Environ Health Perspect

DOI: 10.1289/EHP12013

Supplemental Material

The Association between Long-Term DDT or DDE Exposures and an Altered Sperm Epigenome—a Cross-Sectional Study of Greenlandic Inuit and South African VhaVenda Men

Ariane Lismer, Xiaojian Shao, Marie-Charlotte Dumargne, Christine Lafleur, Romain Lambrot, Donovan Chan, Gunnar Toft, Jens Peter Bonde, Amanda J. MacFarlane, Riana Bornman, Natalie Aneck-Hahn, Sean Patrick, Janice M. Bailey, Christiaan de Jager, Vanessa Dumeaux, Jacquetta M. Trasler, and Sarah Kimmins

Table of Contents

Figure S3. Identification of predicted persistent DNAme regions from sperm to the pre-implantation embryo.

Figure S4. Genic and transposable element characterization of differentially enriched H3K4me3 peaks.

Figure S5. Examples of peaks with deH3K4m3 that overlap promoters, transposable elements and / or putative enhancers.

Figure S6. Identification of predicted persistent H3K4me3 peaks from sperm to the pre-implantation embryo.

Figure S7. Promoters marked by H3K4me3 in sperm and pre-implantation embryos correspond to expressed genes in the embryo.

Additional File- Excel Document

Table S1: Aggregated population data of Greenlandic men from this study (INUENDO cohort; m = 47; Figure S1A; mean +/- SEM or %).

	mean (± SEM) or %
Age (years)	31 ± 0.949
Body mass index (BMI)	26.268 ± 0.487
Smoking (%)	57.44%
Cotinine (ng / mL)	138.878 ± 22.682
DNA fragmentation index (DFI; %)	9.418 ± 0.716
***p,p'*-DDE (ng / mL)**	870.734 ± 134.030

Table S1 Notes: Greenlandic Inuit blood and semen paired samples were selected from the biobank of the INUENDO cohort. The subjects ranged from 20 to 44 years of age (mean age of 31 years), were born in Greenland and all had proven fertility with confirmation of a pregnant partner. Sample selection was based on obtaining a range in p,p'-DDE serum levels (mean 870.734 ± 134.030 ng/mL, Figure S1) (n = 47 selected from cohort of 193 total for MCC-seq). Participants were recruited between May 2002 and February 2004 and eligible men were born in Greenland. Full details on recruitment and the cohort have been previously described. Data was available on smoking (questionnaire and cotinine levels), DNA fragmentation index and Body Mass Index (BMI). Note for adherence with General Data Protection Regulations (GDPR), individual data cannot be published. Semen samples from participants who gave informed consent were collected between May 2002 and February 2004 by masturbation in private room and blood was collected within one week of semen collection except for a subgroup which were collected within one year. The men were asked to abstain from sexual activities for ≥ 2 days before collecting the sample. Immediately after collection, semen samples were kept close to the body to maintain a 37°C temperature when transported to the laboratory. Two cryotubes with 0.2 mL aliquots of undiluted raw semen collected 30 min after liquefaction, were prepared from each semen sample, and long-term storage was at −80°C freezer. The blood samples were centrifuged immediately after collection and sera were stored in a −80°C freezer for later analysis. Samples were analyzed at the department of Occupational and Environmental Medicine in Lund, Sweden as previously described. Briefly, the p,p'-DDE was extracted by solid phase extraction using on-column degradation of the lipids and analysis by gas chromatography mass spectrometry. The relative standard deviations, calculated from samples analyzed in duplicate at different days, were 1% at 1 ng/mL (n = 1,058), 8% at 3 ng/mL (n = 1,058) and 7% at 8 ng/mL (n = 1,058) and the detection limit was 0.1 ng/mL for p,p'-DDE.

Table S2: Aggregated population data of South African men from this study (VhaVenda cohort; m = 51; Figure S1B; mean +/- SEM or %).

	mean (± SEM) or %
Indoor residual spraying (%)	66%
Age (years)	25.3 ± 0.521
Body mass index (BMI)	20.742 ± 0.373
Pesticide self application (%)	50%
Smoking (%)	48%
Drinking (%)	60%
Sperm count (million / mL)	57.421 ± 5.366
Progressive motility (a + b)	53.449 ± 2.181
DNA fragmentation index (DFI; %)	7.703 ± 1.0890
***p,p'*-DDE (ng / mL)**	10462.228 ± 1792.298

