

# Integrative DNA Methylation and Gene Expression Analyses Identify DNA Packaging and Epigenetic Regulatory Genes Associated with Low Motility Sperm

Sara E. Pacheco<sup>1</sup>, E. Andres Houseman<sup>2</sup>, Brock C. Christensen<sup>2</sup>, Carmen J. Marsit<sup>1</sup>, Karl T. Kelsey<sup>1,2</sup>, Mark Sigman<sup>3</sup>, Kim Boekelheide<sup>1\*</sup>

**1** Department of Pathology and Laboratory Medicine, Brown University, Providence, Rhode Island, United States of America, **2** Department of Community Health, Brown University, Providence, Rhode Island, United States of America, **3** Division of Urology, Brown University, Providence, Rhode Island, United States of America

## Abstract

**Background:** In previous studies using candidate gene approaches, low sperm count (oligospermia) has been associated with altered sperm mRNA content and DNA methylation in both imprinted and non-imprinted genes. We performed a genome-wide analysis of sperm DNA methylation and mRNA content to test for associations with sperm function.

**Methods and Results:** Sperm DNA and mRNA were isolated from 21 men with a range of semen parameters presenting to a tertiary male reproductive health clinic. DNA methylation was measured with the Illumina Infinium array at 27,578 CpG loci. Unsupervised clustering of methylation data differentiated the 21 sperm samples by their motility values. Recursively partitioned mixture modeling (RPMM) of methylation data resulted in four distinct methylation profiles that were significantly associated with sperm motility ( $P=0.01$ ). Linear models of microarray analysis (LIMMA) was performed based on motility and identified 9,189 CpG loci with significantly altered methylation ( $Q<0.05$ ) in the low motility samples. In addition, the majority of these disrupted CpG loci (80%) were hypomethylated. Of the aberrantly methylated CpGs, 194 were associated with imprinted genes and were almost equally distributed into hypermethylated (predominantly paternally expressed) and hypomethylated (predominantly maternally expressed) groups. Sperm mRNA was measured with the Human Gene 1.0 ST Affymetrix GeneChip Array. LIMMA analysis identified 20 candidate transcripts as differentially present in low motility sperm, including *HDAC1* (NCBI 3065), *SIRT3* (NCBI 23410), and *DNMT3A* (NCBI 1788). There was a trend among altered expression of these epigenetic regulatory genes and RPMM DNA methylation class.

**Conclusions:** Using integrative genome-wide approaches we identified CpG methylation profiles and mRNA alterations associated with low sperm motility.

**Citation:** Pacheco SE, Houseman EA, Christensen BC, Marsit CJ, Kelsey KT, et al. (2011) Integrative DNA Methylation and Gene Expression Analyses Identify DNA Packaging and Epigenetic Regulatory Genes Associated with Low Motility Sperm. PLoS ONE 6(6): e20280. doi:10.1371/journal.pone.0020280

**Editor:** Brian P. Chadwick, Florida State University, United States of America

**Received:** February 4, 2011; **Accepted:** April 23, 2011; **Published:** June 2, 2011

**Copyright:** © 2011 Pacheco et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The research was funded by 5R01CA121147-5 and the Superfund Basic Research Program National Institutes of Health/The National Institute of Environmental Health Sciences (NIH/NIEHS) grant P42ES013660. SP was supported by T32ES007272-17 for "Training in Environmental Pathology". The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

\* E-mail: Kim\_Boekelheide@Brown.edu

## Introduction

Traditional semen analysis measures sperm concentration, motility, morphology, and semen volume, and is acknowledged to be a poor predictor of fertility, demonstrating remarkable intra- and inter-individual variability [1,2]. Because of these limitations, effort has been devoted to developing sperm molecular biomarkers that may better and more stably reflect sperm function.

DNA methylation is the stable, covalent addition of a methyl group to cytosine that can represent response to environmental cues or exposures that may modify gene expression. Both human and animal studies indicate that abnormal sperm DNA methylation patterns are associated with subfertility, including aberrant methylation of both imprinted [3–11] and non-imprinted genes [4,12,13] in oligospermic men.

In addition to DNA methylation, significant effort is being devoted to developing human sperm mRNAs as biomarkers of infertility [14–30]. The discovery of mRNAs in mature sperm shook the long-held belief that the sole purpose of sperm was to deliver its DNA to the egg [14]. Recent evidence indicates that some of these transcripts may be intentionally transported to the oocyte to aid embryogenesis, since some sperm mRNAs are found to persist in the zygote and are functionally important [14,27,28]. In addition, remnant sperm mRNAs provide a record of the spermatogenic environment and may have clinical applications as novel biomarkers of fertility status [15–26].

In the present study, we utilized high-density array techniques to investigate the hypothesis that alterations to the pattern of sperm DNA methylation or mRNA content are associated with sperm function.

## Materials and Methods

### Ethics Statement

The Committee on the Protection of Human Subjects: Rhode Island Hospital Institutional Review Board 2 (Committee #403908) approved the study and written informed consent was obtained from all participants. Clinical investigation was conducted according to the principles expressed in the Declaration of Helsinki.

### Microarray DataSets

The microarray data discussed in this publication is MAIME compliant and the raw data has been deposited in NCBI's Gene Expression Omnibus (Edgar *et al.*, 2002) as detailed in the MGED Society website <http://www.mged.org/Workgroups/MAIME/maime.html>. This data is accessible through GEO Series accession number GSE26982 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE26982>).

### Patient Population, Semen Analysis, and Sperm Isolation

Study subjects presented for semen evaluation at Rhode Island Hospital's tertiary male reproductive health clinic. Samples were collected from 21 men with unknown fertility status and a range of semen characteristics (Table 1). During the semen analysis, morphology was scored using Kruger strict criteria and total motility was calculated as described in the WHO laboratory manual (2010) [31].

After clinical analysis the samples were divided into one quarter and three quarter aliquots for DNA and RNA isolations, respectively. Each group was processed through an optimized Percoll (GE Healthcare, Uppsala, Sweden) gradient to eliminate debris, non-sperm cells, and dead sperm [32]. Briefly, 1 ml of the

fresh semen was applied to a monolayer of 50% Percoll. After centrifugation, the upper and interface layers containing the dead sperm and other somatic contaminants were aspirated off, leaving the sperm enriched fraction. The sperm fraction was washed with phosphate buffered saline and the purified sperm samples were processed immediately for mRNA and DNA isolation.

Prior to processing the 21 samples, sperm purity was confirmed by the absence of somatic cell contaminants using bright phase microscopy and by the absence of 18/28S ribosomal RNA peaks by RNA gel electrophoresis (data not shown) [19,21].

### DNA Isolation, Bisulfite Modification, and Illumina Infinium HumanMethylation27 BeadChip Array

DNA was isolated from the sperm of the 21 men using a modified protocol in which sperm pellets were lysed for 16 hours in a solution containing Tris (Fisher Scientific, Pittsburgh, PA, USA), DTT (Promega Corporation, Madison, WI, USA), NaCl (EMD Chemicals, Inc., the North American Affiliate of Merck KGaA, Darmstadt, Germany), EDTA (Fisher Scientific, Pittsburgh, PA, USA), SDS (Fisher Scientific, Pittsburgh, PA, USA), Proteinase K (Promega Corporation, Madison, WI, USA), and beta-mercaptoethanol (Sigma-Aldrich, St. Louis, MO, USA) [33]. The DNA was then extracted using phenol/chloroform (Sigma-Aldrich, St. Louis, MO, USA), ethanol precipitated, and bisulfite modified using the EZ DNA Methylation kit (Zymo Research Corporation, Orange, CA, USA). Genome-wide scanning for DNA methylation was performed using the Illumina Infinium HumanMethylation27 BeadChip assay (Illumina, Inc., San Diego, CA, USA) to determine the methylation state at 27,578 CpG sites spanning more than 14,000 genes; and on this array, there are 616 CpGs associated with 187 imprinted genes identified using the array's annotation file (HumanMethylation27\_270596\_v.1.2, [www.Illumina.com](http://www.Illumina.com)). Multiple groups including ours have previously demonstrated the validity of Illumina methylation array data using several different approaches [34–39].

