human reproduction

ORIGINAL ARTICLE *Andrology*

Aberrant DNA methylation patterns of spermatozoa in men with unexplained infertility

Rocío G. Urdinguio[1,†], Gustavo F. Bayón[1,†], Marija Dmitrijeva[1], Estela G. Toraño[1], Cristina Bravo[1], Mario F. Fraga[1,2], Lluís Bassas[3], Sara Larriba[4,*], and Agustín F. Fernández[1,*]

[1]Cancer Epigenetics Laboratory, Institute of Oncology of Asturias (IUOPA), HUCA, Universidad de Oviedo, Oviedo 33006, Spain [2]Department of Immunology and Oncology, National Center for Biotechnology, CNB-CSIC, Cantoblanco, Madrid 28049, Spain [3]Laboratory of Seminology and Embryology, Andrology Service—Fundació Puigvert, Barcelona 08025, Spain [4]Human Molecular Genetics Group—IDIBELL, L'Hospitalet de Llobregat, Barcelona 08908, Spain

*Correspondence address. E-mail: affernandez@hca.es (A.F.F)/slarriba@idibell.cat (S.L)

STUDY QUESTION: Are there DNA methylation alterations in sperm that could explain the reduced biological fertility of male partners from couples with unexplained infertility?

SUMMARY ANSWER: DNA methylation patterns, not only at specific loci but also at Alu Yb8 repetitive sequences, are altered in infertile individuals compared with fertile controls.

WHAT IS KNOWN ALREADY: Aberrant DNA methylation of sperm has been associated with human male infertility in patients demonstrating either deficiencies in the process of spermatogenesis or low semen quality.

STUDY DESIGN, SIZE, DURATION: Case and control prospective study. This study compares 46 sperm samples obtained from 17 normospermic fertile men and 29 normospermic infertile patients.

PARTICIPANTS/MATERIALS, SETTING, METHODS: Illumina Infinium HD Human Methylation 450K arrays were used to identify genomic regions showing differences in sperm DNA methylation patterns between five fertile and seven infertile individuals. Additionally, global DNA methylation of sperm was measured using the Methylamp Global DNA Methylation Quantification Ultra kit (Epigentek) in 14 samples, and DNA methylation at several repetitive sequences (LINE-1, Alu Yb8, NBL2, D4Z4) measured by bisulfite pyrosequencing in 44 sperm samples. A sperm-specific DNA methylation pattern was obtained by comparing the sperm methylomes with the DNA methylomes of differentiated somatic cells using data obtained from methylation arrays (Illumina 450 K) of blood, neural and glial cells deposited in public databases.

MAIN RESULTS AND THE ROLE OF CHANCE: In this study we conduct, for the first time, a genome-wide study to identify alterations of sperm DNA methylation in individuals with unexplained infertility that may account for the differences in their biological fertility compared with fertile individuals. We have identified 2752 CpGs showing aberrant DNA methylation patterns, and more importantly, these differentially methylated CpGs were significantly associated with CpG sites which are specifically methylated in sperm when compared with somatic cells. We also found statistically significant ($P < 0.001$) associations between DNA hypomethylation and regions corresponding to those which, in somatic cells, are enriched in the repressive histone mark H3K9me3, and between DNA hypermethylation and regions enriched in H3K4me1 and CTCF, suggesting that the relationship between chromatin context and aberrant DNA methylation of sperm in infertile men could be locus-dependent. Finally, we also show that DNA methylation patterns, not only at specific loci but also at several repetitive sequences (LINE-1, Alu Yb8, NBL2, D4Z4), were lower in sperm than in somatic cells. Interestingly, sperm samples at Alu Yb8 repetitive sequences of infertile patients showed significantly lower DNA methylation levels than controls.

LIMITATIONS, REASONS FOR CAUTION: Our results are descriptive and further studies would be needed to elucidate the functional effects of aberrant DNA methylation on male fertility.

[†] These authors contributed equally to this work.

Introduction

Human infertility is a disorder affecting 13−15% of couples worldwide, where in 20% of cases the male factor is fully responsible and in another 30−40% it is contributory (Jarow et al., 2002). In the majority of cases, male factor infertility is closely related to decreased semen quality as a consequence of urogenital abnormalities including testicular dysfunction (a large proportion being caused by genetic abnormalities such as karyotype anomalies and Y chromosome microdeletions), varicocele, infections of the genital tract, immunological problems and/or exposure to exogenous chemical or physical agents. However, in some cases the clinical evaluation of the individual is normal, thus the patient is diagnosed with infertility of unknown origin (Irvine, 1998; Hamada et al., 2011).

Epigenetics involves the study of heritable changes affecting gene expression that are not caused by any change in DNA sequence (Holliday, 1987). The best-known epigenetic mark is DNA methylation (Esteller, 2008; Fernandez et al., 2012), a dynamic process that takes place throughout the course of development in multicellular organisms and ensures the maintenance of normal expression patterns. Likewise, DNA methylation alterations related to different human pathologies, developmental processes and ageing have been found (Urdinguio et al., 2009; Fernandez et al., 2012).

In particular, DNA methylation of germ cells is critically involved in many processes, including paternal genomic imprinting (Feinberg et al., 2002), the gene-dosage reduction involved in X chromosome inactivation in females (Payer and Lee, 2008), the silencing of transposable elements (Doerfler, 1991) and several aspects of meiosis, post-meiotic gene silencing and DNA compaction (Oakes et al., 2007a).

The mammalian germ line undergoes extensive epigenetic reprogramming during germ cell maturation and gametogenesis. In males, widespread erasure of DNA methylation takes place in primordial germ cells (Reik et al., 2001; Hajkova et al., 2002) and subsequent de novo DNA methylation occurs during the maturation of germ cells and spermatogenesis, prior to meiosis (Oakes et al., 2007a). As a result, the pattern of sperm DNA methylation is unique and hypomethylated compared with any other somatic cell (Eckhardt et al., 2006; Oakes et al., 2007b). The main targets of methylation in germ cells are non-CpG island (non-CGI) sequences in both distinct loci and repetitive sequences, but CpG islands (CGIs) can also be methylated (Oakes et al., 2007a). Interestingly, hypomethylated promoters in the mature sperm are the promoters of developmental transcription and signalling factors. In mammals, correct sperm DNA methylation is suggested to be essential for both fertilization and early embryo viability (Li et al., 1992; Okano et al., 1999; Bourc'his and Bestor, 2004; Anway et al., 2005; Yaman and Grandjean, 2006; Carrell and Hammoud, 2010; Romero et al., 2011; Dada et al., 2012; Jenkins and Carrell, 2012) and therefore improved knowledge of the epigenetics of sperm is not only necessary to understand these processes, but may also provide clues to the potential causes of male infertility of unknown origin.

Early studies of sperm DNA methylation analysis were specifically performed in imprinted genes as an increased risk of congenital imprinting diseases in children conceived through assisted reproductive technologies (ART) had been suggested. These studies showed the aberrant methylation patterns of imprinted genes in poor quality sperm (Marques et al., 2004; Kobayashi et al., 2007; Marques et al., 2008; Poplinski et al., 2010). Altered sperm DNA methylation patterns have also been found in non-imprinted genes associated with spermatogenic impairment, such as methylenetetrahydrofolate reductase (MTHFR) (associated with folate metabolism and methylation reactions; Wu et al., 2010b), the cAMP responsive element modulator (CREM), involved in spermatogenesis, (Nanassy and Carrell, 2011) and the Deleted in Azoospermia-like (DAZL) gene which is involved in germ line establishment and gametogenesis (Navarro-Costa et al., 2010).

The emergence of new technologies to analyse DNA methylation has allowed the study of alterations at the whole-genome level. In this manner, aberrant sperm DNA methylation of both imprinted and non-imprinted genes has been identified in infertile individuals with poor quality semen (Houshdaran et al., 2007; Pacheco et al., 2011). Furthermore, nearly 600 genes were found differentially methylated in the testes of humans with spermatogenic disorders, of which two are worth noting; Piwi-like RNA-mediated gene silencing 2 (PIWIL2) and Tudor domain containing 1 (TDRD1), two germ line-specific genes involved in PIWI-interacting RNA (piRNA) processing machinery (Heyn et al., 2012). The results of all these studies suggest that alterations of DNA methylation patterns of the germ cell could affect reproductive success.

Although defective germ cell-DNA methylation patterns have been associated with alterations in semen quality, there is a lack of such epigenetic studies in infertile men with normal sperm parameters. In this project we aim for the first time to conduct a genome-wide analysis of DNA methylation in sperm samples from normozoospermic fertile and infertile men. To address this issue, we used high-throughput 450K methylation arrays, covering the whole genome, to determine whether epigenetic changes in male germ cells could explain differences in reproductive success related to the functional quality of spermatozoa.

Materials and Methods

Ethical approval

Our study recruited semen samples from 49 selected individuals of Caucasian origin. All the participants signed an informed consent form and the project was approved by the Ethical Committee of the Puigvert Foundation.

Subjects of study

Clinical assessment of fertile and infertile individuals was conducted at the Andrology Service of the Fundació Puigvert and included taking a full personal and family medical history to rule out heritable conditions, physical examination—with special emphasis on sexual characters, gonads and genitalia—and a minimum of two semen analyses [performed in accordance with the World Health Organization guidelines (Cooper *et al.*, 2010; WHO, 2010) except for motility assessments, which were done at room temperature using the WHO 1999 four-grade classification of movement]. Spermiograms included volume, pH, sperm concentration, four-category motility assessment, vitality, morphology and antisperm antibodies. Motility and sperm count were done in duplicate aliquots of ≥ 200 cells, and measures were adopted to control for acceptable differences between duplicates. Sperm concentration was performed on diluted, immobilized samples using haemocytometer chambers. Computer-assisted sperm analysis (CASA) was performed on fresh ejaculates with the Integrated Semen Analysis System (ISAS®, version 1.01), Proiser R+D (Valencia, Spain) to obtain objective measurements of sperm kinematics (Pedigo *et al.*, 1989).

Semen samples from 17 fertile men (2 of whom were anonymous donors used in >10 insemination cycles per sperm donor for at least 6 female recipients; the other 15 being volunteers of proven fertility who were going to undergo vasectomy; aged 22–49 years) with normal seminal quality were studied as methylation controls of fertile spermatozoa. In addition, semen samples were obtained from 29 male patients (aged 30–55 years) consulting for couple infertility with no known risk factors and normal or mild defects of semen quality values that were used in husband ART (IVF-ICSI: *in vitro* fertilization- intracytoplasmic sperm injection technique) (Table I). The reference values of fertility were those most recently defined by the WHO (Cooper *et al.*, 2010; WHO, 2010).

