	0.84	0.15	0.04	0.22	0.75
S	SOvN rec	rs3088232	1	BRDT	NA	2.63×10^{-2}	NA	G	C	0.21	0.07	0.28	0.65	0.18	0.02	0.34	0.65
M	A+SOvN add	rs11204546	1	OR2W3	2.42×10^{-4}	4.60×10^{-2}	1.87×10^{-4}	T	C	0.28	0.07	0.41	0.52	0.38	0.15	0.46	0.39
M	SOvN all	rs2059807	19	INSR	1.10×10^{-4}	4.74×10^{-2}	3.24×10^{-4}	T	C	0.30	0.06	0.48	0.46	0.42	0.19	0.48	0.34
M	AvN dom	rs10246939	7	TAS2R38	1.45×10^{-3}	7.37×10^{-2}	7.24×10^{-4}	C	T	0.33	0.10	0.44	0.45	0.48	0.24	0.49	0.27

SNPs are listed in order of significance based on *P*-value in the targeted genotyping study.
Design category: selection of SNPs for study was based on previous microarray results (M), SNPs with previously published associations with infertility (P) or non-synonymous SNPs in spermatogenesis genes (S). Test: describes the comparison made for association testing as well as the genetic model applied for association testing. A, azoospermic; SO, severe oligozoospermic; N, normospermic; AvN, azoospermic versus normospermic; SOvN, severe oligozoospermic versus normospermic; A+SOvN, azoospermic+severe oligozoospermic versus normospermic; add, additive; all, allelic; dom, dominant; rec, recessive; NA, SNPs not selected for follow-up based on initial microarray results; D, minor allele; d, major allele.

Interestingly, only one of these five significant SNPs was selected based on previously published data. In other words, only one SNP from previous re-sequencing studies showed significant association in this study.

Effect of inclusion of moderate oligozoospermic samples on SNP associations

When moderately oligozoospermic samples were included along with the severe oligozoospermic samples, associations for six SNPs improved, whereas associations for eight SNPs declined (Table V).

SNP associations with inclusion of alternative comparisons

When A versus O+N and O versus A+N comparisons were made, associations improved for four SNPs that were significant using the conventional A versus N, O versus N and A+O versus N comparisons. In addition, five more SNPs reached significance with the added statistical power derived from a larger number of control samples (Table V).

Haplogroup associations

A total of seven haplogroups were found to display significant associations with infertile phenotypes at $P < 0.05$ when A versus N, O versus N and A+O versus N comparisons were made excluding moderate oligozoospermia (Table VI). Including the comparisons A versus O+N and O versus A+N in the association testing increased the number of significant associations to nine. Inclusion of moderate oligozoospermic patients in the oligozoospermia group resulted in improved associations for only one haplogroup and reduced associations for three haplogroups (Table VII). Analysis of the number of risk SNPs per sample revealed that, as expected, multiple risk SNPs were present at a higher frequency in cases compared with controls ($P < 0.05$).

Oligozoospermic sperm count and genotype analysis

Comparison of mean sperm counts in oligozoospermic samples displaying a 'risk allele' versus those without for the SNPs most significantly associated with oligozoospermia revealed no difference in sperm count between groups for any of the SNPs analyzed.

Discussion

The data presented here represents the most comprehensive SNP study for male infertility performed to date. In all, genotype data for 147 SNPs across several hundred samples is reported. Of the 147 SNPs evaluated, we report significant associations for a total of 14 SNPs at $P < 0.05$ when all samples and comparisons were included (Table V). It should be noted that the *P*-values for association reported here have not been corrected for multiple comparisons. With 147 independent tests, a *P*-value of 0.0003 would be required to achieve a Bonferroni-adjusted *P*-value of <0.05.

The most significant association found in this study is for rs5911500. This is an intergenic SNP located on the X chromosome ~200 kb from solute carrier family 6, member 14 (SLC6A14). On the basis of the intergenic location of rs5911500, if the SNP is a functional variant responsible for the oligozoospermia phenotype, it is most likely involved in a regulatory capacity. Indeed, many SNPs associated with a variety of complex diseases are located in gene deserts, providing further evidence that uncharacterized functional elements are located in those regions (Easton and Eeles, 2008; Mathew, 2008).