Table S2 Notes: The Vhembe district is a malaria endemic area where housing includes mud, or brick or cement dwellings that are sprayed with DDT or not, to control for malaria. Participants volunteered from 12 villages from the Vhembe district of the Limpopo province of South Africa that either sprayed via indoor residual spraying (n = 33 participants) or not non-sprayed (n = 17 participants), This prospective study was conducted in the same manner as our prior studies and full details on recruitment methods and the questionnaire have been previously described. Semen and blood were collected on the same day in either October 2016, February 2017, or November 2017. Men were excluded from the study if they were less than 18 or more than 40 years of age, appeared intoxicated, reported drug use, or had a neuropsychiatric illness. Physical measurements included height and weight. All participants provided informed consent and were interviewed using a yes or no format questionnaire on their use of insecticides, smoking, and drinking. Fertility status was not queried. Of 247 men enrolled in the study, we selected 50 paired blood and semen samples, from men that ranged from 18 to 32 years of age (mean 25 years). Sample inclusion was based on normal sperm counts (> 15 million/mL), normal sperm DNA fragmentation index, and testing a range of p,p'-DDE serum levels (mean 10,462.228 ± 1,792.298 ng/mL, Figure S1). Note for adherence with General Data Protection Regulations (GDPR) individual data cannot be published. The semen was preserved in Sperm Freeze (LifeGlobal) and stored in liquid nitrogen for transport, followed by long-term storage in a −80°C freezer. Semen analysis was performed according to the World Health Organization standard55 and the DNA fragmentation index was determined as detailed in de Jager et al., 2009. Semen analyses involved a macroscopic and microscopic examination of sperm which takes into account sperm count, sperm motility, and sperm morphology.

Table S3: MethylC-Capture-sequencing (MCC-seq) read statistics of Greenlandic sperm.