Semen volume, count, motility and morphology, including the teratozoospermia index (TZI), as well as the results of the gynaecological assessment of the female partner, for the samples used for the epigenetic studies are summarized in Table I. TZI is defined as the number of abnormalities present per abnormal spermatozoon, these being defects of the head, neck/mid piece and tail defects or presence of cytoplasmic droplets, thus indicating the severity of the morphology alteration (from 1 to 4 value).

Pregnancy outcome from IVF-ICSI treatment of infertile couples is also described for patients (Table I).

Isolation of mature germ cells and DNA extraction

Semen samples were liquefacted and homogenized with a mechanical mixer at 37°C (30–60 min) and subsequently processed with a differential centrifugation technique using density gradients (65–90% Puresperm®, Nidacon International AB, Mölndal, Sweden) to remove somatic contaminants and to enrich the sample in terms of spermatozoa. The isolated germ cells were normalized to a concentration of 1×10^6/ml and processed to obtain sperm DNA (Wizard Genomic Purification kit, Promega, USA) following the manufacturer's instructions for this specific cell type. The DNA extraction included RNAse A treatment of DNA samples.

Genome-wide DNA methylation analysis with high-density arrays

Microarray-based DNA methylation profiling was performed with the Illumina Infinium® Human Methylation450 BeadChip (Illumina Inc., USA) (Bibikova *et al.*, 2011). Bisulfite conversion of DNA was carried out using the EZ DNA Methylation Kit (Zymo Research, Orange, CA) following the manufacturer's procedure, but with the modifications described in the Infinium Assay Methylation Protocol Guide. Processed DNA samples were then hybridized to the BeadChip (Illumina) following the Illumina Infinium HD Methylation Protocol. Genotyping services were provided by the Spanish 'Centro Nacional de Genotipado (CEGEN-ISCIII)' (www.cegen.org).

IDAT files from the microarray were further processed using the R/Bioconductor (version 3.0; open source) package *minfi* (Hansen and Aryee). In order to adjust for the different probe design types present in the 450k architecture, red and green signals from the IDAT files were corrected using the SWAN algorithm (Makismovic *et al.*, 2012). Probes with detection *P*-values over 0.01 in at least two samples were filtered out. In accordance with Du (Du *et al.*, 2010), both beta values and *M*-values were computed and employed across the analysis pipeline. *M*-values were used for all the statistical analyses, assuming homoscedasticity, while beta values were mostly used for the intuitive interpretation and visualization of results.

Probes that were found to co-hybridate with probes in the sexual chromosomes (Chen *et al.*, 2013) were removed.

Detection of differentially methylated probes

In order to identify CpG sites which were differentially methylated (dmCpGs) between the sperm samples of fertile and infertile men, a robust moderated *t*-test implemented in the R/Bioconductor package *limma* (Smyth, 2005) was performed. False discovery rate (FDR) was controlled using the Benjamini–Hochberg procedure, and a significance level of 0.05 employed. An additional threshold of effect size was applied, meaning that only those probes with the strongest differences between groups (the top 70%) were selected. The application of this threshold is essential to remove differences deriving from technical artefacts and consequently ensure a more biologically sound statistical data analysis (Pan *et al.*, 2005). dmCpGs were defined as hypermethylated or hypomethylated when methylation values were, respectively, higher or lower in infertile samples compared with fertile controls.

Genomic region analysis

The probes in the microarray were assigned to a genomic region according to their position relative to the transcript information extracted from the R/Bioconductor package TxDb.Hsapiens.UCSC.hg19.knownGene (Carlson). A probe was said to be in a *Promoter* region when it was located inside the first exon, the 5′-UTR or a region up to 2 kbp upstream of the transcription start site (TSS) of any given transcript. Similarly, a probe found inside any intron or any exon other than the first was labelled as *Intragenic*. *Intergenic* probes were determined as those which did not fall into either of the two previous categories. According to this definition, a probe could be in both a *Promoter* and an *Intragenic* region at the same time, for different transcripts. A contingency table was built for each selected subset of probes and a given genomic region, with one variable indicating whether a given probe belonged or not to the subset, and the other indicating whether a given probe was labelled with the selected region. Significance of the association was determined by a χ^2-test. A significance level of 0.05 was used to determine if a subset was dependent with respect to a given genomic region. Odds ratio (OR) was used as a measure of effect size.

CGI status analysis

The CGI locations used in the analyses were obtained from the R/Bioconductor package FDb.InfiniumMethylation.hg19 (Triche). This dataset contains

Table I Clinical characteristics of semen samples from fertile controls and infertile patients included in the study.

Code ID	Methylation assay	Clinical intervention	IVF-ICSI outcome	Age (years)	Semen volume (ml)	Sperm count ($\times 10^6$/ml)	Total sperm count ($\times 10^6$)	Progressive motility (%)	Normal morphology (%)	Teratozoospermia index (TZI)	Age female partner (years)	Female gynaecological assessment
Control-1	1,3	Vasectomy	–	41	4.5	174	783.0	46	11	1.69	–	N (by default)
Control-2	1,3	Vasectomy	–	49	4.5	122	549.0	46	11	1.78	–	N (by default)
Control-3	1,3	Vasectomy	–	35	6.0	32	192.0	23	6	1.64	–	N (by default)
Control-4	1,3	Vasectomy	–	39	1.5	96	144.0	58	10	1.21	–	N (by default)
Control-5	1,3	Vasectomy	–	36	3.5	42	147.0	42	6	1.49	–	N (by default)
Control-6	2,3	Vasectomy	–	45	2.8	89	244.8	67	10	1.52	–	N (by default)
Control-7	2,3	Vasectomy	–	39	3.3	81	267.3	32	6	1.50	–	N (by default)
Control-8	2,3	Vasectomy	–	48	5.5	89	489.5	67	10	1.56	–	N (by default)
Control-9	2,3	Vasectomy	–	39	4.0	65	260.0	69	5	1.45	–	N (by default)
Control-10	2,3	Vasectomy	–	42	5.8	29	168.2	71	5	1.56	–	N (by default)
Control-11	2,3	Vasectomy	–	44	3.0	250	750.0	60	8	1.65	–	N (by default)
Control-12	2,3	Vasectomy	–	38	4.6	70	322.0	54	16	1.33	–	N (by default)
Control-13	3	Vasectomy	–	34	2.0	29	58.0	24	0	1.44	–	N (by default)
Control-14	3	Vasectomy	–	35	4.0	75	300.0	46	8	1.39	–	N (by default)
Control-15	3	Vasectomy	–	44	1.0	120	120.0	68	9	1.45	–	N (by default)
Control-16	3	Semen donor	–	25	3.3	80	264.0	58	24	1.30	35.2[a]	–
Control-17	3	Semen donor	–	22	4.8	78	370.5	39	14	1.32	34.1[a]	–
Patient-1	1	IVF-ICSI	No pregnancy	42	2.0	36	72.0	50	1	1.58	41	N
Patient-2	1	IVF-ICSI	Pregnancy-birth (twin)	37	4.8	43	204.3	42	9	1.51	37	N
Patient-3	1,3	IVF-ICSI	No pregnancy	40	6.1	10	61.0	37	2	1.49	39	N
Patient-4	1,3	IVF-ICSI	No pregnancy	36	6.6	18	118.8	52	1	1.63	35	N
Patient-5	1,3	IVF-ICSI	No pregnancy	36	5.5	91	500.5	48	5	1.67	37	N
Patient-6	1,3	IVF-ICSI	Pregnancy-birth	54	3.7	25	92.5	18	1	1.57	38	N
Patient-7	1,3	IVF-ICSI	Pregnancy-miscarriage	39	3.9	17	66.3	46	1	1.59	35	Treated dysovulation
Patient-8	2,3	IVF-ICSI	No pregnancy	45	5.0	95	475.0	40	7	1.52	40	N
Patient-9	2,3	IVF-ICSI	No pregnancy	31	2.5	98	245.0	62	7	1.43	26	Tubal obstruction
Patient-10	2,3	IVF-ICSI	No pregnancy	33	3.0	123	369.0	69	7	1.48	35	N
Patient-11	2,3	IVF-ICSI	No pregnancy	37	2.5	171	427.5	71	8	1.61	36	Tubal obstruction
Patient-12	2,3	IVF-ICSI	No pregnancy	55	2.8	92	257.6	36	6	1.52	38	N
Patient-13	2,3	IVF-ICSI	No pregnancy	34	3.6	203	722.7	55	6	1.46	34	N
Patient-14	2,3	IVF-ICSI	No pregnancy	36	4.5	61	274.5	53	8	1.57	37	N
Patient-15	3	IVF-ICSI	Pregnancy-birth	45	2.5	56	140.0	50	6	1.46	39	Mild endometriosis