### Imprinted Genes

A list of 187 imprinted genes in the human genome was compiled based on information from three sources: (1) experimentally determined imprinted genes listed in two databases (<http://www.genemprint.com/databases/> and <http://icg.otago.ac.nz/home.html>) (n = 62); (2) imprinted genes identified using the ChIP-SNP method (n = 27) [40]; and (3) protein-coding genes from the 156 putatively imprinted sequences that correspond to known genes listed by NCBI (n = 106) [41]. Taken together, a final list of 187 imprinted genes is identified from these three sources (Table S1).

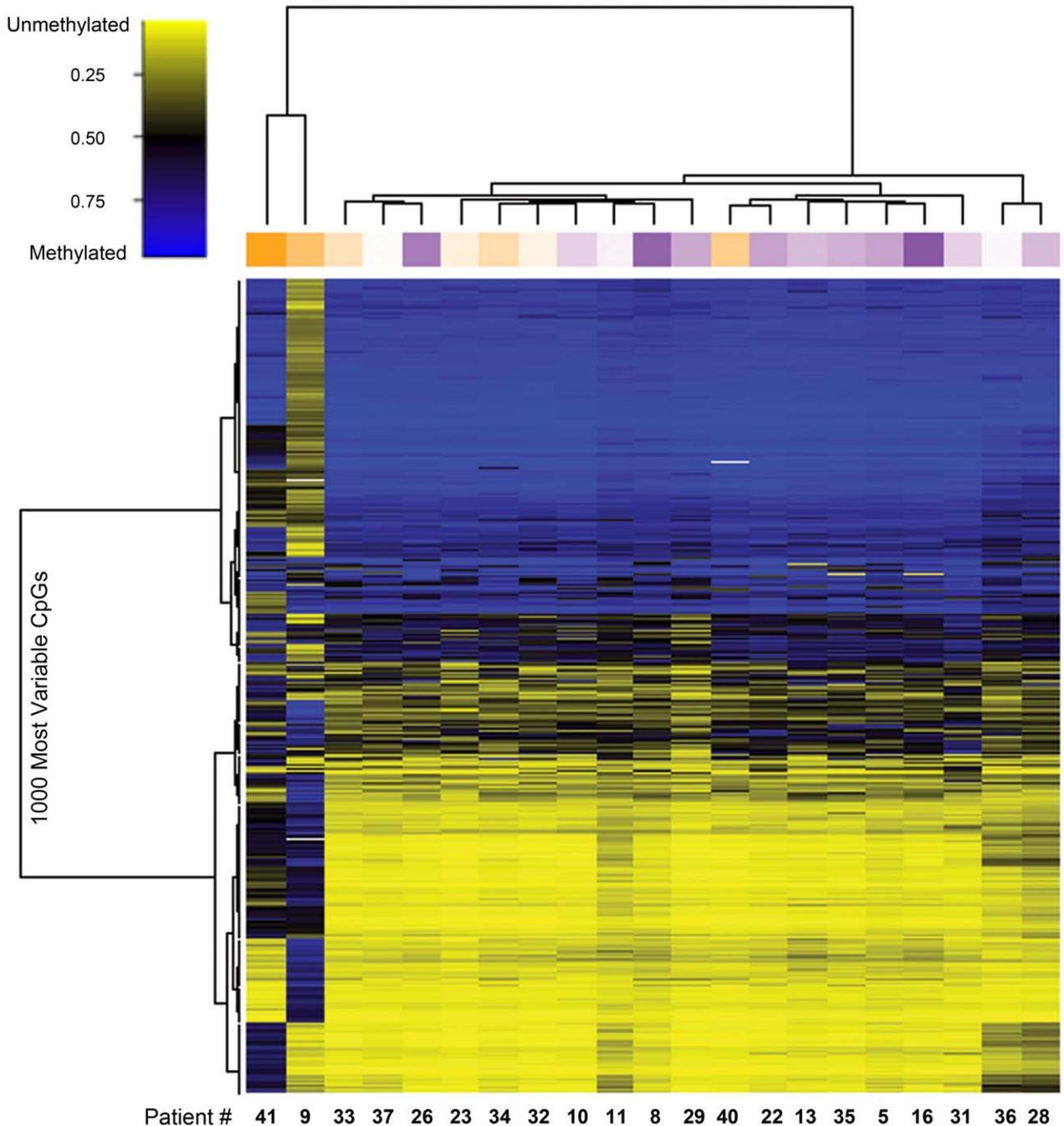
### mRNA Isolation and Affymetrix GeneChip Human Gene 1.0 ST Array

Sperm mRNA was extracted from 18 of the 21 men using a modified Stat 60 (IsoTex Diagnostics, Inc., Friendswood TX, USA) protocol in addition to components of Qiagen's RNeasy kit (Qiagen Sciences, Germantown, MD, USA). Using the Brown Genomics Core Facility, the isolated sperm mRNA was processed and hybridized to Affymetrix GeneChip Human Gene 1.0 ST Arrays (Affymetrix, Santa Clara, CA, USA), providing whole-transcript coverage of 28,869 genes by ~26 probes spread across the length of each gene. The probe cell intensity data from the Affymetrix GeneChips was normalized and annotated using Affymetrix Expression Console as recommended by the manufacturer. The application uses the RMA-Sketch workflow analysis as the default to create CHP files. The CHP log<sub>2</sub> expression files were then merged in Expression Console with the annotation file

**Table 1.** Semen Parameters of Subjects Examined.

Subject ID	Motility (%)	Morphology (%)	Count ( $\times 10^6$ /ml)
16	77	11	132
8	74	4.5	18
26	70	6	80
5	64	8	44
22	64	5.5	67
29	63	7	174
35	62	8	103
13*	60	8	26
28	60	10	72
10	57	10	58
31*	57	4	30
11	52	8	49
36	51	3	44
37	49	1	9
32	46	8.5	23
23	45	1	67
33*	41	6	1.6
34	39	4	20
40	34	4	72
9	30	3.5	7
41	21	0	15

Note: Patients with \* were excluded from the mRNA analysis due to low yield.  
doi:10.1371/journal.pone.0020280.t001



**Figure 1. Unsupervised clustering of the 1,000 most variable CpG loci average beta values.** The dendrograms above the heatmap show unsupervised clustering based on the methylation data alone, using a Euclidean metric with Ward's method of hierarchical clustering. Patients are represented by column ( $n=21$ ) and CpG loci ( $n=1000$ ) by row. Each cell represents the CpG level of methylation for one site in one sample. The methylation scale indicates the level of methylation: yellow = sample predominantly unmethylated (0–49% methylated); black = 50% of the sample is methylated; blue = sample predominantly methylated (51–100%). The column annotation track shows motility: orange = low, purple = high. doi:10.1371/journal.pone.0020280.g001