MCC-seq read statistics for Greenlandic sperm

sample ID	total number of raw reads	number of aligned reads	% alignment	duplicate reads	% duplication rate	uniquely mapped reads	% useful aligned reads	on target reads	% on target rate	% lambda conversion rate	% human conversion rate	estimated average genome coverage	median CpG coverage
GL1	181680158	157259762	86.89	61639054	39.19	95620708	52.63	48009594	50.20	99.72	99.70	32.64	27
GL2	175944092	149233162	85.20	53323384	35.73	95909778	54.51	49239546	51.33	99.70	99.66	33.58	28
GL3	178649926	154756672	87.00	70970714	45.85	83785958	46.89	39344583	46.95	99.69	99.67	26.46	22
GL4	167291332	142944290	85.88	52649422	36.83	90294868	53.97	45432346	50.31	99.70	99.68	30.97	26
GL5	164570866	143102564	87.25	59129190	41.31	83973374	51.02	40761107	48.54	99.71	99.67	27.50	23
GL6	188295242	162507622	86.81	61489688	37.83	101017934	53.64	51305178	50.78	99.72	99.69	34.89	29
GL7	195501454	169837798	87.22	73710382	43.40	96127416	49.16	46545955	48.42	99.72	99.71	31.46	26
GL8	180807174	157178592	87.17	62573770	39.81	94604822	52.32	46143380	48.77	99.69	99.69	31.01	27
GL9	146699408	126421552	86.72	47091952	37.24	79329600	54.07	40855975	51.50	99.67	99.66	28.23	22
GL10	159046420	134727866	85.50	50362998	37.38	84364868	53.04	43882959	52.01	99.66	99.67	30.36	24
GL11	159090000	137569006	86.98	52970652	38.50	84598354	53.17	43375683	51.27	99.70	99.68	30.00	23
GL12	159149022	135496636	86.21	45627278	33.67	89869358	56.46	47603802	52.97	99.64	99.64	33.46	25
GL13	169178528	144755760	86.70	52799972	36.47	91955788	54.35	51422765	55.92	99.69	99.67	36.09	27
GL14	131866644	113128296	86.21	42287564	37.38	70840732	53.71	38455805	54.28	99.67	99.64	26.56	20
GL15	150366520	126857976	85.15	47485342	37.43	79372634	52.78	44505449	56.07	99.64	99.65	31.20	23
GL16	166329096	144298592	87.16	53653578	37.18	90645014	54.49	50237367	55.42	99.70	99.69	34.79	27
GL17	172878272	148357710	86.63	64936142	43.76	83421568	48.25	45718703	54.80	99.70	99.68	32.08	23
GL18	157290790	128790732	82.73	53282686	41.37	75508046	48.00	41622162	55.12	99.63	99.64	29.23	22
GL19	171428294	144195302	85.58	60397550	41.88	83797752	48.88	46751222	55.79	99.64	99.63	33.28	23
GL20	153069742	128120694	84.40	53548250	41.79	74572444	48.71	40756428	54.65	99.66	99.66	28.53	21
GL21	196945488	170515510	87.07	71269082	41.79	99246428	50.39	55661307	56.08	99.65	99.67	38.60	30
GL22	159371566	135921524	85.79	53715436	39.51	82206088	51.58	45573315	55.43	99.65	99.64	31.81	24
GL23	181944930	156192400	86.41	63479972	40.64	92712428	50.95	52183806	56.28	99.64	99.65	36.56	27
GL24	184369896	159257838	87.12	62834336	39.45	96423502	52.29	54064065	56.06	99.68	99.69	37.81	28
GL25	186263000	160372502	86.89	62397138	38.90	97975364	52.60	52832266	53.92	99.68	99.67	36.87	28
GL26	174800592	145551408	84.00	61280446	42.10	84270962	48.20	43760421	51.92	99.65	99.66	30.50	23
GL27	186624372	160301876	86.43	67412512	42.05	92889364	49.77	48367496	52.07	99.64	99.63	33.69	25
GL28	169193044	142809150	85.30	58389514	40.88	84419636	49.89	43833312	51.92	99.63	99.63	30.69	22
GL29	181375148	156410562	86.76	61501150	39.32	94909412	52.32	50018462	52.70	99.67	99.66	34.72	26
GL30	194207136	168310590	87.06	69858428	41.50	98452162	50.69	51014104	51.81	99.65	99.63	34.79	29
GL31	216131432	182932592	85.10	71383468	39.02	111549124	51.61	60391711	54.13	99.70	99.68	42.24	32
GL32	156053220	132652232	85.67	50414002	38.00	82238230	52.69	43621603	53.04	99.67	99.67	30.36	22
GL33	177699876	150931090	85.87	56074238	37.15	94856852	53.38	50523931	53.26	99.69	99.69	35.60	26
GL34	166612028	143178062	86.92	52546224	36.69	90631838	54.39	48415104	53.41	99.63	99.64	34.10	24
GL35	142906640	120993672	85.55	47617302	39.35	73376370	51.34	38369037	52.29	99.60	99.62	26.90	18
GL36	165836442	140657198	85.61	52568120	37.37	88089078	53.11	47010861	53.36	99.66	99.67	32.86	24
GL37	151931590	128414660	85.69	44940876	34.99	83473784	54.94	46025967	55.13	99.63	99.64	32.60	23
GL38	153845984	130687696	85.41	47437890	36.29	83249806	54.11	45568937	54.73	99.61	99.64	31.71	24
GL39	161611580	137961590	86.07	51038314	36.99	86923276	53.78	47123287	54.21	99.67	99.66	33.09	24
GL40	177508108	152846452	86.98	61984024	40.55	90862428	51.18	49106181	54.04	99.69	99.69	34.60	24
GL41	164756238	142168394	86.88	53525514	37.64	88642880	53.80	48074460	54.23	99.65	99.65	33.62	25
GL42	185604606	158247888	86.68	64042076	40.46	94205812	50.75	51912050	55.10	99.63	99.64	36.97	26
GL43	158172182	130920318	83.81	51327554	39.20	79592764	50.32	43356086	54.47	99.63	99.64	30.43	22
GL44	170621636	143715284	85.47	63010998	43.84	80704286	47.30	43252073	53.59	99.64	99.65	30.64	22
GL45	148486600	129933818	88.02	47777818	36.77	82156000	55.32	43810100	53.32	99.65	99.62	30.59	21
GL46	175698374	150598602	86.52	61098752	40.57	89499850	50.93	48165503	53.81	99.64	99.66	33.67	24
GL47	179025470	154206614	86.78	61801232	40.07	92405382	51.61	50202698	54.32	99.64	99.65	35.22	25