Continued

Table I *Continued*

Code ID	Methylation assay	Clinical intervention	IVF-ICSI outcome	Age (years)	Semen volume (ml)	Sperm count (×10⁶/ml)	Total sperm count (×10⁶)	Progressive motility (%)	Normal morphology (%)	Teratozoospermia index (TZI)	Age female partner (years)	Female gynaecological assessment
Patient-16	3	IVF-ICSI	Pregnancy-birth (twin)	39	4.0	90	360.0	60	4	1.60	37	N
Patient-17	3	IVF-ICSI	Pregnancy-birth	39	2.8	206	576.8	40	15	1.32	35	N
Patient-18	3	IVF-ICSI	Pregnancy-birth (twin)	36	2.8	105	288.8	58	15	1.63	33	N
Patient-19	3	IVF-ICSI	Pregnancy-miscarriage	40	1.8	75	131.3	51	12	1.33	35	Mild endometriosis
Patient-20	3	IVF-ICSI	Pregnancy-miscarriage	37	3.5	202	707.0	61	12	1.44	38	Mild endometriosis
Patient-21	3	IVF-ICSI	Pregnancy-miscarriage	36	1.8	42	75.6	59	6	1.34	40	N
Patient-22	3	IVF-ICSI	Pregnancy-birth	32	4.8	196	931.0	48	14	1.35	27	Tubal obstruction
Patient-23	3	IVF-ICSI	No pregnancy	30	5.3	75	393.8	50	9	1.45	38	Tubal obstruction
Patient-24	3	IVF-ICSI	Pregnancy-birth (twin)	37	2.5	87	217.5	54	11	1.54	38	Mild endometriosis
Patient-25	3	IVF-ICSI	Pregnancy-miscarriage	44	4.9	119	583.1	60	12	1.42	39	Dysovulation
Patient-26	3	IVF-ICSI	Pregnancy-miscarriage	48	1.8	164	287.0	55	7	1.62	39	Mild endometriosis
Patient-27	3	IVF-ICSI	Pregnancy-birth	35	1.7	206	342.0	58	12	1.49	32	Dysovulation
Patient-28	3	IVF-ICSI	No pregnancy	41	3.6	71	255.6	47	6	1.46	38	Dysovulation
Patient-29	3	IVF-ICSI	Pregnancy-birth	36	7.3	92	671.6	46	8	1.52	35	Tubal obstruction
Average controls	1			40.0	4.0	93.2	363	43.0	8.8	1.56		
Average patients	1			40.6	4.7	34.3	159.3	41.9	2.9	1.6		
P-value	1			*0.872*	*0.514*	*0.088*	*0.205*	*0.877*	*0.005*	*0.874*		
Average controls	2			42.1	4.1	96.1	357.4	60.0	8.6	1.5		
Average patients	2			38.7	3.4	120.4	395.9	55.1	7.0	1.5		
P-value	2			*0.354*	*0.244*	*0.474*	*0.704*	*0.516*	*0.335*	*0.956*		
Average controls	3			38.5	3.8	89.5	319.4	51.2	9.4	1.5		
Average patients	3			38.9	3.7	103.3	354.5	51.3	7.7	1.5		
P-value	3			*0.853*	*0.925*	*0.445*	*0.606*	*0.985*	*0.284*	*0.753*		

Italic entries correspond to *P*-values.

N, normal; IVF-ICSI, *in vitro* fertilization-intracytoplasmic sperm injection.

Methylation assay 1, methylation array; methylation assay 2, global DNA methylation; methylation assay 3, pyrosequencing of repetitive sequences.

[a]Age average of female recipients is indicated.

all the CGIs distributed along the whole genome. The generation procedure for these CGIs is described by Wu and colleagues (Wu et al., 2010a). CpG shores were defined as the 2 kbp regions flanking a CGI. CpG shelves were defined as the 2 kbp region, either upstream or downstream, of each CpG shore. Probes not belonging to any of the regions thus far mentioned were assigned to the special category Non-CGI. Each probe was assigned to only one category. A 4×2 contingency table was constructed for every subset of probes in order to study the association between the given subset and the different CGI categories. A Chi-squared test was used to determine if any of the categories had a significant association with the given subset. For each of the CGI status levels, a 2×2 contingency table was defined and another Chi-squared test was independently used to evaluate the association of the given subset with each status level, a significance level of 0.05 being employed for all tests. Effect size was reported as the odds ratio (OR) for each of the individual tests.

Histone enrichment analysis

In order to analyse the enrichment of a histone mark on a given subset of probes, we used the information contained in the UCSC Browser Broad Histone track from the ENCODE project (Bernstein et al., 2005, 2006; Mikkelsen et al., 2007; Guttman et al., 2010; Ernst et al., 2011). Histone peak data for every cell line and mark present in the ENCODE project were downloaded from the UCSC Browser (Supplementary Table SI). Small peaks were discarded when they were completely contained within wider peaks. For each combination of cell line and mark, a 2×2 contingency table was built to determine its association with the input subset of probes. Probes in the array were classified according to whether they belonged to the subset or not, and whether they intersected with a significant broad peak for the given combination of cell line and mark. A Fisher exact test was used to determine if the given subset of probes was significantly enriched for each combination of cell line and mark. P-values were corrected for multiple testing using FDR (using the Benjamini–Hochberg method) and a significance level of 0.05 was used to determine which probes had significant enrichment. The base-2 logarithm of the OR was used as a measure of effect size.

Global DNA methylation analysis

Global DNA methylation status was quantified using the Methylamp global DNA methylation quantification ultra kit (Epigentek, Catalog # P-1014B, USA) following the manufacturer's instructions. Briefly, 100 or 200 ng of genomic DNA is used for a 5-methylcytosine (5-mC) quantification. The methylated fraction of DNA is recognized by a 5-methylcytosine antibody and colorimetrically quantified through an ELISA-like reaction.

Bisulfite pyrosequencing

DNA methylation patterns of the repetitive sequences were analysed by bisulfite pyrosequencing. Bisulfite modification of DNA was performed with the EZ DNA Methylation-Gold kit (Zymo Research) following the manufacturer's instructions. Each sequence was amplified with previously described forward and reverse primers (Supplementary Table SII; Bollati et al., 2007; Choi et al., 2009; Martinez et al., 2012). After PCR amplification of the region of interest with the specific primers, pyrosequencing was performed using PyroMark Q24 reagents, vacuum prep workstation, equipment and software (Qiagen, Netherlands).

Identification of specific DNA methylation patterns in spermatozoa

DNA methylation data of blood (Hannum et al., 2013) and brain (neuron and glia) (Guintivano et al., 2013) samples produced with the Illumina Infinium Human Methylation450 were used for comparison with the sperm data in order to identify DNA methylation patterns specific to mature germ cells. DNA methylation beta value data were downloaded from GEO accession numbers GSE40279 and GSE41826. Methylation data for the blood dataset were adjusted for white blood cell heterogeneity using the method described in Houseman et al. (2012). In order to feed this method, we used the original 27k database of purified white blood cell subtypes included in the authors' original implementation of the algorithm.

To identify tissue-specific methylated CpG sites in a given tissue, we looked for CpGs showing mean methylation $>60\%$ in the target tissue and $<40\%$ in the other tissues. Similarly, to identify tissue-specific unmethylated CpG sites we looked for CpGs with methylation $<40\%$ in the target tissue and $>60\%$ in the other tissues.

Circos data track smoothing

In order to plot the CpG information on Circos genome-wide graphs, smoothing was applied to our data. Broad Histone peak information from UCSC was averaged by partitioning the genome into intervals of 200 kbp and assigning to each peak a score that corresponded to the average of the broad peak scores found within it. CpG locations were not smoothed but rather stacked on several lines. This does not mean that any CpG has a higher score than other, but simply that the higher the stack of markers, the higher the density of CpGs in the region.

Statistical analyses

Statistical analyses were performed using R/Bioconductor (version 3.0; open source). To identify CpG sites showing differential methylation values between the sperm samples of fertile and infertile men, a robust moderated t-test implemented in the R/Bioconductor package limma was performed. FDR was controlled using the Benjamini–Hochberg procedure.

Significant associations between dmCpGs and specific genomic locations were determined by a χ^2-test. OR was used as a measure of effect size.
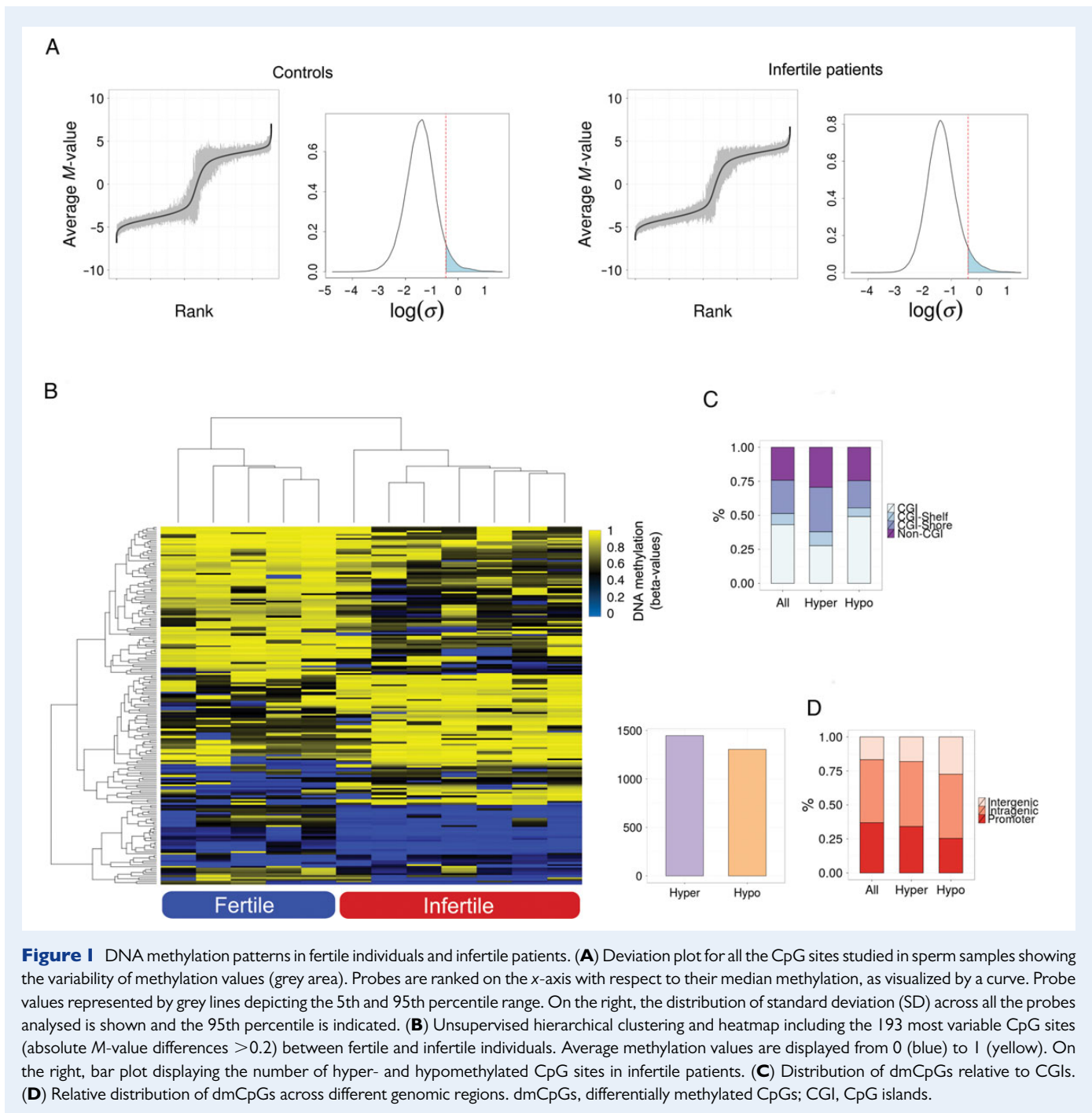
A Fisher exact test was used to analyse the enrichment of dmCpGs on an specific chromatin mark. P-values were corrected for multiple testing using FDR, and the base-2 logarithm of the OR was used as a measure of effect size.