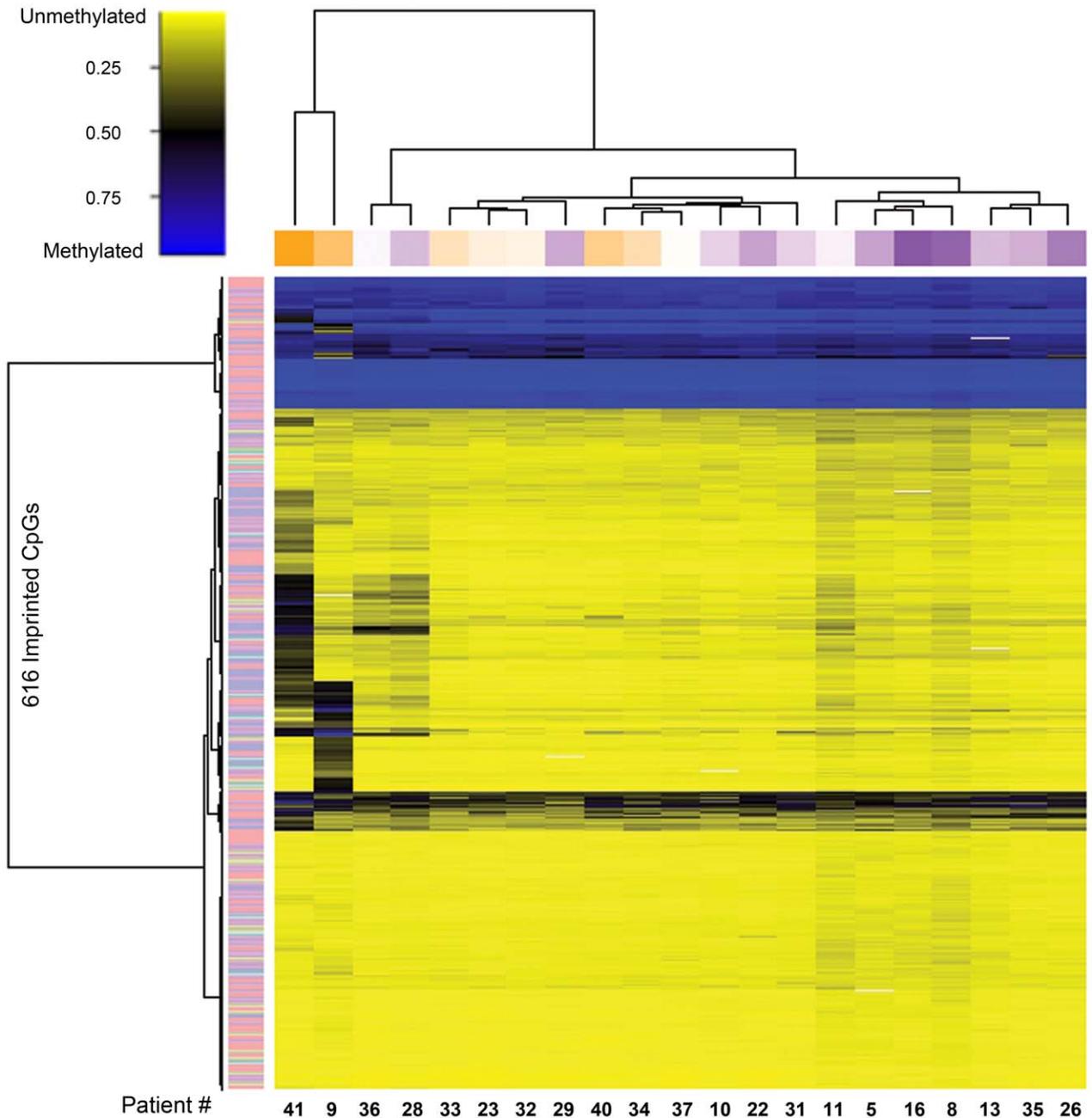
and the annotated log<sub>2</sub> results were exported as a text file for third-party downstream analysis.

### Statistical Analyses

Aside from array normalization procedures, the R software environment (R Foundation for Statistical Computing, Vienna, Austria) was used for all statistical analysis.

### Recursively Partitioned Mixture Modeling

Recursively partitioned mixture modeling (RPMM) profiles were fit to the entire Infinium array using previously described methods [42]. This method builds classes of samples based upon the similarity of methylation profiles by recursively splitting samples into parsimoniously differentiated classes. The classes are identified by pattern of branching into right (R) or left (L) arms. Permutation tests



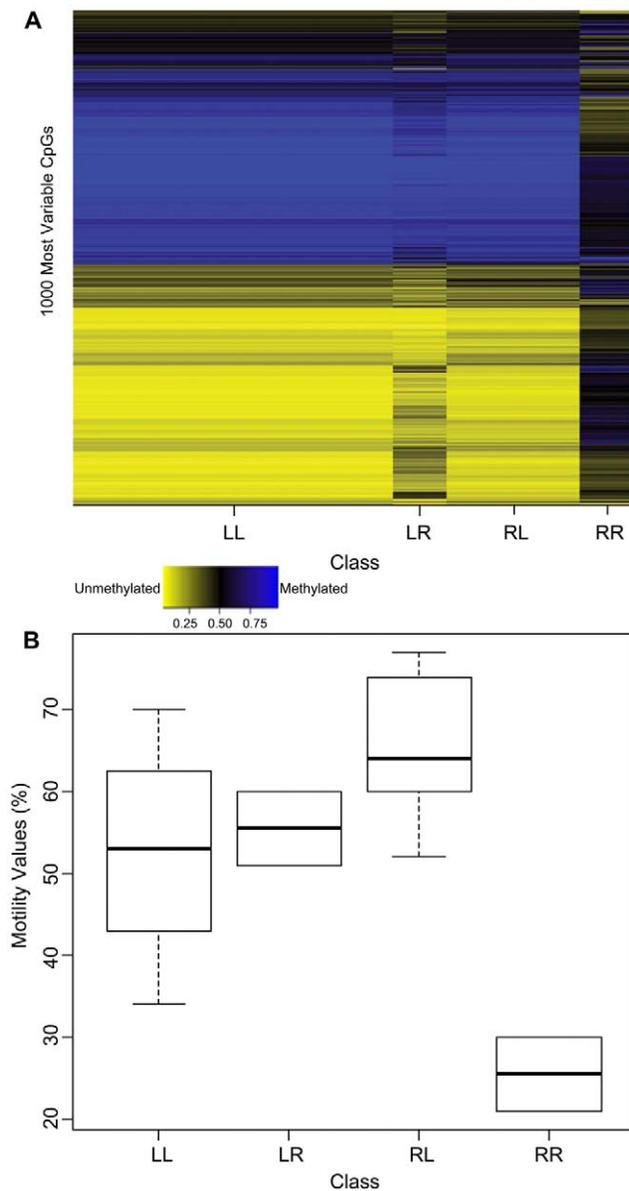
**Figure 2. Heatmap displaying the methylation status of CpG loci related to known and predicted imprinted genes.** The dendrograms above the heatmap show unsupervised clustering based on the methylation data alone, using a Euclidean metric with Ward's method of hierarchical clustering. Patients are represented by column ( $n=21$ ) and CpG loci by row ( $n=616$ ). Each cell represents the CpG level of methylation for one site in one sample. The methylation scale indicates the level of methylation: yellow = sample predominantly unmethylated (0–49% methylated); black = 50% of the sample is methylated; blue = sample predominantly methylated (51–100%). The column annotation track shows motility: orange = low, purple = high. The row annotation track shows type of imprinting: pink = maternally expressed, light blue = paternally expressed, light green = not determined.

doi:10.1371/journal.pone.0020280.g002

(5,000 permutations run with the Kruskal-Wallis [KW] test statistic) were used to test associations between RPMM class and the 3 clinical fertility variables: count, motility and morphology, using the values in Table 1. Our test statistic was the maximum of the KW test statistic, and the null distribution for this test statistic was obtained by the permutation. Semen parameters were considered significantly associated with RPMM profiles when  $P < 0.02$ , after Bonferroni correction for multiple comparisons.

### Quantitative Analysis of the DNA Methylation Status of All CpGs

The LIMMA procedure [43] (R package limma) utilized a matrix design containing the 21 samples and their corresponding percent motility values listed in Table 1 to fit a simple linear regression model for each CpG dinucleotide. This univariately tests each CpG for association between methylation and sperm motility. LIMMA results provided estimates of strength and direction of association



**Figure 3. RPMM classes.** (A) RPMM displaying 1,000 most variable CpGs by class. Each column represents a class generated by RPMM (left left (LL), left right (LR), right left (RL), and right right (RR)). The width of the column represents the number of patients distributed in that class. The rows represent the average beta values for each CpG within the class. The methylation scale indicates the level of methylation: yellow = sample predominantly unmethylated (0–49% methylated); black = 50% of the sample is methylated; blue = sample predominantly methylated (51–100%). (B) Boxplot comparing the motility values for each class.

doi:10.1371/journal.pone.0020280.g003

between CpG methylation and sperm motility and were adjusted for multiple comparisons with the *q*value package in R [44]. CpGs with positive slopes were interpreted as hypomethylated in low motility sperm and CpGs with negative slopes were interpreted as hypermethylated in low motility sperm.