Table S3 Notes: Generalized linear regression models (GLMs) were built using the methylation proportion inferred from the combination of methylated reads and unmethylated reads as a binomially distributed response variable to look for associations between DNAme in sperm and *p.p'*-DDE serum levels. Continuous and binarized *p.p'*-DDE effects were both explored and models were adjusted for BMI, smoking status and age. For the downstream analyses, we opted for the continuous regression *p,p'*-DDE model but for visualization purposes, samples from each cohort were separated into *p,p'*-DDE exposure tertiles in Figure 1D. For some CpGs, the number of individuals with sufficient sequencing coverage (>= 20X) was low (e.g. < 30 samples); these CpGs were removed from our analyses, to minimize the impact of low measurement accuracy. Non-variable CpGs (standard deviation = 0) were also removed to reduce the multiple testing burden. For the South African cohort, from a total number of 2,354,599 CpGs with 20X coverage and covered by at least one sample, we obtained 1,573,641 CpGs with 20X coverage and these CpGs were covered by more than 30 samples (66.8% of total CpGs were retained after removing low coverage CpGs). Furthermore, 3,327 CpGs showed non-variable methylation (0.14% of total CpGs, or 0.2% after removing low coverage CpGs). For the Greenlandic cohort, from a total number of 2,458,895 CpGs with 20X coverage and covered by at least one sample, we obtained 1,728,019 CpGs with 20X coverage and these CpGs were covered by more than 30 samples (70.3% of total CpGs were retained after removing low coverage CpGs). Furthermore, 4,034 CpGs showed non-variable methylation (0.16% of total CpGs, or 0.23% after removing low coverage CpGs). R function glm() and the binomial family were used to fit each model, and p-values for variables of interest were obtained accordingly. The obtained p-values were then corrected by estimating the false discovery rate q-values using the Bioconductor/R package qvalue (version 2.16). We defined significant associated DMCs when q-values were less than 0.01.

Table S4: MethylC-Capture-sequencing (MCC-seq) read statistics of South African sperm.

MCC-seq read statistics for South African sperm

sample ID	total number of raw reads	number of aligned reads	% alignment	duplicate reads	% duplication rate	uniquely mapped reads	% useful aligned rate	on target reads	% on target rate	% lambda conversion rate	% human conversion rate	estimated average genome coverage	median CpG coverage
SA50	365352836	305623810	84.91	156844488	51.31	148779322	40.72	55955008	37.60	NA	99.57	37.05	33
SA14	193077232	152762212	79.99	89836498	58.80	62925714	32.59	38422160	61.05	99.58	99.51	26.40	21
SA42	196736572	163908490	83.66	86701034	52.89	77207456	39.24	28203697	36.52	99.59	99.53	18.97	16
SA22	261495372	220665024	84.68	151561826	68.68	69103198	26.42	44625782	64.57	99.76	99.72	30.43	26
SA38	231930594	193585834	83.67	107399848	55.47	86185986	37.16	31254251	36.26	99.65	99.56	20.92	18
SA33	209431290	177339526	84.89	118106414	66.59	59233112	28.28	38172616	64.44	99.59	99.54	26.05	22
SA28	242483090	202382056	83.67	111387542	55.03	90994514	37.52	33736511	37.07	99.59	99.53	22.57	19
SA5	221361692	188578928	85.45	125856156	66.73	62722772	28.33	41027214	65.41	99.56	99.54	28.06	23
SA24	269311886	226083342	84.15	154133230	68.17	71950112	26.71	45617534	63.40	99.76	99.72	30.99	26
SA48	273384618	227351438	83.36	127116978	55.91	100234460	36.66	36747457	36.66	99.58	99.53	24.49	21
SA49	169976438	135938408	80.37	76946170	56.60	58992238	34.70	38288491	64.90	99.58	99.53	26.28	21
SA29	253034146	215003006	85.11	144695914	67.29	70307092	27.78	45343066	64.49	99.60	99.54	30.83	26
SA30	216384578	171090080	79.16	96013044	56.11	75077036	34.69	42771502	56.97	99.76	99.71	28.76	24
SA25	165695404	132944284	80.40	74015392	55.67	58928892	35.56	35780426	60.71	99.54	99.54	24.37	20
SA6	230340258	191674704	83.44	101830734	53.12	89843970	39.00	32926012	36.64	99.74	99.70	22.05	19
SA13	179253634	143843194	80.73	83749512	58.18	60193682	33.58	36464543	60.57	99.53	99.50	24.95	20
SA39	198667440	168096468	84.86	112433928	66.88	55662540	28.01	36474178	65.52	99.65	99.53	24.94	21
SA36	233701296	195458280	83.84	134482190	68.80	60976090	26.09	39135774	64.18	99.58	99.53	26.69	22
SA11	231160808	183940214	79.79	105285266	57.23	78654948	34.02	47779566	60.74	99.72	99.69	32.55	27
SA15	198647642	156211652	78.82	87842690	56.23	68368962	34.41	43833507	64.11	99.66	99.61	29.93	25
SA16	146615384	115661404	79.09	61731856	53.37	53929548	36.78	34402953	63.79	99.52	99.48	23.48	19
SA40	178423068	141700966	79.67	79448418	56.06	62252548	34.89	40554691	65.14	99.74	99.71	27.76	23
SA23	184485462	147726458	80.56	84533366	57.22	63193092	34.25	41147062	65.11	99.65	99.61	28.36	22
SA17	219816784	185262826	84.43	123760534	66.80	61502292	27.97	39622689	64.42	99.61	99.52	26.94	23
SA7	246265388	208350866	84.77	137881924	66.17	70468942	28.61	45793063	64.98	99.63	99.55	31.23	26
SA46	225334670	180056706	80.16	101862214	56.57	78194492	34.70	47387038	60.60	99.64	99.54	32.36	26
SA35	246468720	204694458	83.50	111117462	54.28	93576996	37.96	34292583	36.64	99.71	99.54	23.05	19
SA41	168843170	135417608	80.49	78436486	57.92	56981122	33.74	37306969	65.47	99.54	99.52	25.58	21
SA45	162607236	130003292	80.18	76113600	58.54	53889692	33.14	34420555	63.87	99.59	99.54	23.52	19
SA9	212933480	170849634	80.37	96109530	56.25	74740104	35.10	43490992	58.18	99.52	99.53	29.30	25
SA37	163893514	131582718	80.87	71569072	54.39	60013646	36.61	37466344	62.42	99.50	99.51	25.71	21
SA43	176638244	140560244	80.12	79650186	56.66	60910058	34.48	40191265	65.98	99.59	99.54	27.69	23
SA44	170398356	136934006	80.60	74394394	54.32	62539612	36.70	40024617	63.99	99.62	99.54	27.32	23
SA20	233642672	186793146	80.19	107285752	57.43	79507394	34.02	47939152	60.29	99.54	99.50	32.65	26
SA2	208555364	166954384	80.32	95343882	57.10	71610502	34.33	43274475	60.43	99.75	99.72	29.53	24
SA10	202005416	167943162	83.35	90190310	53.70	77752852	38.49	28802710	37.04	99.55	99.52	19.31	16
SA34	275301646	232266602	84.62	154792160	66.64	77474442	28.14	50835486	65.61	99.54	99.54	34.78	29
SA31	174732480	140234332	80.90	81029606	57.78	59204726	33.88	38545979	65.10	99.57	99.53	26.52	21
SA47	251465784	212787118	84.81	143893752	67.62	68893366	27.39	44696184	64.87	99.56	99.54	30.50	25
SA1	237533930	197494806	83.28	105185380	53.25	92309426	38.86	33383581	36.16	99.59	99.53	22.28	19
SA18	218788138	185029588	84.77	121931690	65.89	63097898	28.83	41132555	65.18	99.77	99.72	28.12	23
SA8	243661626	201542780	83.32	103394460	51.30	98148320	40.28	36281864	36.96	99.77	99.69	24.43	21
SA3	246353842	205456180	83.73	111071952	54.06	94384228	38.31	34219864	36.25	99.69	99.61	23.03	19
SA32	164420988	129407060	78.98	72889208	56.32	56517852	34.37	36189198	64.03	99.54	99.52	24.73	20
SA12	180102686	143263464	79.91	78548496	54.82	64714968	35.93	42109868	65.06	99.55	99.53	28.86	24
SA19	221527108	175683756	79.50	99754044	56.78	75929712	34.27	44613533	58.75	99.58	99.54	30.19	25
SA26	223486058	186246558	83.64	97751586	52.48	88494972	39.59	32004116	36.16	99.63	99.55	21.52	18
SA4	204007332	163536638	80.26	95761028	58.55	67775610	33.22	40099076	59.16	99.58	99.52	27.07	23
SA27	245357324	201557508	82.68	108761644	53.96	92795864	37.82	33999008	36.63	99.72	99.70	22.97	19