The non-parametric Kruskal–Wallis and Wilcox tests were used to analyse differences in methylation levels (global and repetitive regions) in sperm groups compared with somatic cells. A value of $P < 0.05$ was considered significant.

Results

Alterations of sperm DNA methylation are found in infertile individuals

To identify genomic regions showing differences in sperm DNA methylation patterns between fertile and infertile individuals, we performed methylation arrays (methylation assay 1 in Table I) of 12 sperm samples and analysed the site-specific methylation status of 485 577 CpG sites across the human genome (Bibikova et al., 2011; Sandoval et al., 2011). In order to analyse differential methylation patterns, we divided the samples into two groups; one composed of five fertile individuals (samples Control-1 to 5) and the other, seven infertile patients (samples Patient-1 to 7). Only one of the female partners (Patient-7 partner) had a potential factor (dysovulation, corrected with medication) that may influence the fertility of the couple. The rest of the women presented no known risk factors. It is worth taking into account that we decided to establish semen diagnosis based exclusively on the first semen analysis, because fertile controls only delivered a single sample. Nevertheless, some of the subsequent semen samples collected for

Figure 1 DNA methylation patterns in fertile individuals and infertile patients. (**A**) Deviation plot for all the CpG sites studied in sperm samples showing the variability of methylation values (grey area). Probes are ranked on the *x*-axis with respect to their median methylation, as visualized by a curve. Probe values represented by grey lines depicting the 5th and 95th percentile range. On the right, the distribution of standard deviation (SD) across all the probes analysed is shown and the 95th percentile is indicated. (**B**) Unsupervised hierarchical clustering and heatmap including the 193 most variable CpG sites (absolute *M*-value differences >0.2) between fertile and infertile individuals. Average methylation values are displayed from 0 (blue) to 1 (yellow). On the right, bar plot displaying the number of hyper- and hypomethylated CpG sites in infertile patients. (**C**) Distribution of dmCpGs relative to CGIs. (**D**) Relative distribution of dmCpGs across different genomic regions. dmCpGs, differentially methylated CpGs; CGI, CpG islands.

this study in the infertile group showed some deviation from initial values with respect to sperm morphology. It is of note, however, that abnormal sperm from infertile individuals showed a similar low severity of morphological alteration (TZI values < 1.7) as those from fertile individuals (Table I).

The first observation indicated that, although methylation patterns are well preserved, some CpG sites exhibited higher inter-individual variability (5% showed *M*-values SD > 0.6), irrespective of the group of samples analysed (Fig. 1A), which confirms variation in DNA methylation of the male germ line across unrelated individuals (Flanagan *et al.*, 2006). Statistical analysis showed 2752 dmCpGs between fertile and infertile

men. Of these, 1447 CpG sites were hypermethylated while 1305 were hypomethylated in infertile patients (Supplementary Fig. S1 and Supplementary Table SIII). Hierarchical clustering of DNA methylation data for the most variable CpG sites highlights the differences between the fertile controls and infertile patients (Fig. 1B).

To study, from a functional genomics point of view, the characteristics of the dmCpG sites we first determined their distribution in CGI and non-CGI regions (Wu *et al.*, 2010a). Interestingly, while hypermethylated CpG sites were preferentially enriched in CGI-shores ($P < 0.001$; OR = 1.50), hypomethylated CpG sites were enriched in CGIs ($P < 0.001$, OR = 1.27) (Fig. 1C). Intergenic regions showed a significantly increased

proportion of hypomethylated CpG sites ($P < 0.001$, OR = 1.83) while promoter regions presented a decreased proportion of both hypermethylated ($P = 0.039$, OR = 0.89) and hypomethylated ($P < 0.001$, OR = 0.53) CpGs in infertile patients (Fig. 1D). Gene ontology (GO) analysis of the dmCpGs, using DAVID gene ontology annotation groups, showed that genes around the hypermethylated CpG sites were enriched for a cell adhesion related term (GO term: homophilic cell adhesion; FDR < 0.01), while genes around the hypomethylated CpG sites did not show any significant term enrichment for functional (GO) categories.

Given that the establishment of germ cell-DNA methylation marks involved in paternal genomic imprinting is critical during spermatogenesis, we next compared these dmCpGs with those associated with imprinted genes (Pacheco et al., 2011). A total of 8746 CpG sites belonging to 183 imprinted genes were present in the 450K array. We found that 54 CpG sites associated with 48 imprinted genes were aberrantly methylated in infertile patients. Specifically, 33 CpG sites (related to 28 genes) were hypermethylated and 21 CpG sites (related to 28 genes) were hypomethylated; 8 genes showed both hyper- and hypomethylation (Supplementary Tables SIV and SV). In addition we compared our results with those obtained by Pacheco and collaborators, who analysed 619 CpGs associated with imprinted genes from samples of sperm with low motility using 27K Illumina arrays, and found two common CpGs aberrantly methylated [associated with insulin-like growth factor 2 (IGF2) and heat shock 70 kDa protein 6 (HSPA6) genes] in both studies.

To study other genomic features which may provide clues about the mechanisms underlying the aberrant methylation changes in infertile men, we investigated whether the differentially methylated regions were among those targeted by any specific histone mark in somatic cells. We compared our methylation data with previously published data on a range of histone modifications and chromatin modifiers in 10 differentiated somatic cells obtained from healthy individuals (see Materials and Methods). We found statistically significant ($P < 0.001$) associations between DNA hypomethylation and the repressive histone mark H3K9me3 in most differentiated ENCODE cell lines (Fig. 2A and B). However, DNA hypermethylation was associated with H3K4me1 and the CCCTC-binding factor (CTCF) ($P < 0.001$) (Fig. 2A and B). We also investigated whether the hyper- and hypomethylated CpGs in the sperm of infertile patients were associated with the genomic regions that are known to be enriched for nucleosomes, H2AZ or the post-translational histone marks H3K4me3 and H3K27me3 in human sperm (Hammoud et al., 2009), and we found a significant association between hypermethylated CpGs in sperm of infertile patients and those genomic regions enriched for H3K4me3 ($P = 0.036$; OR: 1.40).

Differentially methylated regions are associated with sperm-specific DNA methylation sites in infertile individuals

For the further characterization of the regions associated with the dmCpG sites found in our study, we performed a comparative analysis between dmCpG sites and regions with sperm-specific DNA methylation (see Materials and Methods). To identify regions with tissue-specific DNA methylation we first compared the above-mentioned sperm methylomes of the fertile subjects (5 samples) with the DNA methylomes of differentiated somatic cells. We used data obtained from methylation analysis of blood (8 samples) and neural and glial cells (8 samples of each) from the same type of methylation arrays (Illumina

450K) deposited in public databases (Guintivano et al., 2013; Hannum et al., 2013). To reduce confounding factors in the blood dataset, we first corrected for cellular heterogeneity (Houseman et al., 2012).

The results of these comparisons showed that, as expected, the levels of DNA methylation of the germ cells were lower than those of somatic cells (average M-values: sperm = −0.31, blood = −0.08, neuron = 0.02, glia = 0.02) (Supplementary Fig. S2 and Fig. S3A), whilst inter-individual variability of sperm was higher than observed in differentiated somatic cells [mean of log (SD): sperm = −1.41; blood = −1.66; glia = −1.80; neuron = −1.77] (Supplementary Fig. S3A). Furthermore, we observed that sperm have a large number of CpG sites showing extreme values (89.5%), both unmethylated (<40 methylation %) and methylated (>60 methylation %), while the differentiated somatic cells showed a large number of probes with intermediate methylation values (~33%) (Supplementary Fig. S3A). In addition, in sperm cells a larger number of probes showed specific DNA methylation patterns (unmethylated 21 139; methylated 1842) than in any of the somatic cell types (see Materials and Methods) (Supplementary Fig. S3B and Supplementary Table SVI). Intriguingly, tissue-specific unmethylated CpGs were enriched in non-CGIs in all somatic cell types, whilst in sperm they were enriched in CGI-shores, and, conversely, tissue-specific methylated CpGs were enriched in CGI-shores in all the somatic cells whilst being enriched in non-CGIs in sperm (Supplementary Fig. S4), highlighting the peculiarity of this cell type in terms of DNA methylation patterns.

Most importantly though, we found a significant association between those CpG sites with specific DNA methylation in sperm and the dmCpG sites (hyper- and hypomethylated) identified in infertile men ($P < 0.001$; Fisher's exact test, OR = 1.43), which suggests that the alterations of sperm-specific DNA methylation patterns could be associated with male infertility. A more detailed analysis showed that only hypomethylated CpGs were significantly associated with both sperm-specific methylated ($P = 0.038$, OR: 2.00) or unmethylated ($P < 0.001$; OR: 2.08) CpGs.

Global DNA methylation patterns: comparative analysis between fertile and infertile individuals

To evaluate changes in global 5-mC levels (methylation assay 2 in Table I) that might be associated with male infertility we analysed a different set of samples from seven fertile individuals (samples Control-6 to 12) and seven normospermic infertile patients (samples Patient-8 to 14). Only two of the female partners (those of Patient-9 and 11) had a potential factor (tubal obstruction) that could influence the fertility of the couple. We found no significant differences in global 5-mC levels between fertile and infertile sperm samples (Fig. 3A).