Three SNPs located within receptor protein genes were significantly associated with azoospermia or oligozoospermia. The SNP rs2059807, associated with oligozoospermia, is an intronic SNP located on chromosome 19 of the insulin receptor gene (INSR). Rs11204546 is a non-synonymous SNP located in the olfactory receptor, family 2, subfamily W, member 3 gene (OR2W3) on chromosome 1. This polymorphism is associated with both azoospermia and severe oligozoospermia and results in a conservative Methionine to Valine amino acid change. Rs10246939 is another non-synonymous SNP located in the taste receptor, type 2, member 38 gene (TAS2R38) on chromosome 7 and is associated specifically with azoospermia. This SNP also results in a conservative amino acid change of Isoleucine to Valine. Although none of these receptors have been directly implicated in spermatogenesis, receptor proteins of a variety of classes have been shown to be present in the testes (Oonk and Grootegoed, 1987; Thomas *et al.*, 1996).

SNPs with significant associations with spermatogenic failure were also found in testis expressed 15 (TEX15), fas ligand (FASLG), bromodomain, testis-specific (BRDT) and lysine-specific demethylase 3A (JMJD1A). The associated SNPs in TEX15, BRDT and JMJD1A all result in conservative amino acid substitutions, and the SNP in FASLG is a non-coding SNP located in the promoter region of the gene. Each of these genes is specifically involved in spermatogenesis.

TEX15 was recently shown to be necessary for spermatogenesis but not oogenesis. Knockout of the gene in mice resulted in early meiotic arrest in males resulting in a complete lack of germ cells (Yang *et al.*, 2008). Further analysis revealed TEX15 is required for normal chromosomal synapsis and that the protein is likely involved in the loading of DNA repair machinery at the point of DNA double-strand breaks, with gene loss resulting in the meiotic arrest as observed in *Tex15*-deficient male mice (Yang *et al.*, 2008). Although the associated SNPs result in conservative amino acid changes to the protein, a direct link to impaired spermatogenesis in unclear.

The SNP rs763110 in the promoter region of FASLG, a gene involved in germ cell apoptosis induction, was recently shown by meta-analysis to be associated with cancer susceptibility (Zhang *et al.*, 2009). This SNP was selected for study based on a recent report of its association with male infertility (Wang *et al.*, 2009). In this report, azoospermic and severe oligozoospermic men together more frequently carried the risk genotype than controls (Wang *et al.*, 2009). In contrast, we found the SNP to associate with severe oligozoospermia, with no apparent association with azoospermia.

The other two SNPs with significant associations with spermatogenic failure both reside in genes involved in chromatin remodeling. The SNP rs3088232, located in the gene BRDT is associated with severe oligozoospermia in our study. BRDT is a testis-specific protein that interacts with acetylated lysine residues of histones and other proteins (Pivot-Pajot *et al.*, 2003). Hyperacetylation of histones occurs in male germ cells prior to histone-to-prtotamine replacement.

Table V SNPs with significant associations with azoospermia, severe oligozoospermia or moderate oligozoospermia including alternate comparisons