#### mRNA Content Analysis of Candidate Transcripts

The transcript presence of the 276 candidate genes was tested using the same statistical strategy as the CpG analysis except here the design matrix was limited to the 18 samples with array data

and the slopes were transformed into fold change values. The Affymetrix platform yielded a dataset with ~28,000 transcripts to assess. However, sperm contain a limited transcriptome (~5000 transcripts) with few (~400) consistently expressed in sperm [22]. Therefore, we assessed 276 genes where an *a priori* hypothesis for association with subfertility existed based on previous reports. The analysis included 177 imprinted genes (10 of the 187 potential imprinted genes were not present on the Affymetrix array) as well as 99 candidate genes with biallelic expression (Table S1 and Table S2) [10,11,13,24,26,29,45–49].

#### Statistical Analysis Comparing Associations Among RPMM Classes and Candidate Genes

Associations among the RPMM classes and the normalized gene expression values for candidate transcripts were calculated with the KW test statistic utilizing the strategy employed previously. Messenger RNAs were considered significantly associated with RPMM class when  $P < 0.02$ , after adjusting for multiple comparisons using the Bonferroni correction.

## Results

### Sperm DNA Methylation Profiles Cluster by Motility

Unsupervised clustering of sperm DNA methylation data for the 1,000 most variable CpG loci on the array highlights the methylation differences among the 21 individual men (Figure 1). As shown in the column annotation track, the clustering differentiated men based upon the motility of their sperm, with high motility samples (dark purple) clustering together and low motility samples (dark orange) clustering together, with intermediate shades between. The DNA methylation of CpGs within imprinted genes is established during spermatogenesis and maintained in mature spermatozoa. In addition, several laboratories have shown alterations at imprinted loci to occur more frequently in men with sperm abnormalities [3–8,10,11]. Thus, we hypothesized that imprinted loci may be specifically targeted for aberrant methylation in low motility sperm and separately clustered the 616 CpG loci associated with the 187 imprinted genes present on the array. We observed the same overall trend, with high motility samples clustering together and low motility samples clustering together (Figure 2).

### Sperm DNA Methylation Profiles are Significantly Associated with Motility

Recursively partitioned mixture modeling (RPMM) was performed on raw methylation data to organize the sperm samples into methylation classes based on similarity. The algorithm first separated the 21 sperm profiles into two different branches left (L) and right (R) and then further subdivided each branch into right and left branches resulting in 4 total classes: left left (LL), left right (LR), right left (RL) and right right (RR) (Figure 3, A). In Figure 3 (B) we plotted methylation class-specific sperm motility values: samples in methylation class RR had the lowest median motility, and methylation class was significantly associated with motility after adjusting for multiple comparisons ( $P = 0.01$ ). The association between RPMM methylation class and sperm morphology approached statistical significance ( $P = 0.09$ ), though methylation class was not associated with sperm count ( $P = 0.29$ ).

### Thousands of CpG Loci are Significantly Altered in Low Motility Sperm

Linear models of microarray analysis (LIMMA) was used to univariately test each CpG for association with motility. 9,189 of 27,578 CpGs (34%) had significantly altered methylation associated

**Table 2.** Imprinted Genes with Aberrant DNA Methylation.

Known Imprinted				Predicted Imprinted											
Gene	E	#	MS	Gene	E	#	MS	Gene	E	#	MS	Gene	E	#	MS
<i>BMPR2</i>	N	1	–	<i>MAPK12</i>	N	2	–	<i>ALDH1L1</i>	M	1	–	<i>LMX1B</i>	M	1	–
<i>CCNE1</i>	N	1	–	<i>MEG3*</i>	M	3	+	<i>ANKRD11</i>	M	2	–	<i>LY6D</i>	P	2	+
<i>CD44</i>	N	2	–	<i>MEST*</i>	P	3	–/+	<i>APBA1</i>	P	2	–	<i>MYEOV2</i>	P	1	–
<i>CDKN1C</i>	M	3	–	<i>MKRN3</i>	P	1	+	<i>BRP44L</i>	P	1	–	<i>OTX1</i>	M	2	–
<i>COPG2</i>	P	2	–	<i>NDN</i>	P	3	+	<i>COL9A3</i>	M	2	–/+	<i>PEX10</i>	M	2	–/+
<i>CTAG2</i>	N	2	+	<i>NEDD9</i>	N	1	–	<i>DVL1</i>	M	1	–	<i>PHPT1</i>	M	1	–
<i>CTNND2</i>	N	1	–	<i>NGFB</i>	N	1	–	<i>EGFL7</i>	P	2	–	<i>PPAP2C</i>	M	2	+
<i>CYR61</i>	N	1	–	<i>NNAT</i>	P	5	+	<i>FAM59A</i>	P	2	–	<i>PRDM16</i>	P	1	–
<i>DIRAS3*</i>	P	11	+	<i>PCNA</i>	N	2	–	<i>FAM70B</i>	M	1	–	<i>PTPN14</i>	M	1	–
<i>DLK1</i>	P	1	+	<i>PEG10</i>	P	8	–/+	<i>FASTK</i>	M	1	–	<i>PURG</i>	P	2	–
<i>DLX5</i>	M	3	–/+	<i>PHLDA2</i>	M	4	–	<i>FOXF1</i>	M	1	–	<i>PYY2</i>	P	2	+
<i>GABRA5</i>	P	1	–	<i>PLAGL1*</i>	P	5	+	<i>GATA3</i>	P	1	–	<i>RPL22</i>	P	1	–
<i>GFI1</i>	P	1	–	<i>SDHD</i>	P	1	+	<i>HES1</i>	P	1	–	<i>SALL1</i>	M	1	–
<i>GRB10</i>	N	5	–/+	<i>SGCE</i>	P	2	+	<i>HIST3H2BB</i>	M	1	–	<i>SLC4A2</i>	P	2	–
<i>H19*</i>	M	4	–/+	<i>SHANK2</i>	M	1	–	<i>HOXA11</i>	M	1	–	<i>SOX8</i>	P	1	–
<i>HYMAI</i>	P	1	–	<i>SLC22A18</i>	M	7	–/+	<i>HOXA5</i>	M	1	+	<i>TIGD1</i>	P	1	–
<i>IGF2*</i>	P	1	–	<i>SNRPN*</i>	P	6	+	<i>HOXB2</i>	M	1	–	<i>TMEM60</i>	P	2	–
<i>IGF2AS</i>	P	4	–	<i>TCEB3C</i>	M	3	+	<i>HOXB3</i>	M	1	+	<i>VAX2</i>	M	1	–
<i>IL1B</i>	N	2	–	<i>TP73</i>	M	2	–/+	<i>HOXC4</i>	M	1	–	<i>WDR8</i>	M	1	–
<i>ILK</i>	N	2	–	<i>UBE3A</i>	M	1	–	<i>HSPA6</i>	M	1	+	<i>ZFP36L2</i>	M	1	–
<i>KCNQ1DN</i>	M	2	+	<i>WT1</i>	P	9	–/+	<i>IFITM1</i>	M	1	–	<i>ZNF550</i>	P	1	–
<i>L3MBTL</i>	P	2	+	<i>ZIM2</i>	P	5	+	<i>LDB1</i>	M	1	–				
<i>LASS4</i>	N	2	–	<i>ZNF264</i>	M	3	–/+								
<i>LMO1</i>	N	1	–	<i>ZNF331</i>	P	1	–								
<i>MAGEL2</i>	P	3	+												

Note: E = parent with expressed allele; # = number of significantly altered loci in low motility samples; MS = methylation status of the loci; M = maternally expressed; P = paternally expressed; N = parent of origin not determined; – = loci hypomethylated in low motility samples; + = loci hypermethylated in low motility samples; –/+ = more than two loci were altered and some of the loci were hypomethylated and some of were hypermethylated. Genes with \* have been previously reported differentially methylated in sperm.

doi:10.1371/journal.pone.0020280.t002

with motility after adjusting for multiple comparisons ( $Q < 0.05$ ) (Table S3). Of these, 1,827 CpGs (20%) were hypermethylated in the low motility samples, whereas 7,362 CpGs (80%) were hypomethylated.