In addition, since an association between DNA methylation of repetitive elements and total genomic 5-methylcytosine has been described (Ehrlich, 2002; Yang et al., 2004; Weisenberger et al., 2005), we analysed the methylation patterns of several repetitive sequences (methylation assay 3 in Table I), both at whole-genome level and at four specific repetitive loci, in sperm from 17 fertile (samples Control-1 to 17) and 27 infertile subjects (samples Patient-3 to 29) and compared them with differentiated somatic cells (blood and brain). Successful outcome in IVF-ICSI treatment was also taken into account (12 resulted in no pregnancy and 15 resulted in pregnancy; Table I) as success in IVF-ICSI would
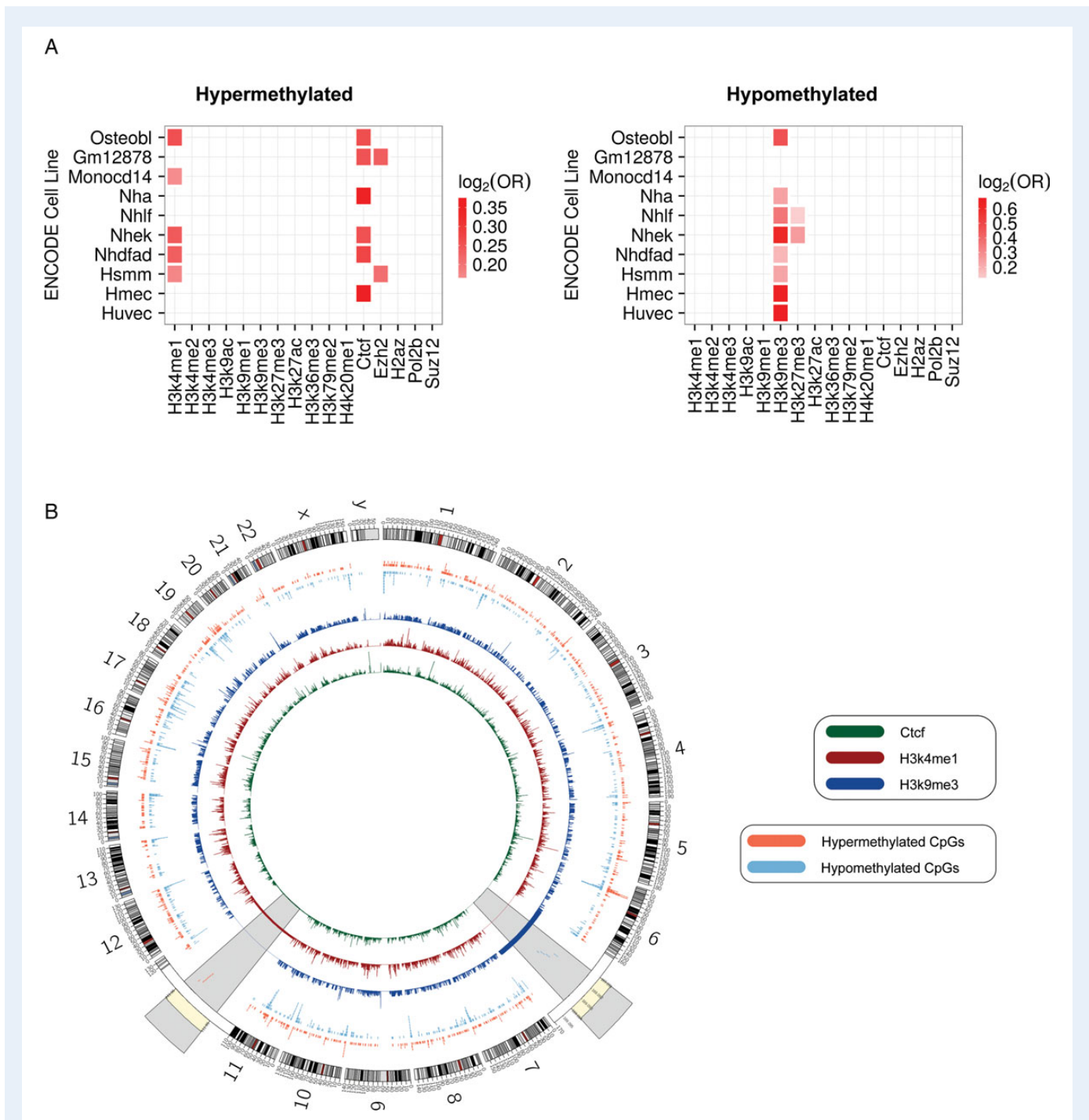
Figure 2 Chromatin signatures associated with aberrant DNA methylation in infertile patients. (**A**) Heatmaps showing significant enrichment of hyper-and hypomethylated CpG sites identified in infertile individuals, with different histone marks and chromatin modifiers contained in the UCSC Browser Broad Histone track from the ENCODE project. Colour code indicates significant enrichment based on log2 OR. (**B**) Circular representation of genome-wide DNA methylation changes in the infertile patients indicating whether the CpGs were hypermethylated (red) or hypomethylated (blue). Inner tracks display chromatin marks (Ctcf, H3K4me1 and H3K9me3) generated for osteoblast cells, and associated with differentially methylated regions. Broad histone peak information is averaged in 200 kbp genomic windows and represented as histogram tracks. Two regions of chromosomes 6 and 11 are magnified (grey area) in order to detail the associations between hypo- or hypermethylated DNA regions with specific chromatin signatures.

provide extra indications for the lack of a phenotype associated with the female for the selected couples.

Specifically, we determined the methylation status of four repetitive sequences, namely: LINE-1, an interspersed repeat found throughout the human genome and used as an estimate of global methylation levels (Yang *et al.*, 2004); Alu Yb8, a relatively young subfamily of the Alu short interspersed elements (SINEs) (Carroll *et al.*, 2001); D4Z4, a macrosatellite found in the subtelomeric regions (Chadwick, 2009)
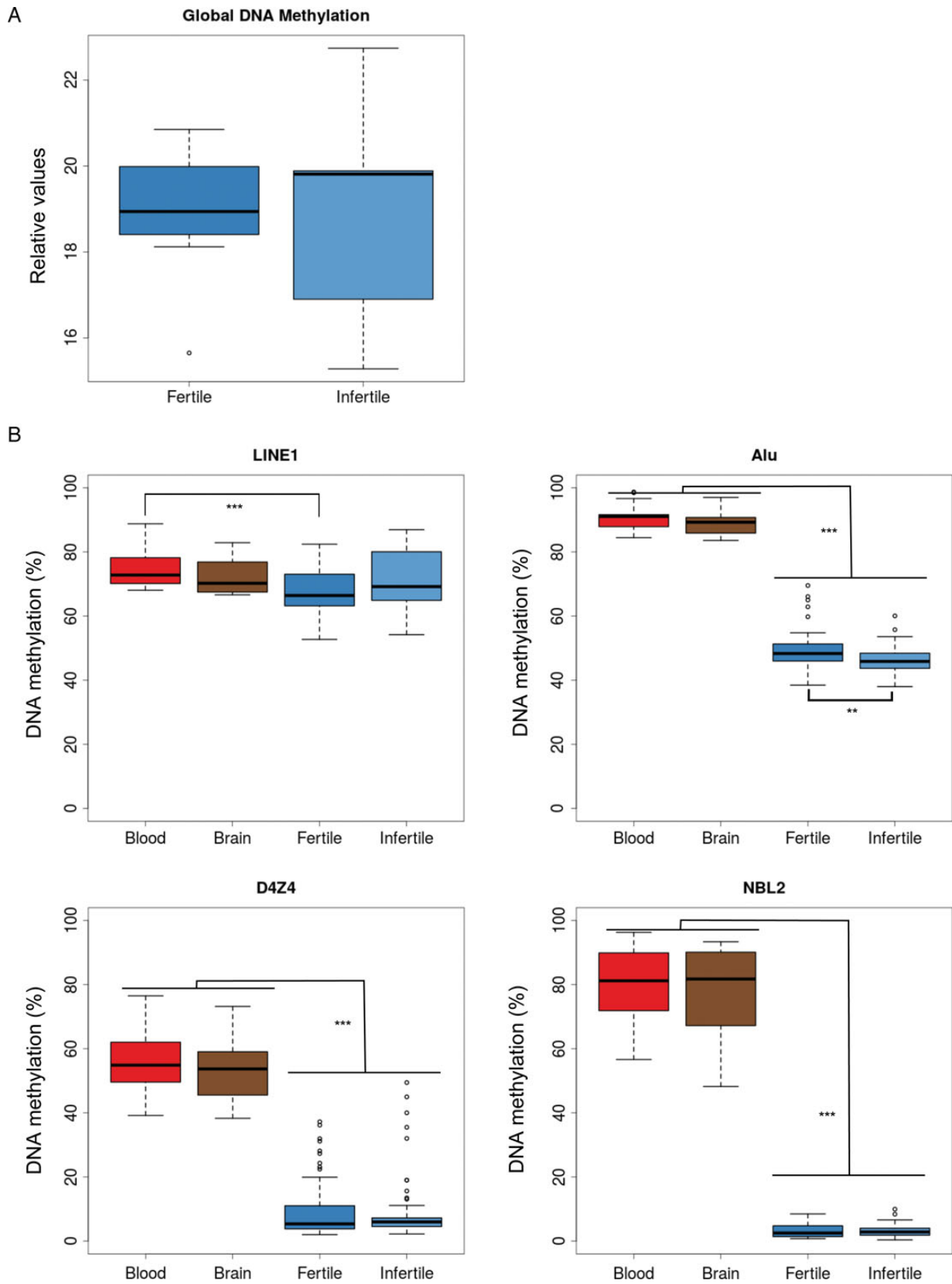
Figure 3 Global DNA methylation patterns in sperm. (**A**) Global DNA methylation levels of sperm from fertile individuals and normospermic infertile patients obtained in a colorimetric assay. (**B**) DNA methylation values of several repetitive regions (LINE-1, Alu Yb8, NBL-2 and D4Z4) measured by bisulfite pyrosequencing in sperm (controls and patients) and somatic cells (blood and brain). ***$P < 0.001$; **$P < 0.01$.

and NBL-2, a complex tandem repeat found in the centromeric regions of acrocentric chromosomes (Nishiyama *et al.*, 2005). Our results showed no differences between fertile and infertile subjects in any of the repetitive sequences analysed, with the exception of Alu Yb8, where sperm of infertile individuals showed significantly lower Alu methylation levels ($P = 0.0011$) (Fig. 3B). Furthermore, the comparative analysis of DNA methylation of repetitive sequences (LINE-1, Alu Yb8, NBL2, D4Z4) showed no statistical differences between the patient subgroups with respect to IVF-ICSI outcome.

However, we did find huge DNA methylation differences between germ and somatic cells (blood, brain) in most of the repetitive sequences analysed, and in all cases sperm samples showed lower average methylation values (Fig. 3B). NBL-2 showed the largest differences (average methylation: sperm = 3.2%, blood = 80.2 and brain = 76.3; $P < 0.001$). Alu (average methylation: sperm = 47.5%, blood = 90.6 and brain = 89.3; $P < 0.001$) and D4Z4 (average methylation: sperm = 7.9%, blood = 55.6, and brain = 53.6; $P < 0.001$) also showed great differences, and LINE-1 displayed the lowest difference, with only the comparison with blood cells being statistically significant (average methylation: sperm = 69.8%, blood = 74.6 and brain = 72.6) (Fig. 3B). These results agree with the DNA methylation patterns of different repetitive regions found in human sperm and human embryonic stem cells (hESCs) (Molaro *et al.*, 2011) as well as in human sperm and B cells (Krausz *et al.*, 2012).