Design category	Test	SNP ID	Chromosome	Gene	Microarray P	Targeted genotyping P A/SO/N samples	Targeted genotyping P A/O/N samples	Microarray + Targeted genotyping combined P A/SO/N samples	Microarray + Targeted genotyping combined P A/O/N samples	Minor Allele	Major Allele	Cases				Controls			
												% D	% DD	% Dd	% dd	% D	% DD	% Dd	% dd
P	OvA+N dom	rs763110*	1	FASLG	NA	1.60×10^{-4}	2.80×10^{-4}	NA	NA	A	G	0.50	0.19	0.60	0.20	0.36	0.12	0.49	0.39
S	AvO+N rec	rs34605051*	2	JMJD1A	NA	7.70×10^{-4}	3.88×10^{-3}	NA	NA	G	A	0.18	0.08	0.21	0.71	0.16	0.01	0.30	0.69
M	OvN add	rs5911500*	X	LOC203413	4.77×10^{-5}	4.21×10^{-3}	1.45×10^{-2}	8.323×10^{-7}	2.91×10^{-6}	T	C	0.08	0.08	0.00	0.92	0.27	0.27	0.00	0.73
M	AvO+N dom	rs10246939*	7	TAS2R38	2.13×10^{-4}	5.13×10^{-3}	2.72×10^{-3}	1.58×10^{-5}	7.82×10^{-6}	C	T	0.33	0.10	0.44	0.45	0.50	0.24	0.51	0.25
S	OvA+N rec	rs3088232*	1	BRDT	NA	6.38×10^{-3}	9.99×10^{-3}	NA	NA	G	C	0.21	0.07	0.28	0.65	0.18	0.02	0.34	0.65
S	A+OvN all	rs323344*	8	TEX15	NA	8.22×10^{-3}	3.99×10^{-3}	NA	NA	C	A	0.08	0.01	0.14	0.84	0.15	0.04	0.22	0.75
S	A+OvN all	rs323345*	8	TEX15	NA	1.21×10^{-2}	5.68×10^{-3}	NA	NA	G	A	0.09	0.01	0.15	0.84	0.15	0.04	0.22	0.75
S	AvO+N dom	rs5764698	22	SMC1B	NA	1.93×10^{-2}	8.76×10^{-3}	NA	NA	A	C	0.53	0.24	0.59	0.18	0.47	0.24	0.45	0.31
P	OvA+N rec	rs1801131	1	MTHFR	NA	2.36×10^{-2}	8.07×10^{-2}	NA	NA	C	A	0.29	0.14	0.31	0.55	0.28	0.07	0.42	0.51
S	OvA+N rec	rs631357	1	KIF17	NA	2.40×10^{-2}	1.52×10^{-2}	NA	NA	C	G	0.19	0.04	0.29	0.66	0.16	0.01	0.30	0.69
S	AvO+N dom	rs35397110	X	USP26	NA	2.45×10^{-2}	3.13×10^{-2}	NA	NA	A	G	0.09	0.09	0.00	0.91	0.03	0.03	0.00	0.97
S	AvO+N all	rs2030259	2	JMJD1A	NA	2.61×10^{-2}	2.64×10^{-2}	NA	NA	A	G	0.31	0.14	0.35	0.51	0.23	0.07	0.32	0.61
M	A+OvN add	rs11204546*	1	OR2W3	2.42×10^{-4}	4.60×10^{-2}	9.45×10^{-2}	1.87×10^{-4}	6.67×10^{-4}	T	C	0.28	0.07	0.41	0.52	0.38	0.15	0.46	0.39
M	OvN all	rs2059807*	19	INSR	1.10×10^{-3}	4.74×10^{-2}	2.32×10^{-1}	3.24×10^{-4}	4.26×10^{-3}	T	C	0.30	0.06	0.48	0.46	0.42	0.19	0.48	0.34

SNPs are listed in order of significance based on *P*-value in the targeted genotyping study.
Design category: selection of SNPs for study was based on previous microarray results (M), SNPs with previously published associations with infertility (P) or non-synonymous SNPs in spermatogenesis genes (S). Test: describes the comparison made for association testing as well as the genetic model applied for association testing. A, azoospermic; SO, severe oligozoospermic; O, oligozoospermic and severe oligozoospermic combined; N, normospermic; AvN, azoospermic versus normospermic; OvN, severe oligozoospermic versus normospermic, A+OvN, azoospermic + oligozoospermic versus normospermic; AvO+N, azoospermic versus oligozoospermic + normospermic; OvA+N, oligospermic versus azoospermic + normospermic; add, additive; all, allelic; dom, dominant; rec, recessive; *, denotes SNPs presented in Table IV; NA, SNPs not selected for follow-up based on initial microarray results; D, minor allele; d, major allele.

Table VI Haplogroups significantly associated with azoospermia, oligozoospermia or all infertiles

Design Category	Comparison	First Marker	Chromosome	Gene	Haplotype	EM cases freq.	EM controls freq.	χ^2 P A/SO/N samples	χ^2 P A/O/N samples
S	OvN	rs323343	8	TEX15	AAAGG	0.08	0.03	6.59×10^{-3}	8.98×10^{-3}
S	A+OvN	rs323343	8	TEX15	ACGGG	0.08	0.14	8.34×10^{-3}	1.83×10^{-3}
M	AvN	rs10246939	7	TAS2R38	AA	0.66	0.54	8.76×10^{-3}	8.76×10^{-3}
M	OvN	rs2863242	2	PAX8	AG	0.01	0.03	3.46×10^{-2}	1.59×10^{-1}
M	OvN	rs4849179	2	PAX8	AAG	0.01	0.03	3.70×10^{-2}	1.44×10^{-1}
S	AvN	rs34904070	16	TNP2	GAACC	0.30	0.22	4.60×10^{-2}	4.60×10^{-2}
S	AvN	rs323344	8	TEX15	CGGG	0.08	0.15	4.66×10^{-2}	4.66×10^{-2}

SNPs are listed in order of significance based on *P*-value in the targeted genotyping study.
Design category: selection of SNPs for study was based on previous microarray results (M), SNPs with previously published associations with infertility (P) or non-synonymous SNPs in spermatogenesis genes (S). Test: describes the comparison made for association testing. A, azoospermic; SO, severe oligozoospermic; O, oligozoospermic and severe oligozoospermic combined; N, normospermic; AvN, azoospermic versus normospermic; OvN, oligozoospermic versus normospermic; A+OvN, azoospermic + oligozoospermic versus normospermic; D, minor allele; d, major allele.