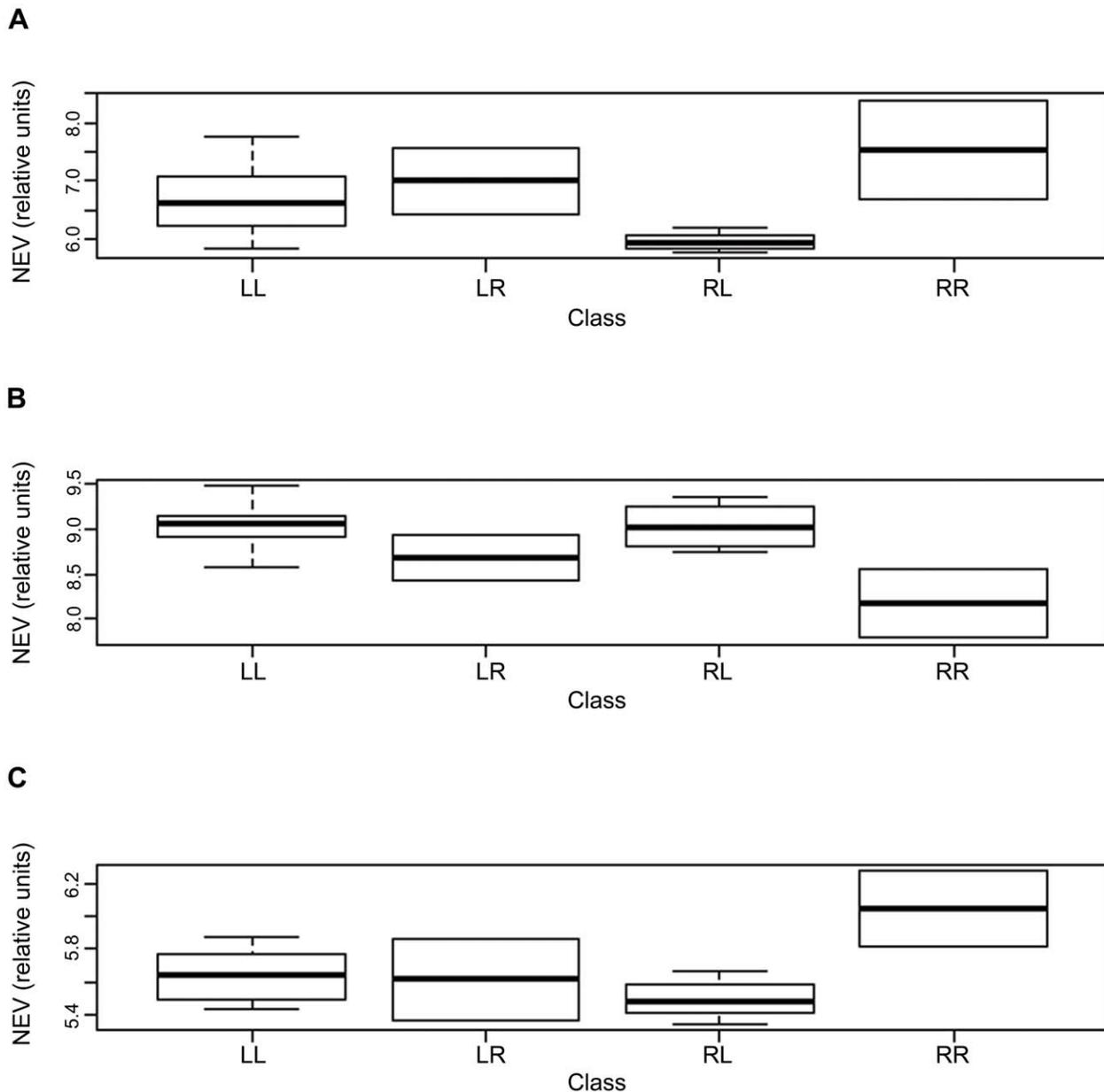
Because establishing proper methylation marks within imprinted genes during spermatogenesis is critical, we next restricted our analysis to CpGs associated with imprinted genes. Of the 616 CpGs associated with imprinted genes, 194 CpGs (31.5%) had significant associations with motility, similar to the distribution of the array overall. Amongst these loci, 47% ( $n = 92$ ) were hypermethylated in the low motility samples, whereas 53% ( $n = 102$ ) were hypomethylated. The majority of hypomethylated CpGs were on maternally expressed genes (45%), followed by paternally expressed (33%) and those with undetermined parent of expression (22%). Conversely, the majority of hypermethylated CpGs were associated with paternally expressed genes (70%), with the remainder maternally expressed (26%), and of undetermined parental expression (4%). The 194 loci corresponded to 92 genes, with 11 genes showing both hyper- and hypomethylated loci (Table 2).

Aberrant promoter methylation in genes related to spermatogenesis and epigenetic regulation have recently been identified in sperm from men with poor semen quality [3–13]. Thus, we next performed an analysis restricted to array CpGs associated with

genes related to spermatogenesis and epigenetic regulation. Of the 147 CpGs on the array associated with genes involved in spermatogenesis, 39% ( $n = 58$ ) were significantly altered in low motility sperm (similar to the 34% of CpGs associated with low motility in array-wide tests, Table 2). Among these 58 CpG loci, 71% ( $n = 41$ ) were hypomethylated and 29% ( $n = 17$ ) were hypermethylated in low motility samples. There were 50 CpG loci associated with epigenetic regulatory genes identified on the array, and only 26% ( $n = 13$ ) had significantly altered methylation in low motility sperm samples. Of these, 61.5% ( $n = 8$ ) were hypomethylated and 38.5% ( $n = 5$ ) were hypermethylated (Table 3).

### mRNA Content is Altered in Low Motility Sperm

Focusing on imprinted mRNAs and candidate biallelic mRNAs, LIMMA analysis was performed to identify differentially expressed transcripts, conditioning on motility. Twenty genes were identified as significant after adjusting for false discovery rate ( $Q < 0.05$ ) (Table S4). These included 11 imprinted genes (*GLI3* (NCBI 2737), *APAB1* (NCBI 320), *CTNND2* (NCBI 1501), *FERMT2* (NCBI 10979), *PHPT1* (NCBI 29085), *SNRPN* (NCBI 6638), *PPP1R9A* (NCBI 55607), *CDH18* (NCBI 603019), *ALDH1L1* (NCBI 10840), *LDB1* (NCBI 8861), and *PEX10* (NCBI 5192)), six



**Figure 4. Boxplots comparing DNA methylation profiles and gene expression values for the 3 epigenetic regulators.** Each panel of boxplots (A–C) compares expression data for a particular gene: (A) *HDAC1*; (B) *SIRT3*; and (C) *DNMT3A*. The x-axis represents RPMM classes. The y-axis represents normalized gene expression values (NEV). doi:10.1371/journal.pone.0020280.g004

genes associated with spermatogenesis (*SERPINA5* (NCBI 5104), *ACE* (NCBI 1636), *FANCC* (NCBI 2176), *PCSK4* (NCBI 57460), *CYP19A1* (NCBI 1588), and *FAS* (NCBI 355)), and three epigenetic regulatory genes (*HDAC1*, *DNMT3A*, and *SIRT3*). *HDAC1*, *DNMT3A*, *LDB1* and *FAS* showed increased mRNA content in the low motility samples, whereas the remaining 16 showed decreased mRNA content.