Discussion

Sperm and testicular DNA methylation profiles of specific genes or genomic regions from fertile and infertile human males have been compared in several studies to date (Marques *et al.*, 2004; Houshdaran *et al.*, 2007; Kobayashi *et al.*, 2007; Marques *et al.*, 2008; Poplinski *et al.*, 2010; Pacheco *et al.*, 2011; Heyn *et al.*, 2012). These analyses were, however, either restricted to a small number of genes or imprinted regions, or considered a larger number of genes but focused only on promoter regions. What is more, only infertile patients presenting alterations associated with either the process of spermatogenesis or semen quality were studied. In contrast, in this work we compare, for the first time, and at genome-wide level, the DNA methylation patterns of sperm from fertile individuals with the sperm of patients with unexplained infertility. Furthermore, the confounding role of the significant proportion of female causes that contribute to reproductive failure was controlled since sperm was selected from couples with no known significant female risk factor.

We analysed gene promoters and intragenic and intergenic regions and identified alterations in the methylation of DNA in 2752 CpGs. It is of particular note that a high level of variation in DNA methylation was observed in mature germ cells across unrelated individuals, even among fertile individuals, and may reflect differing levels of sperm function. Specifically, we found no alterations in the DNA methylation patterns of genes of the PIWI pathway, previously found to be associated with severe defects of sperm production (Heyn *et al.*, 2012). Neither did we find evidence for the previously described alterations in imprinted genes such as imprinted maternally expressed transcript (*H19*) or mesoderm specific transcript (*MEST*) (Marques *et al.*, 2004, 2008; Kobayashi *et al.*, 2007; Poplinski *et al.*, 2010; Pacheco *et al.*, 2011) or non-imprinted genes such as *MTHFR* (Houshdaran *et al.*, 2007; Wu *et al.*, 2010b; Pacheco *et al.*, 2011) which have been associated with male infertility due to poor semen quality. However, we extended our study on imprinted genes and found a small number of CpGs showing aberrant

DNA methylation associated with 48 imprinted genes in infertile patients. Subsequently, we compared our results with data obtained in a previous study (Pacheco *et al.*, 2011) where 619 CpGs associated with imprinted genes were analysed in samples from patients with low sperm motility, and, we only found two common CpG sites aberrantly methylated. Interestingly, one of the genes associated with these CpGs (*IGF2*) has been found also aberrantly methylated in abnormal sperm (Boissonnas *et al.*, 2010; Poplinski *et al.*, 2010; Pacheco *et al.*, 2011). Unlike in other studies to date, here we analysed patients with normal or mild defects of semen quality, and our results show previously unidentified alterations in DNA methylation of CpG sites located at specific genes and genomic regions. This suggests that the mechanistic origins of these alterations in DNA methylation in individuals with unexplained infertility may be different from those associated with the onset of alterations in patients with low semen quality.

We used several genomic approaches to further study the peculiarities of these dmCpGs that might give some clue to their contribution to male infertility. In the case of hypomethylated CpGs, besides being preferentially located in CGI, they are more abundant in intergenic regions and depleted in promoters, all of which suggests that these CpGs are located in the clusters of CGI II and III recently described by Zeng and collaborators, which have been associated with tissue-specific DNA methylation (Zeng *et al.*, 2014). Intriguingly, hypermethylated related genes were also enriched for a cell adhesion ontology term, opening new avenues for investigating the functional role that aberrant DNA hypermethylation may have in cases of unexplained male infertility. This could be at least partly related to defects in functional capacity of sperm to bind the oocyte in infertile patients. It could also give evidence that some of these defects in methylation are shared with sperm morphology deficiencies, from results of other studies in which associations between teratozoospermia and deficiency in adhesion molecules were found (Glander and Schaller, 1993).

Since the aberrant DNA methylation appeared to be related to changes in tissue-specific methylation, we aimed to identify CpGs which were specifically methylated in sperm when compared with somatic cells (brain and blood). These sperm-specific methylated regions were subsequently compared with the dmCpGs in infertile individuals. In addition to showing different DNA methylation patterns between somatic and germ cells, as previously described (Eckhardt *et al.*, 2006; Oakes *et al.*, 2007b; Krausz *et al.*, 2012), we found a statistically significant association between sperm dmCpGs, mainly hypomethylated CpGs in individuals with unexplained infertility and CpGs that showed sperm-specific DNA methylation, which suggests that the alterations of the mechanisms that establish the sperm-specific epigenetic program could be involved in the fertilizing quality of sperm in unexplained human male infertility.

In addition, we analysed the genomic location of these dmCpGs in the context of chromatin and found associations between aberrant DNA methylation and specific histone marks previously identified in somatic and sperm cells. Interestingly, our results showed a significant association between hypomethylated CpGs in sperm and regions strongly enriched in repressive histone marks such as H3K9me3 in somatic cells, whereas hypermethylated regions were associated with H3K4me1 and CTCF. Since there is no reason why these chromatin marks should necessarily be the same in somatic cells and sperm cells, these associations may merely indicate that during germ cell development these dmCpGs are associated with regions with some distinct feature that leads to a differential 'vulnerability' to hypo- or hyper methylation in association with

infertility. On the other hand, if the chromatin marks in somatic cells associated with aberrant DNA methylation in sperm were the same in the germ cell, DNA hypermethylation of CTCF binding sites could be indicative of alterations in the architecture and function of the sperm genome of infertile patients, since CTCF binds to DNA sequences in a methylation-sensitive manner (Wang et al., 2012), and it has been shown that CTCF appears to play a significant role in chromatin organization, as well as in the regulation of gene expression (Phillips and Corces, 2009; Wang et al., 2012; Ong and Corces, 2014), which is especially relevant in mammalian sperm (Arpanahi et al., 2009; Carone et al., 2014).

In addition, we analysed associations between aberrant DNA methylation with the histone marks identified in human sperm and available in public data bases (Hammoud et al., 2009), and found a significant association between hypermethylated CpGs in the sperm of infertile patients and genomic regions enriched for H3K4me3. Taken all together, our findings suggest that the relationship between chromatin context and the aberrant DNA methylation of sperm in infertile men could be locus-dependent. Future studies analysing the complete maps of histone post-translational marks of sperm chromatin in normospermic infertile patients and fertile individuals will elucidate whether these alterations in DNA methylation are also associated with alterations of other specific histone marks, and whether they could affect chromatin compaction, as suggested in other studies analysing subfertile individuals (Steilmann et al., 2010; La Spina et al., 2014)

Apart from locus-specific DNA methylation differences, we analysed for the first time global DNA methylation changes between normospermic fertile and infertile patients. Our results showed no differences in global methylation between the groups, in contrast to the results previously found in infertile patients with poor quality of sperm (Tunc and Tremellen, 2009), suggesting that global DNA methylation changes are related to spermatogenic efficiency and the semen quality of infertile patients.

We also analysed DNA methylation changes between groups in several repetitive elements across the genome, including LINE-1, considered to represent global DNA methylation (Yang et al., 2004). The dynamics of the DNA methylation of repetitive DNA elements during epigenetic reprogramming of primordial germ cells are gender specific (Lees-Murdock and Walsh, 2008; Sasaki and Matsui, 2008), and the functional role of the DNA methylation of retrotransposons, particularly in male germ cells, has been described (Bourc'his and Bestor, 2004). Our results related to the DNA methylation of LINE-1 retrotransposon showed no differences between groups and agree with previous results analysing differences in infertile men exhibiting low sperm concentrations (Kobayashi et al., 2007; Marques et al., 2008). In contrast, LINE-1 repetitive sequences have been found to be hypomethylated in infertile patients with severe spermatogenic disorders (Heyn et al., 2012) associated with the epigenetic inactivation of piRNA-processing genes: PIWIL2 and TDRD1. Our results, however, not only failed to show differences in the DNA methylation patterns of PIWIL2 and TDRD1 between groups (data not shown), but neither did we find differences in methylation of LINE-1 between normospermic fertile and infertile patients.

Interestingly, we did find a significant DNA methylation decrease in another retrotransposon, AluYb8, in infertile patients. The results from previous works focused on infertile individuals showing low seminal quality had been unable to clarify the associations between male infertility and the methylation of Alu sequences (Kobayashi et al., 2007; El Hajj et al., 2011); Kobayashi and collaborators did not find differences between healthy controls and infertile patients (Kobayashi et al., 2007), whereas El Hajj and collaborators, in line with the results of our study, showed that average methylation values in Alu sequences were lower in infertile men with abnormal semen parameters (El Hajj et al., 2011). The differences between these two studies could be attributed to several causes. First different methodologies were used to measure methylation; one case used combined bisulfite restriction analysis (COBRA) assay and the other, bisulfite pyrosequencing. Second the difference could be the consequence of the specific Alu sequence analysed. In our work we specifically analysed the Alu Yb8 subfamily, which is relatively young and more susceptible to retrotransposon activity, thus requiring stricter control by epigenetic silencing mechanisms. The role of SINE elements in the regulation of gene expression has recently been investigated in murine models and been shown to play a role in the activity of downstream gene promoters (Estecio et al., 2012). Methylation of SINEs in the transcriptional regulation of genes specifically expressed in testis has also been suggested to have a role (Ichiyanagi et al., 2011). Thus, future studies are needed to elucidate the functional effects of aberrant methylation of Alu sequences in sperm from infertile individuals.

We also analysed and compared, for the first time, the DNA methylation patterns of pericentromeric (NBL2) and subtelomeric (D4Z4) repetitive elements in both control and infertile patients and found low methylation values in these regions, but no significant differences between groups. That said, the results of the analysis of DNA methylation in both global and repetitive sequences should be considered carefully, because the presence of mild reproductive risk factors in some females might be masking potentially fertile individuals.