Table VII Haplogroups significantly associated with azoospermia, severe oligozoospermia or moderate oligozoospermia including alternate comparisons

Design category	Test	First marker	Chromosome	Gene	Haplotype	EM cases freq.	EM controls freq.	χ^2 P A/SO/N samples	χ^2 P A/O/N samples
M	AvO+N	rs10246939*	7	TAS2R38	AA	0.66	0.50	4.26×10^{-4}	3.05×10^{-4}
S	OvN	rs323343*	8	TEX15	AAAGG	0.08	0.03	6.59×10^{-3}	8.98×10^{-3}
S	A+OvN	rs323343*	8	TEX15	ACGGG	0.08	0.14	8.34×10^{-3}	1.83×10^{-3}
S	AvO+N	rs34904070*	16	TNP2	GAACC	0.30	0.21	1.59×10^{-2}	2.38×10^{-2}
S	AvO+N	rs2030259	2	JMJD1A	GAA	0.51	0.61	1.60×10^{-2}	4.21×10^{-2}
M	OvA+N	rs4849179*	2	PAX8	AAG	0.01	0.03	1.92×10^{-2}	7.40×10^{-2}
S	AvO+N	rs2030259	2	JMJD1A	AAA	0.31	0.23	2.60×10^{-2}	4.58×10^{-2}
M	OvA+N	rs2863242*	2	PAX8	AG	0.01	0.04	3.21×10^{-2}	1.04×10^{-1}
S	AvN	rs323344*	8	TEX15	CGGG	0.08	0.15	4.66×10^{-2}	4.66×10^{-2}

Design category: selection of SNPs for study was based on previous microarray results (M), SNPs with previously published associations with infertility (P) or non-synonymous SNPs in spermatogenesis genes (S). Test: describes the comparison made for association testing as well as the genetic model applied for association testing. A, azoospermic; SO, severe oligozoospermic; O, oligozoospermic and severe oligozoospermic combined; N, normospermic; AvN, azoospermic versus normospermic; OvN, oligozoospermic versus normospermic, A+OvN, azoospermic + oligozoospermic versus normospermic; AvO+N, azoospermic versus oligozoospermic + normospermic; OvA+N, oligospermic versus azoospermic + normospermic; *, denotes haplogroups presented in Table VI.

It was recently reported that BRDT binds specifically to histone H4 tails with at least two acetylation marks (Moriniere *et al.*, 2009). The importance of this protein in spermatogenesis makes it an intriguing candidate for disruption of spermatogenesis.

JMJD1A is a histone demethylase that specifically demethylates mono- and di-methylated histone H3 lysine 9. Knockout of *Jmjd1a* in mice results in reduced testis weight and severe oligozoospermia resulting in infertility (Liu *et al.*, 2010). It was also found that *Jmjd1a* knockout resulted in reduced expression of a number of genes important in chromatin remodeling and spermatid elongation (Liu *et al.*, 2010). As with the other SNPs identified in this study, the functional significance of rs34605051 is not yet known.

As with the majority of associations reported for other GWAS, the involvement of the associated SNPs identified in this study is not immediately clear. The variants identified may not be the actual functional variants, but may be in linkage disequilibrium with the functional variant, as has been determined for the majority of SNPs identified by GWAS (Lettre and Rioux, 2008). Alternatively, the SNPs identified may affect uncharacterized functional elements within the genome.

In a few cases, haplogroup analysis detected significant associations in genes that were not found when evaluating individual SNPs such as haplogroups in paired box 8 (PAX8) and transition protein 2 (TNP2), however, in most cases no statistical power was gained by evaluating haplogroups as compared with individual SNPs. Although our data indicate that cases were significantly more likely than controls to contain multiple independent risk SNPs, much larger studies will be necessary to accurately characterize the combined effects of multiple independent loci on spermatogenic defects.