#### Integration of Epigenetic and Transcript Data

It is known that major modifications in chromatin organization occur in spermatid nuclei during spermatogenesis, leading to the high degree of packaging in the sperm head. Chromatin compaction ensues when the histones surrounding the DNA are replaced by protamines, and this occurs in parallel with

transcriptional arrest [45]. Therefore, nuclear packaging and transcript content are interrelated. To determine whether altered expression of epigenetic regulatory genes was associated with methylation profiles we plotted the methylation class-specific gene expression values for the three epigenetic regulatory genes (*HDAC1*, *SIRT3*, and *DNMT3A*) with significantly altered expression in low motility sperm (Figure 4). Among methylation classes, expression values for *HDAC1*, *SIRT3*, and *DNMT3A* were most altered in class RR, the class with lowest motility sperm (increased expression for *HDAC1* and *DNMT3A*, and decreased expression for *SIRT3*). For all three genes, the association between mRNA expression level and methylation class membership approached significance after adjusting for multiple comparisons (*HDAC1*,  $P=0.03$ ; *SIRT3*,  $P=0.06$ ; and *DNMT3A*,  $P=0.07$ ).

**Table 3.** Genes Associated with Spermatogenesis and Epigenetic Regulation with Aberrant DNA Methylation.

Gene	#	MS	Gene	#	MS	Gene	#	MS	Gene	#	MS
ADAM2	1	+	CSF1	2	–	KIT	1	–	SIRT5	1	–
ADAMTS2	1	–	CTCF	1	–	LIMK2	1	–	SIRT7	2	–
AR	4	–/+	CTCF	1	+	MLH1	2	+	SLC12A2	1	–
ATM	5	–/+	DAZL*	1	+	MORC1	2	+	SPO11	2	+
BCL2	2	–/+	DDX4	1	+	MSH5	1	–	STRBP	1	–
BCL2L2	1	–	DHH	1	–	MTHFR*	1	+	STYX	1	–
BCL6	2	–	DMC1	1	–	PCSK4	1	–	SYCP3	2	+
BRDT	1	+	DNMT3A	1	–	PIK3CG	1	–	TBPL1	1	–
BMP8B	1	+	DNMT3B	1	+	PMS2	2	–	TERT	1	–
BSG	1	–	EGR4	2	–/+	RXRβ	1	–	TUSC2	2	–
CCNA1	1	–	HCLS1	1	–	SIAH1	2	–	VDAC3	1	+
CDKN2C	1	–	HDAC2	1	–	SIRT1	2	–/+			
CREM	2	–	INPP5B	1	–	SIRT2	1	–			

Note: # = Number of significantly altered loci in low motility samples; MS = methylation status of the loci; – = loci hypomethylated in low motility samples; + = loci hypermethylated in low motility samples; –/+ = more than two loci were altered and some of the loci were hypomethylated and some of were hypermethylated. Genes with \* have been previously reported differentially methylated in sperm.  
doi:10.1371/journal.pone.0020280.t003

## Discussion

Currently, the evaluation of male infertility relies upon physical exam and semen and hormone analyses; although quick and relatively inexpensive, these physiologic measurements often do not explain the underlying cause of infertility nor predict the usefulness of various therapeutic interventions. Therefore, new approaches are needed to identify the etiologies of male infertility. Recent data suggest that sperm DNA methylation abnormalities and alterations in sperm mRNA content are found in infertile men [3–8,10,11,13–30,50]. Here we extend these studies by performing integrative analysis of sperm DNA methylation and mRNA content using genome-wide approaches to identify significant associations among these profiles and semen parameters.

Due to the unreliable nature of classifying men into abnormal and normal groups during a semen analysis, we used a data driven approach to first qualitatively assess associations among sperm DNA methylation and our patient population. Unsupervised clustering indicated that there was an association between DNA methylation and motility status. This was true both for all of the CpGs on the array and the imprint-only subset.

RPM separated the 21 men into four classes based on similarity of DNA methylation array data. The median motility values were calculated for each class and the results suggested that the methylation profiles were associated with motility. Comparing the DNA methylation heatmap to the class versus motility boxplot indicates that the low motility class has the most aberrantly methylated CpGs. Overall, these data suggest that low motility sperm have increased hypomethylation relative to high motility sperm. We used LIMMA to identify the significantly altered CpGs conditioned on changes in motility for all CpGs on the array: over one-third of the CpGs (and almost half of the genes represented on the array) were significantly differentially methylated in the low motility samples and the majority of these were hypomethylated. The high prevalence of aberrantly methylated CpGs suggests a genome-wide DNA methylation defect in the low motility sperm. It has been previously hypothesized that the aberrant sperm DNA methylation could be due to abnormal chromatin compaction, inefficient DNA methyltransferases, and/or failure to maintain or

acquire the correct methylation marks during spermatogenesis and our results are consistent with this literature [11–13,29].

We initially focused on CpGs mapping to imprinted genes because of their plasticity during spermatogenesis, biological relevance following conception and development, and because previous studies have identified imprinted loci as aberrantly methylated in abnormal sperm [3–8]. In our data, the distribution of significantly hyper- and hypomethylated imprinted loci was nearly equal. Expanding the imprinting analysis to the gene level identified 92 genes with altered CpG methylation, seven of which (*DIRAS3* (NCBI 9077), *H19* (NCBI 283120), *IGF2* (NCBI 3481), *MEST/PEG1* (NCBI 4232), *PLAGL1/ZAC* (NCBI 5325), *MEG3/GTL2* (NCBI 55384), and *SNRPN*) have already been noted as aberrantly methylated in abnormal sperm [3–8,10,11]. The methylation status of two genes, (*PEG3* (NCBI 5178) and *LIT1/KCNQ1OT1* (NCBI 10984)) has been inconsistently reported in the literature [3,8,10]. We observed no statistical differences for these genes between the low and high motility sperm, which is consistent with the results published by Sato, et al. [10]. In fact, our study confirmed all of the DNA methylation results reported in the aforementioned study.

To further clarify the potential functional alterations to imprinted genes and critical epigenetic regulatory genes, we evaluated sperm mRNA content of 177 imprinted genes and 99 other transcripts where an *a priori* hypothesis for association with male subfertility or epigenetic regulation exists. Twenty genes were identified as demonstrating significantly altered transcript levels in low motility sperm. All of the mRNAs except *HDAC1*, *DNMT3A*, *LBD1*, and *FAS* were present in decreased amounts in low motility sperm, and we did not observe altered mRNA content for *BRDT*, which was previously reported to have increased expression in subfertile patients [29].

Integration of epigenetic and expression data revealed a relationship between transcript content of three epigenetic regulatory genes (*HDAC1*, *SIRT3*, and *DNMT3A*) and methylation class. *HDAC1* is the predominant histone deacetylase (HDAC) during spermatogenesis. Histone hyperacetylation is required for the histone to protamine exchange and is facilitated by the degradation of HDAC1 in elongated spermatids [51]. If HDAC1

is in excess, one could hypothesize that the histones are not being replaced by protamines, leading to an “immature” sperm chromatin structure, with less compact DNA. Therefore, incomplete or incorrect nuclear compaction may influence overall sperm maturation and be reflected in the physiological endpoint of motility.

*SIRT3* is a class III histone deacetylase and this HDAC family is similar to the yeast Sir2 protein which has been associated with chromatin silencing and also plays roles in cellular metabolism and aging [46]. In mammals, however, *SIRT3* is targeted to the mitochondria and functions to induce the expression of the antioxidant MnSOD to eliminate reactive oxygen species (ROS) generated during oxidative phosphorylation [52]. Recent studies have found that increased ROS in sperm have deleterious effects on sperm motility parameters which ultimately have adverse effects on fertility [53]. Therefore, the decrease in *SIRT3* mRNA in the low motility sperm may reflect reduced MnSOD and increased intracellular ROS during spermatogenesis, leading to a diminished fertility potential.