Table S4 Notes: Generalized linear regression models (GLMs) were built using the methylation proportion inferred from the combination of methylated reads and unmethylated reads as a binomially distributed response variable to look for associations between DNAme in sperm and *p.p'*-DDE serum levels. Continuous and binarized *p.p'*-DDE effects were both explored and models were adjusted for BMI, smoking status and age. For the downstream analyses, we opted for the continuous regression *p,p'*-DDE model but for visualization purposes, samples from each cohort were separated into *p,p'*-DDE exposure tertiles in Figure 1D. For some CpGs, the number of individuals with sufficient sequencing coverage (>= 20X) was low (e.g. < 30 samples); these CpGs were removed from our analyses, to minimize the impact of low measurement accuracy. Non-variable CpGs (standard deviation = 0) were also removed to reduce the multiple testing burden. For the South African cohort, from a total number of 2,354,599 CpGs with 20X coverage and covered by at least one sample, we obtained 1,573,641 CpGs with 20X coverage and these CpGs were covered by more than 30 samples (66.8% of total CpGs were retained after removing low coverage CpGs). Furthermore, 3,327 CpGs showed non-variable methylation (0.14% of total CpGs, or 0.2% after removing low coverage CpGs). For the Greenlandic cohort, from a total number of 2,458,895 CpGs with 20X coverage and covered by at least one sample, we obtained 1,728,019 CpGs with 20X coverage and these CpGs were covered by more than 30 samples (70.3% of total CpGs were retained after removing low coverage CpGs). Furthermore, 4,034 CpGs showed non-variable methylation (0.16% of total CpGs, or 0.23% after removing low coverage CpGs). R function glm() and the binomial family were used to fit each model, and p-values for variables of interest were obtained accordingly. The obtained p-values were then corrected by estimating the false discovery rate q-values using the Bioconductor/R package qvalue (version 2.16). We defined significant associated DMCs when q-values were less than 0.01.