In an effort to better characterize the genetic basis for different categories of spermatogenic failure, we performed a number of different comparisons for association testing. In the majority of published male infertility SNP association studies the primary comparison

is normospermic or known fertile controls compared with infertile men. The infertile group sometimes includes only non-obstructive azoospermic men or only severely oligozoospermic men, but often azoospermic and oligozoospermic groups are combined. In combining azoospermic and oligozoospermic samples for association testing, the assumption is that the same SNPs contribute to both pathologies. We performed these same comparisons (A versus N, O versus N and A+O versus N), however, we also evaluated the strength of associations with alternate comparisons (A versus O+N and O versus A+N). In doing so, we test the assumption that azoospermia and oligozoospermia are disorders with completely separate genetic backgrounds. There is not a compelling amount of evidence to support or refute either of these assumptions, so we report here the significant associations for all of the comparisons made.

A striking result of the current study is the failure of this study to validate essentially all of the previously published SNPs associated with azoospermia or oligozoospermia. Due to limitations of the assay, we were not able to design probes to genotype all of the previously published male infertility SNPs, however, we did perform association testing on 17 previously reported SNPs and found a significant association for only one of those SNPs in our study group. In addition, we performed follow-up association testing for 70 SNPs with strong or marginal associations with azoospermia or oligozoospermia based on our pilot microarray study, and of those SNPs, associations were strengthened for only four.

There are a number of possible reasons for this low rate of validation. Our pilot and follow-up studies both included Caucasians of European decent; however, many of the previously published re-sequencing studies included men of other ethnicities (Table II). SNP and haplotype frequencies vary widely between ethnic groups. It is, unfortunately, not uncommon for associations reported in one study to fail validation in subsequent studies as a result of insufficient study power, genetic homogeneity between or within studies, or because the initially reported association was spurious (Krausz and Giachini, 2007; Liu *et al*., 2008). This again emphasizes the need for rigorous follow-up of reported associations in new patient groups and by means of meta-analyses to achieve additional power when possible (Tuttelmann *et al*., 2007; Nuti and Krausz, 2008).

On the basis of these findings, it is likely that male factor infertility is similar in nature to the majority of complex diseases studied to date in that the disease is multigenic and no single SNP is responsible for an appreciable proportion of male factor infertility cases (Manolio *et al*., 2009). It seems unlikely that a few SNPs of moderate or large effect are responsible for spermatogenic failure, rather that a large number of rare variants of small effect are responsible, as has been proposed for other complex diseases (Manolio *et al*., 2009). Given the wide phenotypic spectrum of male infertility or even non-obstructive azoospermia, it is not surprising that SNPs with strong effect have not been identified. In order to detect real associations for such small effect SNPs, much larger genome-wide studies will be necessary.

To date, genome-wide studies have successfully identified hundreds of genetic variants associated with over 80 different diseases or traits (Hindorff *et al*., 2009). Mounting evidence based on larger and larger genome-wide studies suggests genotypes for tens or even hundreds of thousands of cases and controls will be required to capture an appreciable proportion of the loci responsible for the heritable component of many common complex diseases (Panoutsopoulou and Zeggini, 2009). For example, the genetic component of type 2 diabetes (T2D) has been studied extensively over the past few decades. Early gene re-sequencing efforts and genome-wide linkage scans successfully identified four important loci associated with T2D over a span of almost 30 years (McCarthy and Zeggini, 2009). In contrast, the first genome-wide SNP association study of T2D was published in 2007 (Sladek *et al*., 2007), and in <2 years a total of 15 new risk variants have been identified (McCarthy and Zeggini, 2009).

Although this illustrates the incredible power to identify risk alleles by genome-wide genotyping, a total of six studies cumulatively genotyping over 7000 cases and 13 000 controls were employed (McCarthy and Zeggini, 2009). Even with the large effort made to understand the genetic component of T2D, the variants identified to date still only account for about 6% of T2D heritability (Manolio *et al*., 2009).

Similar results have been obtained for other complex diseases including neurological disorders (Simon-Sanchez and Singleton, 2008; Bertram and Tanzi, 2009), autoimmune diseases (Lettre and Rioux, 2008; Graham *et al*., 2009), cardiovascular disease (Arking and Chakravarti, 2009) and a number of different cancers (Savas and Liu, 2009) following multiple large genome-wide studies.