Together, these results suggest that alterations of DNA methylation, both globally and at locus-specific level, and therefore of the mechanisms that produce them are different in normospermic infertile patients compared with infertile individuals with spermatogenic impairment. As in the case of DNA methylation at specific loci, we also performed a comparative analysis of DNA methylation of the repetitive regions mentioned above in order to analyse differences between germ and somatic cells, and to ascertain the peculiarities of the former. It has previously been shown that mammalian testes have more hypomethylated loci than somatic cells (Oakes et al., 2007b). In addition, the DNA methylation of repetitive sequences has been found to be particularly low in sperm compared with undifferentiated somatic cells (hESC) (Molaro et al., 2011). To further characterize the methylation patterns of these repetitive sequences in sperm, we also performed methylation analysis of some of these regions in differentiated somatic cells (blood and brain). The greatest difference in methylation levels between sperm and somatic cells was observed for NBL-2. This is also the case when looking at other centromeric repeats (Molaro et al., 2011), suggesting therefore that an overall hypomethylation of the centromeric region is characteristic of sperm. Indeed, Yamagata and collaborators have previously proposed using methylation levels of the centromeric region in order to discriminate between germ and somatic cell lineages (Yamagata et al., 2007). Moreover, although pericentromeric regions are usually regarded as transcriptionally poor (Copenhaver et al., 1999; Nagaki et al., 2004), the expression of a variety of genes has been shown in the testis, with almost half the cases being unique to the tissue (She et al., 2004). In the case of D4Z4, although to date no studies have concentrated on its methylation in sperm, Jian Li and collaborators have reported enrichment in 'methylation deserts' in telomeric regions (Li et al., 2012). Thus,

the low methylation values found here for D4Z4 are probably, as in the case of NBL-2, reflective of overall hypomethylation occurring in that region and highlights the increased risk of structural mutations in germ cells due to these methylation deserts (Li et al., 2012).

Although Alu Yb8 methylation values in sperm were higher than those found for the other repetitive elements studied, sperm and somatic cells showed large differences, confirming the findings of previous studies using other analysis techniques (Hellmann-Blumberg et al., 1993; Kochanek et al., 1993). Finally, unlike the other repetitive sequences, LINE-1 showed high levels of methylation, with significant differences only being found between sperm and blood cells. When comparing studies based on the same technique (El Hajj et al., 2011; Heyn et al., 2012), the methylation values of this sequence in sperm were similar. This finding is partially supported by the results of Molaro and collaborators (Molaro et al., 2011), who have studied methylation along the full length of LINE-1 and discovered a higher percentage of hypomethylated regions in sperm compared with hESCs. Although there are major differences between the results of Molaro and collaborators and our own, these could be explained because the former analysed undifferentiated somatic cells while differentiated somatic cells were considered in this work, and because we only studied three CpGs, thereby only reflecting methylation for specific locations on LINE-1.

In conclusion, DNA methylation patterns of spermatozoa are significantly different to those found in other somatic cells such as blood or brain. In this work we have analysed for the first time, at genome-wide resolution, the DNA methylation profiles of the sperm of patients with unexplained infertility versus that of fertile individuals, and we have identified almost 3000 CpGs which display aberrant methylation. Our data show that these changes are precisely associated with regions of sperm-specific methylation, thereby suggesting that DNA methylation is involved in the control of the functional capacity of germ cells. Further studies are necessary to elucidate the mechanisms relating to the origin of these alterations, and to determine their significance and functional consequences for male infertility.

Conflict of interest

The authors confirm that they have no conflict of interest to declare.

References

Anway MD, Cupp AS, Uzumcu M, Skinner MK. Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science* 2005; **308**:1466−1469.

Arpanahi A, Brinkworth M, Iles D, Krawetz SA, Paradowska A, Platts AE, Saida M, Steger K, Tedder P, Miller D. Endonuclease-sensitive regions of human spermatozoal chromatin are highly enriched in promoter and CTCF binding sequences. *Genome Res* 2009; **19**:1338−1349.

Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, McMahon S, Karlsson EK, Kulbokas EJ, Gingeras TR et al. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* 2005; **120**:169−181.

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 2006; **125**:315−326.

Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL et al. High density DNA methylation array with single CpG site resolution. *Genomics* 2011; **98**:288−295.

Boissonnas CC, Abdalaoui HE, Haelewyn V, Fauque P, Dupont JM, Gut I, Vaiman D, Jouannet P, Tost J, Jammes H. Specific epigenetic alterations of IGF2-H19 locus in spermatozoa from infertile men. *Eur J Hum Genet* 2010; **18**:73−80.

Bollati V, Baccarelli A, Hou L, Bonzini M, Fustinoni S, Cavallo D, Byun HM, Jiang J, Marinelli B, Pesatori AC et al. Changes in DNA methylation patterns in subjects exposed to low-dose benzene. *Cancer Res* 2007; **67**:876−880.

Bourc'his D, Bestor TH. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* 2004; **431**:96−99.

Carlson M. TxDb.Hsapiens.UCSC.hg19.knownGene: Annotation package for TranscriptDb object(s). R package version 2.9.2.

Carone BR, Hung JH, Hainer SJ, Chou MT, Carone DM, Weng Z, Fazzio TG, Rando OJ. High-resolution mapping of chromatin packaging in mouse embryonic stem cells and sperm. *Dev Cell* 2014; **30**:11−22.

Carrell DT, Hammoud SS. The human sperm epigenome and its potential role in embryonic development. *Mol Hum Reprod* 2010; **16**:37−47.

Carroll ML, Roy-Engel AM, Nguyen SV, Salem AH, Vogel E, Vincent B, Myers J, Ahmad Z, Nguyen L, Sammarco M et al. Large-scale analysis of the Alu Ya5 and Yb8 subfamilies and their contribution to human genomic diversity. *J Mol Biol* 2001; **311**:17−40.

Cooper TG, Noonan E, von Eckardstein S, Auger J, Baker HW, Behre HM, Haugen TB, Kruger T, Wang C, Mbizvo MT et al. World Health Organization reference values for human semen characteristics. *Hum Reprod Update* 2010; **16**:231−245.

Copenhaver GP, Nickel K, Kuromori T, Benito MI, Kaul S, Lin X, Bevan M, Murphy G, Harris B, Parnell LD et al. Genetic definition and sequence analysis of Arabidopsis centromeres. *Science* 1999; **286**:2468−2474.

Chadwick BP. Macrosatellite epigenetics: the two faces of DXZ4 and D4Z4. *Chromosoma* 2009; **118**:675−681.

Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* 2013;**8**:203–209.

Choi SH, Worswick S, Byun HM, Shear T, Soussa JC, Wolff EM, Douer D, Garcia-Manero G, Liang G, Yang AS. Changes in DNA methylation of tandem DNA repeats are different from interspersed repeats in cancer. *Int J Cancer* 2009;**125**:723–729.

Dada R, Kumar M, Jesudasan R, Fernandez JL, Gosalvez J, Agarwal A. Epigenetics and its role in male infertility. *J Assist Reprod Genet* 2012;**29**:213–223.

Doerfler W. Patterns of DNA methylation—evolutionary vestiges of foreign DNA inactivation as a host defense mechanism. A proposal. *Biol Chem Hoppe Seyler* 1991;**372**:557–564.

Du P, Zhang X, Huang C-C, Jafari N, Kibbe WA, Hou L, Lin SM. Comparison of Beta-value and *M*-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* 2010;**11**:587.

Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, Burton J, Cox TV, Davies R, Down TA et al. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet* 2006;**38**:1378–1385.

Ehrlich M. DNA methylation in cancer: too much, but also too little. *Oncogene* 2002;**21**:5400–5413.

El Hajj N, Zechner U, Schneider E, Tresch A, Gromoll J, Hahn T, Schorsch M, Haaf T. Methylation status of imprinted genes and repetitive elements in sperm DNA from infertile males. *Sex Dev* 2011;**5**:60–69.

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 2011;**473**:43–49.

Estecio MR, Gallegos J, Dekmezian M, Lu Y, Liang S, Issa JP. SINE retrotransposons cause epigenetic reprogramming of adjacent gene promoters. *Mol Cancer Res* 2012;**10**:1332–1342.

Esteller M. Epigenetics in cancer. *N Engl J Med* 2008;**358**:1148–1159.

Feinberg AP, Cui H, Ohlsson R. DNA methylation and genomic imprinting: insights from cancer into epigenetic mechanisms. *Semin Cancer Biol* 2002;**12**:389–398.

Fernandez AF, Assenov Y, Martin-Subero JI, Balint B, Siebert R, Taniguchi H, Yamamoto H, Hidalgo M, Tan AC, Galm O et al. A DNA methylation fingerprint of 1628 human samples. *Genome Res* 2012;**22**:407–419.

Flanagan JM, Popendikyte V, Pozdniakovaite N, Sobolev M, Assadzadeh A, Schumacher A, Zangeneh M, Lau L, Virtanen C, Wang SC et al. Intra- and interindividual epigenetic variation in human germ cells. *Am J Hum Genet* 2006;**79**:67–84.

Glander HJ, Schaller J. Semen samples with teratozoospermia show a lower percentage of spermatozoa with detectable adhesion molecules (AM). *Int Urol Nephrol* 1993;**25**:485–489.

Guintivano J, Aryee MJ, Kaminsky ZA. A cell epigenotype specific model for the correction of brain cellular heterogeneity bias and its application to age, brain region and major depression. *Epigenetics* 2013;**8**:290–302.

Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, Fan L, Koziol MJ, Gnirke A, Nusbaum C et al. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat Biotechnol* 2010;**28**:503–510.

Hajkova P, Erhardt S, Lane N, Haaf T, El-Maarri O, Reik W, Walter J, Surani MA. Epigenetic reprogramming in mouse primordial germ cells. *Mech Dev* 2002;**117**:15–23.

Hamada A, Esteves SC, Agarwal A. Unexplained male infertility: potential causes and management. *Human Androl* 2011;**1**:2–16.

Hammoud SS, Nix DA, Zhang H, Purwar J, Carrell DT, Cairns BR. Distinctive chromatin in human sperm packages genes for embryo development. *Nature* 2009;**460**:473–478.

Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, Klotzle B, Bibikova M, Fan JB, Gao Y et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell* 2013;**49**:359–367.

Hansen KD, Aryee M. minfi: Analyse Illumina's 450k methylation arrays. R package version 1.7.15.

Hellmann-Blumberg U, Hintz MF, Gatewood JM, Schmid CW. Developmental differences in methylation of human Alu repeats. *Mol Cell Biol* 1993;**13**:4523–4530.

Heyn H, Ferreira HJ, Bassas L, Bonache S, Sayols S, Sandoval J, Esteller M, Larriba S. Epigenetic disruption of the PIWI pathway in human spermatogenic disorders. *PLoS One* 2012;**7**:e47892.