Table S5: DNA methylation (DNAme) “hotspots” in Greenlandic sperm (related to Figure 1E)

DNAme "hotspots" in Greenlandic sperm (Fig. 1E)

hypermethylated hotspots

hotspot information								DMC annotations within a hotspot			
chr	start	end	total # of CpGs	total # of DMCs	DMC ratio (%)	DMCs with DNAme Loss	DMCs with DNAme Gain	# of DMCs per closest gene	closest gene name	distance to the closest gene (in bp)	cis-methylation QTL CpGs
1	121000000	122000000	112	34	30.36	0	34	34	SRGAP2-AS1	1009	0
2	90000000	91000000	92	21	22.83	2	19	21	MIR4436A	996407	1
2	147000000	148000000	146	23	15.75	0	23	23	PABPC1P2	182793	14
								21	LINC00692	224625	0
3	26000000	27000000	270	73	27.04	0	73	52	LRRC3B	26619	1
								98	WDR1	13627	27
4	10000000	11000000	1111	139	12.51	1	138	40	ZNF518B	974	7
								1	CLNK	120618	0
4	14000000	15000000	247	27	10.93	0	27	2	LINC01085	8883	0
								25	LINC00504	126200	0
4	49000000	50000000	18	4	22.22	0	4	4	CWH43	45991	0
								18	LINC02562	1531	0
4	76000000	77000000	546	64	11.72	0	64	42	LINC02483	47	0
								1	RCHY1	75415	0
								3	NAAA	15534	0
4	78000000	79000000	537	108	20.11	0	108	43	CCNG2	36378	17
								65	CXCL13	1250	3
5	144000000	145000000	129	26	20.16	0	26	11	KCTD16	657680	0
								15	PRELID2	431068	0
6	87000000	88000000	373	38	10.19	0	38	36	HTR1E	78173	0
								2	SMIM8	34164	0
								1	TAC1	41326	0
7	97000000	98000000	907	91	10.03	0	91	4	MIR5692A1	8402	0
								86	BHLHA15	125	31
8	132000000	133000000	492	100	20.33	0	100	23	ADCY8	39941	0
								77	EFR3A	51369	2
								3	LOC101927822	135849	0
8	135000000	136000000	446	109	24.44	0	109	102	ZFAT-AS1	196528	2
								4	LOC101927845	109021	0
								5	LOC101927845	156695	0
8	136000000	137000000	621	226	36.39	0	226	143	LINC01591	1147	1
								78	KHDRBS3	71117	27
								45	COL22A1	182291	0
8	140000000	141000000	955	110	11.52	0	110	64	KCNK9	33258	10
								1	PEG13	117196	0
11	42000000	43000000	198	20	10.10	0	20	4	LINC02740	75723	0
								16	HNRNPKP3	293428	0
11	50000000	51000000	175	23	13.14	0	23	23	LOC441601	19252	1
11	87000000	88000000	179	31	17.32	0	31	30	LINC02711	83707	1
								1	CTSC	74012	0
								1	LINC02297	12227	0
								7	POTEG	539	1
14	19000000	20000000	118	13	11.02	1	12	3	DUXAP10	35783	0
								1	LNCRNA-ATB	68065	1
								1	BMS1P22	1216	0

hypomethylated hotspots

hotspot information								DMC annotations within a hotspot			
chr	start	end	total # of CpGs	total # of DMCs	DMC ratio (%)	DMCs with DNAme Loss	DMCs with DNAme Gain	# of DMCs per closest gene	closest gene name	distance to the closest gene (in bp)	cis-methylation QTL CpGs
8	58000000	59000000	618	137	22.17	122	15	123	LINC01606	64	120
								14	LOC286177	2632	11

common hotspot DMCs in Greenland and South African cohorts

Table S5 Notes: “Hotspot” or cluster analysis was performed by calculating the ratio of DMCs with DNAme gain or loss over the total number of CpGs found within 1 Mb non-overlapping bins over the genome; densities >10% (termed clusters) were extracted for further analysis. To investigate genetic effects on DNAme of CpGs within the clusters, methylation quantitative trait locus (meQTLs) analyses were performed. The genotype profiles of SNPs within all the candidate clusters as well as all the DMCs were extracted. By considering possible SNP cis-effects within 250 kb of a CpG (i.e., a 500-kb window), meQTLs were calculated using Bioconductor/R package MatrixEQTL (version 2.3) with default parameters. The reported p-values were corrected using Benjamini-Hochberg false discovery rate (FDR).