Although these data represent a promising and potentially important group of associations, clearly much larger studies will be necessary to validate these findings and identify new functional variants associated with male infertility. It remains to be seen whether studies evaluating hundreds or a few thousand infertile men will be sufficient to identify risk variants with confidence or whether studies similar in magnitude to the previously discussed diseases will be required, but it is clear that studies need to be expanded significantly if we want to obtain answers. In addition structural variants such as duplications or deletions in the genome have been found to be important risk factors in a number of complex diseases (Frazer *et al*., 2009; Wain *et al*., 2009). In the case of male infertility, gr/gr deletion of the Y chromosome (which removes half of the AZFc genes) has been confirmed as a significant risk factor for impaired spermatogenesis (Repping *et al*., 2003; Giachini *et al*., 2008; Visser *et al*., 2009). Other types of variation may also prove to be important in some cases of male factor infertility and should be evaluated in future studies.

In summary, this is the first follow-up study of a small, pilot genome-wide association study of azoospermia and oligozoospermia. We have further evaluated the SNPs identified in the GWAS, along with SNPs identified from re-sequencing studies and non-synonymous SNPs from spermatogenesis genes, in Caucasian men of European descent and have identified several SNPs of potential relevance to oligozoospermia and azoospermia. Nevertheless, the study also highlights the need for future large-scale genome-wide association studies with increased statistical power, the need for structural variation studies to identify relevant copy number variations, and the need for genome sequencing of individuals to identify rare variants that are likely to be responsible for a significant proportion of spermatogenic defects. Such studies are becoming technologically practical, but will require profound improvements in collaboration and funding.

Funding

This work has been partially supported by a grant from the Italian Telethon Grant GGP08204, to C.K. and from the Spanish Health Ministry (FIS-08/1138) to C.K. and E.R.-C.

References

A ZC, Yang Y, Zhang SZ, Li N, Zhang W. Single nucleotide polymorphism C677T in the methylenetetrahydrofolate reductase gene might be a genetic risk factor for infertility for Chinese men with azoospermia or severe oligozoospermia. *Asian J Androl* 2007;**9**:57–62.

Arking DE, Chakravarti A. Understanding cardiovascular disease through the lens of genome-wide association studies. *Trends Genet* 2009;**25**: 387–394.

Aschim EL, Giwercman A, Stahl O, Eberhard J, Cwikiel M, Nordenskjold A, Haugen TB, Grotmol T, Giwercman YL. The RsaI polymorphism in the estrogen receptor-beta gene is associated with male infertility. *J Clin Endocrinol Metab* 2005;**90**:5343–5348.

Aston KI, Carrell DT. Genome-wide study of single-nucleotide polymorphisms associated with azoospermia and severe oligozoospermia. *J Androl* 2009;**30**:711–725.

Bentz EK, Hefler LA, Denschlag D, Pietrowski D, Buerkle B, Tempfer CB. A polymorphism of the interleukin-1 beta gene is associated with sperm pathology in humans. *Fertil Steril* 2007;**88**:751–753.

Bertram L, Tanzi RE. Genome-wide association studies in Alzheimer's disease. *Hum Mol Genet* 2009;**18**:R137–R145.

Bezold G, Lange M, Peter RU. Homozygous methylenetetrahydrofolate reductase C677T mutation and male infertility. *N Engl J Med* 2001; **344**:1172–1173.

Bult CJ, Eppig JT, Kadin JA, Richardson JE, Blake JA. The Mouse Genome Database (MGD): mouse biology and model systems. *Nucleic Acids Res* 2008;**36**:D724–D728.

Christensen GL, Wooding SP, Ivanov IP, Atkins JF, Carrell DT. Sequencing and haplotype analysis of the activator of CREM in the testis (ACT) gene in populations of fertile and infertile males. *Mol Hum Reprod* 2006; **12**:257–262.

Dhillon VS, Shahid M, Husain SA. Associations of MTHFR DNMT3b 4977bp deletion in mtDNA and GSTM1 deletion, and aberrant CpG island hypermethylation of GSTM1 in non-obstructive infertility in Indian men. *Mol Hum Reprod* 2007;**13**:213–222.

Dohle GR, Halley DJ, Van Hemel JO, van den Ouwel AM, Pieters MH, Weber RF, Govaerts LC. Genetic risk factors in infertile men with severe oligozoospermia and azoospermia. *Hum Reprod* 2002;**17**:13–16.

Easton DF, Eeles RA. Genome-wide association studies in cancer. *Hum Mol Genet* 2008;**17**:R109–R115.