The literature also suggests that oxidative stress itself can impede the process of DNA methylation, resulting in a hypomethylated phenotype [54]. Interestingly, we observed global hypomethylation in the low motility sperm even though we saw increased DNMT3A transcript presence in the low motility sperm. Because DNMT3A is the DNA methyltransferase responsible for *de novo* methylation, our data suggests a failure of the low motility sperm to acquire the proper methylation patterns.

Although we were limited by sample size, we used a powerful integrative approach to simultaneously examine sperm DNA methylation and mRNA content utilizing two high density array techniques. We found that: (1) low motility sperm have genome-wide DNA hypomethylation that may be due to a failure of the sperm to complete chromatin compaction properly because of

increased *HDAC1* presence; (2) low motility sperm have reduced *SIRT3* mRNA content which might be related to increased subcellular ROS during spermatogenesis leading to the abnormal motility phenotype; and (3) this oxidative stress may be impeding the ability of DNMT3A to set the correct methylation marks which would also contribute to the hypomethylated phenotype. Our results suggest that additional integrative studies including larger sample sizes as well as prospective studies of fertility following these integrated molecular assessments have great potential to advance our understanding of the molecular features of sperm associated with fertility status.

## Supporting Information

### Table S1 Imprinted Genes.

(DOC)

### Table S2 Genes Associated with Spermatogenesis and Epigenetic Regulation.

(DOC)

### Table S3 Aberrant CpGs in Low Motility Sperm.

(DOC)

### Table S4 Aberrant mRNA Transcripts in Low Motility Sperm.

(DOC)

## Author Contributions

Conceived and designed the experiments: SEP EAH BCC CJM KTK MS KB. Performed the experiments: SEP BCC. Analyzed the data: SEP EAH BCC CJM KTK MS KB. Contributed reagents/materials/analysis tools: SEP EAH BCC CJM KTK MS KB. Wrote the paper: SEP EAH BCC CJM KTK MS KB.

## References

- Leushuis E, van der Steeg JW, Steures P, Repping S, Bossuyt PM, et al. (2010) Reproducibility and reliability of repeated semen analyses in male partners of subfertile couples. *Fertil Steril*;S0015-0282(10)00458-9 [pii] 10.1016/j.fertnstert.2010.03.021.
- Swan SH (2006) Semen quality in fertile US men in relation to geographical area and pesticide exposure. *Int J Androl* 29: 62–68; discussion 105–108. IJA620 [pii] 10.1111/j.1365-2605.2005.00620.x.
- Hammoud SS, Purwar J, Plueger C, Cairns BR, Carrell DT (2009) Alterations in sperm DNA methylation patterns at imprinted loci in two classes of infertility. *Fertil Steril*;S0015-0282(09)03689-9 [pii] 10.1016/j.fertnstert.2009.09.010.
- Houshdaran S, Cortessis VK, Siegmund K, Yang A, Laird PW, et al. (2007) Widespread epigenetic abnormalities suggest a broad DNA methylation erasure defect in abnormal human sperm. *PLoS ONE* 2: e1289. 10.1371/journal.pone.0001289.
- Marques CJ, Carvalho F, Sousa M, Barros A (2004) Genomic imprinting in disruptive spermatogenesis. *Lancet* 363: 1700–1702. 10.1016/S0140-6736(04)16256-9S0140-6736(04)16256-9 [pii].
- Marques CJ, Costa P, Vaz B, Carvalho F, Fernandes S, et al. (2008) Abnormal methylation of imprinted genes in human sperm is associated with oligozoospermia. *Mol Hum Reprod* 14: 67–74. gam093 [pii] 10.1093/molehr/gam093.
- Poplinski A, Tuttelmann F, Kanber D, Horsthenke B, Gromoll J (2009) Idiopathic male infertility is strongly associated with aberrant methylation of MEST and IGF2/H19 ICR1. *Int J Androl*;IJA1000 [pii] 10.1111/j.1365-2605.2009.01000.x.
- Kobayashi H, Sato A, Otsu E, Hiura H, Tomatsu C, et al. (2007) Aberrant DNA methylation of imprinted loci in sperm from oligospermic patients. *Hum Mol Genet* 16: 2542–2551. ddm187 [pii] 10.1093/hmg/ddm187.
- Pathak S, Kedia-Mokashi N, Saxena M, D'Souza R, Maitra A, et al. (2008) Effect of tamoxifen treatment on global and insulin-like growth factor 2-H19 locus-specific DNA methylation in rat spermatozoa and its association with embryo loss. *Fertil Steril*;S0015-0282(08)01484-2 [pii] 10.1016/j.fertnstert.2008.07.1709.
- Sato A, Hiura H, Okae H, Miyauchi N, Abe Y, et al. (2011) Assessing loss of imprint methylation in sperm from subfertile men using novel methylation polymerase chain reaction Luminex analysis. *Fertil Steril* 95: 129–134. 134 e121–124. S0015-0282(10)01029-0 [pii] 10.1016/j.fertnstert.2010.06.076.
- Boissonnas CC, Abdalaoui HE, Haelewyn V, Fauque P, Dupont JM, et al. (2010) Specific epigenetic alterations of IGF2-H19 locus in spermatozoa from infertile men. *Eur J Hum Genet* 18: 73–80. ehg2009117 [pii] 10.1038/ejhg.2009.117.
- Wu W, Shen O, Qin Y, Niu X, Lu C, et al. (2010) Idiopathic male infertility is strongly associated with aberrant promoter methylation of methylenetetrahydrofolate reductase (MTHFR). *PLoS ONE* 5: e13884. 10.1371/journal.pone.0013884.
- Navarro-Costa P, Nogueira P, Carvalho M, Leal F, Cordeiro I, et al. (2010) Incorrect DNA methylation of the DAZL promoter CpG island associates with defective human sperm. *Hum Reprod* 25: 2647–2654. deq200 [pii] 10.1093/humrep/deq200.
- Krawetz SA (2005) Paternal contribution: new insights and future challenges. *Nat Rev Genet* 6: 633–642.
- Lalancette C, Miller D, Li Y, Krawetz SA (2008) Paternal contributions: new functional insights for spermatozoal RNA. *J Cell Biochem* 104: 1570–1579. 10.1002/jcb.21756.
- Lalancette C, Platts AE, Johnson GD, Emery BR, Carrell DT, et al. (2009) Identification of human sperm transcripts as candidate markers of male fertility. *J Mol Med* 87: 735–748. 10.1007/s00109-009-0485-9.
- Miller D (1997) RNA in the ejaculate spermatozoon: a window into molecular events in spermatogenesis and a record of the unusual requirements of haploid gene expression and post-meiotic equilibration. *Mol Hum Reprod* 3: 669–676.
- Miller D (2007) Spermatozoal RNA as reservoir, marker and carrier of epigenetic information: implications for cloning. *Reprod Domest Anim* 42 Suppl 2: 2–9. RDA883 [pii] 10.1111/j.1439-0531.2007.00883.x.
- Miller D, Ostermeier GC (2006) Towards a better understanding of RNA carriage by ejaculate spermatozoa. *Hum Reprod Update* 12: 757–767. dml037 [pii] 10.1093/humupd/dml037.
- Miller D, Ostermeier GC (2006) Spermatozoal RNA: Why is it there and what does it do? *Gynecol Obstet Fertil* 34: 840–846. S1297-9589(06)00232-3 [pii] 10.1016/j.gyobfe.2006.07.013.
- Miller D, Ostermeier GC, Krawetz SA (2005) The controversy, potential and roles of spermatozoal RNA. *Trends Mol Med* 11: 156–163. S1471-4914(05)00048-1 [pii] 10.1016/j.molmed.2005.02.006.