Table S6: DNA methylation (DNAme) "hotspots" in South African sperm (related to Figure 1E)

DNAme "hotspots" in South African sperm (Fig. 1E)

hypermethylated hotspots

hotspot information								DMC annotations within a hotspot			
chr	start	end	total # of CpGs	total # of DMCs	DMC ratio (%)	DMCs with DNAme Loss	DMCs with DNAme Gain	# of DMCs per closest gene	closest gene name	distance to the closest gene (in bp)	cis-methylation QTL CpGs
2	56000000	57000000	178	21	11.80	0	21	21	LOC100129434	6971	0
								147	WDR1	1658	9
4	10000000	11000000	1043	201	19.27	7	194	46	ZNF518B	2452	2
								8	CLNK	64525	4
								26	LOC100505912	28253	0
4	22000000	23000000	339	36	10.62	0	36	6	MIR12115	38971	0
								4	GBA3	53717	0
								94	CTD-2194D22.4	118473	2
								118	LOC100506858	9711	19
5	2000000	3000000	2004	414	20.66	22	392	92	LSINCT5	8762	10
								15	C5orf38	36025	7
								95	LOC105374620	2596	6
5	129000000	130000000	86	12	13.95	3	9	12	CHSY3	373188	0
6	64000000	65000000	259	33	12.74	1	32	23	LGSN	120517	23
								10	SCAT8	216975	1
								4	IMPG1	235744	0
6	77000000	78000000	132	16	12.12	0	16	7	LINC02540	116767	0
								5	HTR1B	399843	0
6	81000000	82000000	153	22	14.38	1	21	22	BCKDHB	312079	1
								102	SMOC2	158876	5
								144	LOC105378146	20	54
								4	LOC101929460	13266	0
6	169000000	170000000	1911	285	14.91	21	264	1	LOC102724357	25746	0
								5	LINC01615	6395	0
								21	LINC02544	3465	10
								1	LOC101929523	10322	0
								7	THBS2	35742	5
								2	TAC1	41326	1
								2	OCM2	44499	0
7	97000000	98000000	807	88	10.90	2	86	1	LMTK2	19564	0
								50	BHLHA15	52	3
								32	BRI3	1544	0
								1	BAIAP2L1	51766	1
								1	SNRPN	21357	1
								1	SNORD116-3	113	1
								1	SNORD116-5	276	1
								1	SNORD116-21	131	1
								2	SNORD115-1	1088	0
								5	SNORD115-12	106	0
								13	SNORD115-6	28	0
								13	SNORD115-7	22	1
								10	SNORD115-8	44	1
								4	SNORD115-12	66	0
								6	SNORD115-5	2	1
								12	SNORD115-10	5	1
								12	SNORD115-36	12	0
								11	SNORD115-29	72	0
								8	SNORD115-9	66	1
								6	SNORD115-5	2	0
								12	SNORD115-13	5	1
								3	SNORD115-14	5	0
								11	SNORD115-15	5	1
								22	SNORD115-16	72	1
								9	SNORD115-19	22	2
								5	SNORD115-18	5	4
								3	SNORD115-17	344	0
								6	SNORD115-19	22	0
								17	SNORD115-18	5	0
								6	SNORD115-20	82	0
								11	SNORD115-21	7	0
								22	SNORD115-22	44	0
								3	PWAR4	197	0
15	25000000	26000000	1373	443	32.27	9	434	7	SNORD115-23	38	0
								10	SNORD115-24	62	0
								10	SNORD115-25	104	0
								13	SNORD115-26	22	0
								16	SNORD115-27	16	0
								1	SNORD115-28	74	0
								3	SNORD115-11	21	0
								13	SNORD115-36	12	0
								3	SNORD115-30	163	0
								16	SNORD115-31	18	0
								2	SNORD115-32	142	0
								8	SNORD115-33	28	0
								7	SNORD115-34	23	0
								2	SNORD115-35	134	0
								4	SNORD115-36	12	0
								1	SNORD115-29	72	0
								11	SNORD115-37	151	0
								3	SNORD115-38	44	0
								12	SNORD115-39	3	0
								12	SNORD115-21	7	0
								5	SNORD115-40	72	1
								6	SNORD115-41	37	1
								13	SNORD115-15	5	0
								6	SNORD115-42	5	0
								3	SNORD115-36	12	0
								1	SNORD115-44	1259	0
								1	SNORD115-45	1705	0
								1	SNORD109B	49354	0
								8	UBE3A	2270	0
								9	MIR4715	130173	0
								8	LINC01893	137079	0
								39	LOC101927989	3325	0
								11	LINC00908	1843	0
								7	LINC00683	1567	0
18	74000000	75000000	1781	192	10.78	3	189	1	LINC01879	7559	0
								1	ZNF236-DT	18267	0
								122	MSP	20068	1
								3	GALR1	4583	1
								49	LINC01029	294923	3
18	76000000	77000000	2432	325	13.36	4	321	254	SALL3	4582	35
								22	ATP9B	29849	8
								1	TAFA5	68447	0
22	49000000	50000000	1276	226	17.71	5	221	3	MIR4535	19084	1
								25	LINC01310	35344	0
								197	MIR3667	3543	9