Ebisch IM, van Heerde WL, Thomas CM, van der Put N, Wong WY, Steegers-Theunissen RP. C677T methylenetetrahydrofolate reductase polymorphism interferes with the effects of folic acid and zinc sulfate on sperm concentration. *Fertil Steril* 2003;**80**:1190–1194.

Excoffier L, Slatkin M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol* 1995; **12**:921–927.

Frazer KA, Murray SS, Schork NJ, Topol EJ. Human genetic variation and its contribution to complex traits. *Nat Rev Genet* 2009;**10**:241–251.

Galan JJ, Buch B, Cruz N, Segura A, Moron FJ, Bassas L, Martinez-Pineiro L, Real LM, Ruiz A. 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**:910–918.

Galan JJ, De Felici M, Buch B, Rivero MC, Segura A, Royo JL, Cruz N, Real LM, Ruiz A. Association of genetic markers within the KIT and KITLG genes with human male infertility. *Hum Reprod* 2006;**21**: 3185–3192.

Gazquez C, Oriola J, de Mateo S, Vidal-Taboada JM, Ballesca JL, Oliva R. A common protamine 1 promoter polymorphism (−190 C- > A) correlates with abnormal sperm morphology and increased protamine P1/P2 ratio in infertile patients. *J Androl* 2008;**29**:540–548.

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.

Graham RR, Hom G, Ortmann W, Behrens TW. Review of recent genome-wide association scans in lupus. *J Intern Med* 2009;**265**: 680–688.

Hammoud S, Emery BR, Dunn D, Weiss RB, Carrell DT. Sequence alterations in the YBX2 gene are associated with male factor infertility. *Fertil Steril* 2009;**91**:1090–1095.

Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA* 2009;**106**:9362–9367.

Ji G, Gu A, Xia Y, Lu C, Liang J, Wang S, Ma J, Peng Y, Wang X. ERCC1 and ERCC2 polymorphisms and risk of idiopathic azoospermia in a Chinese population. *Reprod Biomed Online* 2008;**17**:36–41.

Krausz C, Giachini C. Genetic risk factors in male infertility. *Arch Androl* 2007;**53**:125–133.

Lee HC, Jeong YM, Lee SH, Cha KY, Song SH, Kim NK, Lee KW, Lee S. Association study of four polymorphisms in three folate-related enzyme genes with non-obstructive male infertility. *Hum Reprod* 2006; **21**:3162–3170.

Lettre G, Rioux JD. Autoimmune diseases: insights from genome-wide association studies. *Hum Mol Genet* 2008;**17**:R116–R121.

Liu YJ, Papasian CJ, Liu JF, Hamilton J, Deng HW. Is replication the gold standard for validating genome-wide association findings? *PLoS One* 2008;**3**:e4037.

Liu Z, Zhou S, Liao L, Chen X, Meistrich M, Xu J. The Jmjd1a demethylase-regulated histone modification is essential for crem-regulated gene expression and spermatogenesis. *J Biol Chem* 2010;**285**:2758–2770.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A *et al.* Finding the missing heritability of complex diseases. *Nature* 2009;**461**:747–753.

Mathew CG. New links to the pathogenesis of Crohn disease provided by genome-wide association scans. *Nat Rev Genet* 2008;**9**:9–14.

Matzuk MM, Lamb DJ. The biology of infertility: research advances and clinical challenges. *Nat Med* 2008;**14**:1197–1213.

McCarthy MI, Zeggini E. Genome-wide association studies in type 2 diabetes. *Curr Diab Rep* 2009;**9**:164–171.

Merisalu A, Punab M, Altmae S, Haller K, Tiido T, Peters M, Salumets A. The contribution of genetic variations of aryl hydrocarbon receptor pathway genes to male factor infertility. *Fertil Steril* 2007;**88**:854–859.

Moriniere J, Rousseaux S, Steuerwald U, Soler-Lopez M, Curtet S, Vitte AL, Govin J, Gaucher J, Sadoul K, Hart DJ *et al.* Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. *Nature* 2009;**461**:664–668.

Nuti F, Krausz C. Gene polymorphisms/mutations relevant to abnormal spermatogenesis. *Reprod Biomed Online* 2008;**16**:504–513.

Oonk RB, Grootegoed JA. Identification of insulin receptors on rat Sertoli cells. *Mol Cell Endocrinol* 1987;**49**:51–62.