Holliday R. The inheritance of epigenetic defects. *Science* 1987; **238**:163–170.

Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 2012;**13**:86.

Houshdaran S, Cortessis VK, Siegmund K, Yang A, Laird PW, Sokol RZ. Widespread epigenetic abnormalities suggest a broad DNA methylation erasure defect in abnormal human sperm. *PLoS One* 2007;**2**:e1289.

Ichiyanagi K, Li Y, Watanabe T, Ichiyanagi T, Fukuda K, Kitayama J, Yamamoto Y, Kuramochi-Miyagawa S, Nakano T, Yabuta Y et al. Locus- and domain-dependent control of DNA methylation at mouse B1 retrotransposons during male germ cell development. *Genome Res* 2011; **21**:2058–2066.

Irvine DS. Epidemiology and aetiology of male infertility. *Hum Reprod* 1998;**13** Suppl 1:33–44.

Jarow JP, Sharlip ID, Belker AM, Lipshultz LI, Sigman M, Thomas AJ, Schlegel PN, Howards SS, Nehra A, Damewood MD et al. Best practice policies for male infertility. *J Urol* 2002;**167**:2138–2144.

Jenkins TG, Carrell DT. The sperm epigenome and potential implications for the developing embryo. *Reproduction* 2012;**143**:727–734.

Kobayashi H, Sato A, Otsu E, Hiura H, Tomatsu C, Utsunomiya T, Sasaki H, Yaegashi N, Arima T. Aberrant DNA methylation of imprinted loci in sperm from oligospermic patients. *Hum Mol Genet* 2007;**16**:2542–2551.

Kochanek S, Renz D, Doerfler W. DNA methylation in the Alu sequences of diploid and haploid primary human cells. *EMBO J* 1993;**12**:1141–1151.

Krausz C, Sandoval J, Sayols S, Chianese C, Giachini C, Heyn H, Esteller M. Novel insights into DNA methylation features in spermatozoa: stability and peculiarities. *PLoS One* 2012;**7**:e44479.

La Spina FA, Romanato M, Brugo-Olmedo S, De Vincentiis S, Julianelli V, Rivera RM, Buffone MG. Heterogeneous distribution of histone methylation in mature human sperm. *J Assist Reprod Genet* 2014;**31**:45–49.

Lees-Murdock DJ, Walsh CP. DNA methylation reprogramming in the germ line. *Epigenetics* 2008;**3**:5–13.

Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 1992;**69**:915–926.

Li J, Harris RA, Cheung SW, Coarfa C, Jeong M, Goodell MA, White LD, Patel A, Kang SH, Shaw C et al. Genomic hypomethylation in the human germline associates with selective structural mutability in the human genome. *PLoS Genet* 2012;**8**:e1002692.

Makismovic J, Gordon L, Oshlack A. SWAN: Subset-quantile within array normalization for Illumina Infinium HumanMethylation450 BeadChips. *Genome Biol* 2012;**13**:R44.

Marques CJ, Carvalho F, Sousa M, Barros A. Genomic imprinting in disruptive spermatogenesis. *Lancet* 2004;**363**:1700–1702.

Marques CJ, Costa P, Vaz B, Carvalho F, Fernandes S, Barros A, Sousa M. Abnormal methylation of imprinted genes in human sperm is associated with oligozoospermia. *Mol Hum Reprod* 2008;**14**:67–74.

Martinez JG, Perez-Escuredo J, Castro-Santos P, Marcos CA, Pendas JL, Fraga MF, Hermsen MA. Hypomethylation of LINE-1, and not centromeric SAT-alpha, is associated with centromeric instability in head and neck squamous cell carcinoma. *Cell Oncol (Dordr)* 2012;**35**:259–267.

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim T-K, Koche RP et al. Genome-wide maps of chromatin

state in pluripotent and lineage-committed cells. *Nature* 2007;**448**: 553–560.

Molaro A, Hodges E, Fang F, Song Q, McCombie WR, Hannon GJ, Smith AD. Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. *Cell* 2011;**146**:1029–1041.

Nagaki K, Cheng Z, Ouyang S, Talbert PB, Kim M, Jones KM, Henikoff S, Buell CR, Jiang J. Sequencing of a rice centromere uncovers active genes. *Nat Genet* 2004;**36**:138–145.

Nanassy L, Carrell DT. Analysis of the methylation pattern of six gene promoters in sperm of men with abnormal protamination. *Asian J Androl* 2011;**13**:342–346.

Navarro-Costa P, Nogueira P, Carvalho M, Leal F, Cordeiro I, Calhaz-Jorge C, Goncalves J, Plancha CE. Incorrect DNA methylation of the DAZL promoter CpG island associates with defective human sperm. *Hum Reprod* 2010;**25**:2647–2654.

Nishiyama R, Qi L, Tsumagari K, Weissbecker K, Dubeau L, Champagne M, Sikka S, Nagai H, Ehrlich M. A DNA repeat, NBL2, is hypermethylated in some cancers but hypomethylated in others. *Cancer Biol Ther* 2005;**4**:440–448.

Oakes CC, La Salle S, Smiraglia DJ, Robaire B, Trasler JM. Developmental acquisition of genome-wide DNA methylation occurs prior to meiosis in male germ cells. *Dev Biol* 2007a;**307**:368–379.

Oakes CC, La Salle S, Smiraglia DJ, Robaire B, Trasler JM. A unique configuration of genome-wide DNA methylation patterns in the testis. *Proc Natl Acad Sci USA* 2007b;**104**:228–233.

Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 1999;**99**:247–257.

Ong CT, Corces VG. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet* 2014;**15**:234–246.

Pacheco SE, Houseman EA, Christensen BC, Marsit CJ, Kelsey KT, Sigman M, Boekelheide K. Integrative DNA methylation and gene expression analyses identify DNA packaging and epigenetic regulatory genes associated with low motility sperm. *PLoS One* 2011;**6**:e20280.

Pan KH, Lih CJ, Cohen SN. Effects of threshold choice on biological conclusions reached during analysis of gene expression by DNA microarrays. *Proc Natl Acad Sci USA* 2005;**102**:8961–8965.

Payer B, Lee JT. X chromosome dosage compensation: how mammals keep the balance. *Annu Rev Genet* 2008;**42**:733–772.

Pedigo NG, Vernon MW, Curry TE Jr. Characterization of a computerized semen analysis system. *Fertil Steril* 1989;**52**:659–666.

Phillips JE, Corces VG. CTCF: master weaver of the genome. *Cell* 2009; **137**:1194–1211.

Poplinski A, Tuttelmann F, Kanber D, Horsthemke B, Gromoll J. Idiopathic male infertility is strongly associated with aberrant methylation of MEST and IGF2/H19 ICR1. *Int J Androl* 2010;**33**:642–649.

Reik W, Dean W, Walter J. Epigenetic reprogramming in mammalian development. *Science* 2001;**293**:1089–1093.

Romero Y, Meikar O, Papaioannou MD, Conne B, Grey C, Weier M, Pralong F, De Massy B, Kaessmann H, Vassalli JD et al. Dicer1 depletion in male germ cells leads to infertility due to cumulative meiotic and spermiogenic defects. *PLoS One* 2011;**6**:e25241.

Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, Esteller M. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* 2011;**6**:692–702.

Sasaki H, Matsui Y. Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nat Rev Genet* 2008;**9**:129–140.

She X, Horvath JE, Jiang Z, Liu G, Furey TS, Christ L, Clark R, Graves T, Gulden CL, Alkan C et al. The structure and evolution of centromeric transition regions within the human genome. *Nature* 2004;**430**:857–864.

Smyth GK. Limma: linear models for microarray data. In: Gentleman R, Carey V, Dudoit S, Irizarry R and Huber W (eds). *Bioinformatics and Computational Biology Solutions using R and Bioconductor*. New York: Springer, 2005, 397–420.

Steilmann C, Cavalcanti MC, Bartkuhn M, Pons-Kuhnemann J, Schuppe HC, Weidner W, Steger K, Paradowska A. The interaction of modified histones with the bromodomain testis-specific (BRDT) gene and its mRNA level in sperm of fertile donors and subfertile men. *Reproduction* 2010; **140**:435–443.

Triche TJ. FDb.InfiniumMethylation.hg19: Annotation package for Illumina Infinium DNA methylation array probes. R package version 1.0.1.

Tunc O, Tremellen K. Oxidative DNA damage impairs global sperm DNA methylation in infertile men. *J Assist Reprod Genet* 2009;**26**:537–544.

Urdinguio RG, Sanchez-Mut JV, Esteller M. Epigenetic mechanisms in neurological diseases: genes, syndromes, and therapies. *Lancet Neurol* 2009;**8**:1056–1072.

Wang H, Maurano MT, Qu H, Varley KE, Gertz J, Pauli F, Lee K, Canfield T, Weaver M, Sandstrom R et al. Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res* 2012;**22**:1680–1688.

Weisenberger DJ, Campan M, Long TI, Kim M, Woods C, Fiala E, Ehrlich M, Laird PW. Analysis of repetitive element DNA methylation by MethyLight. *Nucleic Acids Res* 2005;**33**:6823–6836.

World Health Organization. *WHO laboratory manual for the examination and processing of human semen*, 5th edn. World Health Organization, 2010.

Wu H, Caffo B, Jaffee HA, Irizarry RA, Feinberg AP. Redefining CpG islands using hidden Markov models. *Biostatistics* 2010a;**11**:499–514.

Wu W, Shen O, Qin Y, Niu X, Lu C, Xia Y, Song L, Wang S, Wang X. Idiopathic male infertility is strongly associated with aberrant promoter methylation of methylenetetrahydrofolate reductase (MTHFR). *PLoS One* 2010b;**5**:e13884.

Yamagata K, Yamazaki T, Miki H, Ogonuki N, Inoue K, Ogura A, Baba T. Centromeric DNA hypomethylation as an epigenetic signature discriminates between germ and somatic cell lineages. *Dev Biol* 2007; **312**:419–426.

Yaman R, Grandjean V. Timing of entry of meiosis depends on a mark generated by DNA methyltransferase 3a in testis. *Mol Reprod Dev* 2006; **73**:390–397.

Yang AS, Estecio MR, Doshi K, Kondo Y, Tajara EH, Issa JP. A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. *Nucleic Acids Res* 2004;**32**:e38.

Zeng J, Nagarajan HK, Yi SV. Fundamental diversity of human CpG islands at multiple biological levels. *Epigenetics* 2014;**9**:483–491.