22. Ostermeier GC, Dix DJ, Miller D, Khatri P, Krawetz SA (2002) Spermatozoal RNA profiles of normal fertile men. *Lancet* 360: 772–777. S0140-6736(02)09899-9 [pii] 10.1016/S0140-6736(02)09899-9.
23. Ostermeier GC, Miller D, Huntriss JD, Diamond MP, Krawetz SA (2004) Reproductive biology: delivering spermatozoan RNA to the oocyte. *Nature* 429: 154. 10.1038/429154a429154a [pii].
24. Steger K, Wilhelm J, Konrad L, Staff T, Greb R, et al. (2008) Both protamine-1 to protamine-2 mRNA ratio and Bcl2 mRNA content in testicular spermatids and ejaculated spermatozoa discriminate between fertile and infertile men. *Hum Reprod* 23: 11–16. dem363 [pii] 10.1093/humrep/dem363.
25. Wang H, Zhou Z, Xu M, Li J, Xiao J, et al. (2004) A spermatogenesis-related gene expression profile in human spermatozoa and its potential clinical applications. *J Mol Med* 82: 317–324. 10.1007/s00109-004-0526-3.
26. Aoki VW, Liu L, Carrell DT (2006) A novel mechanism of protamine expression deregulation highlighted by abnormal protamine transcript retention in infertile human males with sperm protamine deficiency. *Mol Hum Reprod* 12: 41–50. gah258 [pii] 10.1093/molehr/gah258.
27. Avendano C, Franchi A, Jones E, Oehninger S (2009) Pregnancy-specific {beta}-1-glycoprotein 1 and human leukocyte antigen-E mRNA in human sperm: differential expression in fertile and infertile men and evidence of a possible functional role during early development. *Hum Reprod* 24: 270–277. den381 [pii] 10.1093/humrep/den381.
28. Kempisty B, Antosik P, Bukowska D, Jackowska M, Lianeri M, et al. (2008) Analysis of selected transcript levels in porcine spermatozoa, oocytes, zygotes and two-cell stage embryos. *Reprod Fertil Dev* 20: 513–518. RD07211 [pii].
29. Steilmann C, Cavalcanti MC, Bartkuhn M, Pons-Kuhnemann J, Schuppe HC, et al. (2010) 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* 140: 435–443. REP-10-0139 [pii] 10.1530/REP-10-0139.
30. Bonaparte E, Moretti M, Colpi GM, Nerva F, Contalbi G, et al. (2010) ESX1 gene expression as a robust marker of residual spermatogenesis in azoospermic men. *Hum Reprod* 25: 1398–1403. deq074 [pii] 10.1093/humrep/deq074.
31. World Health Organization DoRHaR (2010) WHO laboratory manual for the examination and processing of human semen. 287 p.
32. Griffin DK, Abruzzo MA, Millie EA, Shecan LA, Feingold E, et al. (1995) Non-disjunction in human sperm: evidence for an effect of increasing paternal age. *Hum Mol Genet* 4: 2227–2232.
33. Doerksen T, Benoit G, Trasler JM (2000) Deoxyribonucleic acid hypomethylation of male germ cells by mitotic and meiotic exposure to 5-azacytidine is associated with altered testicular histology. *Endocrinology* 141: 3235–3244.
34. Bibikova M, Lin Z, Zhou L, Chudin E, Garcia EW, et al. (2006) High-throughput DNA methylation profiling using universal bead arrays. *Genome Res* 16: 383–393. gr.4410706 [pii] 10.1101/gr.4410706.
35. Byun HM, Siegmund KD, Pan F, Weisenberger DJ, Kanel G, et al. (2009) Epigenetic profiling of somatic tissues from human autopsy specimens identifies tissue- and individual-specific DNA methylation patterns. *Hum Mol Genet* 18: 4808–4817. ddp445 [pii] 10.1093/hmg/ddp445.
36. Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, et al. (2009) Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet* 5: e1000602. 10.1371/journal.pgen.1000602.
37. Christensen BC, Kelsey KT, Zheng S, Houseman EA, Marsit CJ, et al. (2010) Breast cancer DNA methylation profiles are associated with tumor size and alcohol and folate intake. *PLoS Genet* 6: e1001043. 10.1371/journal.pgen.1001043.
38. Irizarry RA, Ladd-Acosta C, Carvalho B, Wu H, Brandenburg SA, et al. (2008) Comprehensive high-throughput arrays for relative methylation (CHARM). *Genome Res* 18: 780–790. gr.7301508 [pii] 10.1101/gr.7301508.
39. Ladd-Acosta C, Pevsner J, Sabuncyan S, Yolken RH, Webster MJ, et al. (2007) DNA methylation signatures within the human brain. *Am J Hum Genet* 81: 1304–1315. S0002-9297(07)63779-3 [pii] 10.1086/524110.
40. Maynard ND, Chen J, Stuart RK, Fan JB, Ren B (2008) Genome-wide mapping of allele-specific protein-DNA interactions in human cells. *Nat Methods* 5: 307–309. nmeth.1194 [pii] 10.1038/nmeth.1194.
41. Luedi PP, Dietrich FS, Weidman JR, Bosko JM, Jirtle RL, et al. (2007) Computational and experimental identification of novel human imprinted genes. *Genome Res* 17: 1723–1730. gr.6584707 [pii] 10.1101/gr.6584707.
42. Houseman EA, Christensen BC, Yeh RF, Marsit CJ, Karagas MR, et al. (2008) Model-based clustering of DNA methylation array data: a recursive-partitioning algorithm for high-dimensional data arising as a mixture of beta distributions. *BMC Bioinformatics* 9: 365. 1471-2105-9-365 [pii] 10.1186/1471-2105-9-365.
43. Smyth GK, Yang YH, Speed T (2003) Statistical issues in cDNA microarray data analysis. *Methods Mol Biol* 224: 111–136. 1-59259-364-X-111 [pii] 10.1385/1-59259-364-X:111.
44. Storey JD (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B* 64: 479–498.
45. Dadoune JP, Siffroi JP, Alfonsi MF (2004) Transcription in haploid male germ cells. *Int Rev Cytol* 237: 1–56. 10.1016/S0074-7696(04)37001-4S0074769604370014 [pii].
46. Gray SG, Ekstrom TJ (2001) The human histone deacetylase family. *Exp Cell Res* 262: 75–83. 10.1006/excr.2000.5080S0014-4827(00)95080-8 [pii].
47. Loukinov DI, Pugacheva E, Vatolin S, Pack SD, Moon H, et al. (2002) BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma. *Proc Natl Acad Sci U S A* 99: 6806–6811. 10.1073/pnas.0921236999/10/6806 [pii].
48. Trasler JM (2009) Epigenetics in spermatogenesis. *Mol Cell Endocrinol* 306: 33–36. S0303-7207(09)00037-9 [pii] 10.1016/j.mce.2008.12.018.
49. Matzuk MM, Lamb DJ (2002) Genetic dissection of mammalian fertility pathways. *Nat Cell Biol* 4 Suppl: s41–49. 10.1038/ncb-nm-fertilityS41.
50. Marques CJ, Francisco T, Sousa S, Carvalho F, Barros A, et al. (2010) Methylation defects of imprinted genes in human testicular spermatozoa. *Fertil Steril* 94: 585–594. S0015-0282(09)00470-1 [pii] 10.1016/j.fertnstert.2009.02.051.
51. Choi E, Han C, Park I, Lee B, Jin S, et al. (2008) A novel germ cell-specific protein, SHIP1, forms a complex with chromatin remodeling activity during spermatogenesis. *J Biol Chem* 283: 35283–35294. M805590200 [pii] 10.1074/jbc.M805590200.
52. Schumacker PT (2010) A tumor suppressor SIRTainty. *Cancer Cell* 17: 5–6. S1535-6108(09)00436-X [pii] 10.1016/j.ccr.2009.12.032.
53. du Plessis SS, McAllister DA, Luu A, Savia J, Agarwal A, et al. (2010) Effects of H<sub>2</sub>O<sub>2</sub> exposure on human sperm motility parameters, reactive oxygen species levels and nitric oxide levels. *Andrologia* 42: 206–210. AND980 [pii] 10.1111/j.1439-0272.2009.00980.x.
54. Tunc O, Tremellen K (2009) Oxidative DNA damage impairs global sperm DNA methylation in infertile men. *J Assist Reprod Genet* 26: 537–544. 10.1007/s10815-009-9346-2.