Panoutsopoulou K, Zeggini E. Finding common susceptibility variants for complex disease: past, present and future. *Brief Funct Genomic Proteomic* 2009;**8**:345–352.

Park JH, Lee HC, Jeong YM, Chung TG, Kim HJ, Kim NK, Lee SH, Lee S. MTHFR C677T polymorphism associates with unexplained infertile male factors. *J Assist Reprod Genet* 2005;**22**:361–368.

Pivot-Pajot C, Caron C, Govin J, Vion A, Rousseaux S, Khochbin S. Acetylation-dependent chromatin reorganization by BRDT, a testis-specific bromodomain-containing protein. *Mol Cell Biol* 2003; **23**:5354–5365.

Repping S, Skaletsky H, Brown L, van Daalen SK, Korver CM, Pyntikova T, Kuroda-Kawaguchi T, de Vries JW, Oates RD, Silber S *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.

Savas S, Liu G. Genetic variations as cancer prognostic markers: review and update. *Hum Mutat* 2009;**30**:1369–1377.

Simon-Sanchez J, Singleton A. Genome-wide association studies in neurological disorders. *Lancet Neurol* 2008;**7**:1067–1072.

Singh K, Singh SK, Sah R, Singh I, Raman R. Mutation C677T in the methylenetetrahydrofolate reductase gene is associated with male infertility in an Indian population. *Int J Androl* 2005;**28**:115–119.

Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S *et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 2007;**445**: 881–885.

Stuppia L, Gatta V, Scarciolla O, Colosimo A, Guanciali-Franchi P, Calabrese G, Palka G. The methylenetethrahydrofolate reductase (MTHFR) C677T polymorphism and male infertility in Italy. *J Endocrinol Invest* 2003;**26**:620–622.

Su D, Zhang W, Yang Y, Deng Y, Ma Y, Song H, Zhang S. Mutation screening and association study of the TSSK4 Gene in Chinese infertile men with impaired spermatogenesis. *J Androl* 2008;**29**:374–378.

Suryavathi V, Khattri A, Gopal K, Rani DS, Panneerdoss S, Gupta NJ, Chakravarty B, Deenadayal M, Singh L, Thangaraj K. Novel variants in UBE2B gene and idiopathic male infertility. *J Androl* 2008;**29**:564–571.

Suzuki Y, Sasagawa I, Itoh K, Ashida J, Muroya K, Ogata T. Estrogen receptor alpha gene polymorphism is associated with idiopathic azoospermia. *Fertil Steril* 2002;**78**:1341–1343.

Thomas MB, Haines SL, Akeson RA. Chemoreceptors expressed in taste, olfactory and male reproductive tissues. *Gene* 1996;**178**:1–5.

Tuttelmann 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.

Visser L, Westerveld GH, Korver CM, van Daalen SK, Hovingh SE, Rozen S, van der Veen F, Repping S. Y chromosome gr/gr deletions are a risk factor for low semen quality. *Hum Reprod* 2009;**24**:2667–2673.

Wain LV, Armour JA, Tobin MD. Genomic copy number variation, human health, and disease. *Lancet* 2009;**374**:340–350.

Wang W, Lu N, Xia Y, Gu A, Wu B, Liang J, Zhang W, Wang Z, Su J, Wang X. FAS and FASLG polymorphisms and susceptibility to idiopathic azoospermia or severe oligozoospermia. *Reprod Biomed Online* 2009;**18**:141–147.

Yang F, Eckardt S, Leu NA, McLaughlin KJ, Wang PJ. Mouse TEX15 is essential for DNA double-strand break repair and chromosomal synapsis during male meiosis. *J Cell Biol* 2008;**180**:673–679.

Zhang W, Zhang S, Xiao C, Yang Y, Zhoucun A. Mutation screening of the FKBP6 gene and its association study with spermatogenic impairment in idiopathic infertile men. *Reproduction* 2007;**133**:511–516.

Zhang Z, Xue H, Gong W, Wang M, Yuan L, Han S. FAS promoter polymorphisms and cancer risk: a meta-analysis based on 34 case–control studies. *Carcinogenesis* 2009;**30**:487–493.

Zhoucun A, Zhang S, Yang Y, Ma Y, Zhang W, Lin L. The common variant N372H in BRCA2 gene may be associated with idiopathic male infertility with azoospermia or severe oligozoospermia. *Eur J Obstet Gynecol Reprod Biol* 2006;**124**:61–64.