hypomethylated hotspots

hotspot information								DMC annotations within a hotspot			
chr	start	end	total # of CpGs	total # of DMCs	DMC ratio (%)	DMCs with DNAme Loss	DMCs with DNAme Gain	# of DMCs per closest gene	closest gene name	distance to the closest gene (in bp)	cis-methylation QTL CpGs
3	82000000	83000000	207	23	11.11	22	1	23	LINC02008	511855	12
								3	LINC01324	146292	1
3	164000000	165000000	209	24	11.48	24	0	2	SI	115603	0
								18	SLITRK3	7030	0
								1	LINC01322	51327	0
3	166000000	167000000	136	15	11.03	14	1	14	LOC105374194	131915	1
								1	ZBBX	240042	
5	104000000	105000000	135	21	15.56	21	0	21	RAB9BP1	252017	0
6	95000000	96000000	120	13	10.83	13	0	11	TSG1	626135	0
								2	MANEA-DT	455616	0
								82	LINC01606	33	81
8	58000000	59000000	553	85	15.37	78	7	1	LOC286177	1141	0
								1	LINC00588	191	0
								1	LOC286178	146399	0
9	12000000	13000000	129	14	10.85	14	0	14	LURAP1L-AS1	15919	0
10	107000000	108000000	73	13	17.81	13	0	13	LINC02627	428114	11
11	21000000	22000000	58	9	15.52	9	0	7	NELL1	559766	0
								2	ANO5	697924	0
								12	LINC02737	1085129	0
11	97000000	98000000	156	18	11.54	16	2	6	CNTN5	935423	0
16	33000000	34000000	270	44	16.30	44	0	44	ENPP7P13	11	0
18	27000000	28000000	168	24	14.29	24	0	24	MIR302F	40950	11

common hotspot DMCs in Greenland and South African cohorts

Table S6 Notes: "Hotspot" or cluster analysis was performed by calculating the ratio of DMCs with DNAme gain or loss over the total number of CpGs found within 1 Mb non-overlapping bins over the genome; densities >10% (termed clusters) were extracted for further analysis. To investigate genetic effects on DNAme of CpGs within the clusters, methylation quantitative trait locus (meQTLs) analyses were performed. The genotype profiles of SNPs within all the candidate clusters as well as all the DMCs were extracted. By considering possible SNP cis-effects within 250 kb of a CpG (i.e., a 500-kb window), meQTLs were calculated using Bioconductor/R package MatrixEQTL (version 2.3) with default parameters. The reported p-values were corrected using Benjamini-Hochberg false discovery rate (FDR).

Table S7: H3K4me3 Chromatin Immunoprecipitation followed by sequencing (ChIP-seq) read statistics on South African sperm (related to Figure S4A,B).

H3K4me3 ChIP-seq read statistics on South African sperm (Fig. S4A,B)

sample ID	number of mapped reads	uniquely mapped reads	duplicate reads	% duplication rate	% GC content	average read length
SA14	120638877	86111082	34527795	28.60%	48%	100 bp
SA42	105261054	81415102	23845952	22.70%	46%	98 bp
SA22	99984668	77249990	22734678	22.70%	45%	99 bp
SA38	108189837	85009378	23180459	21.40%	45%	98 bp
SA33	107903751	79331366	28572385	26.50%	46%	97 bp
SA28	121779743	95008155	26771588	22.00%	44%	98 bp
SA5	115224225	91575455	23648770	20.50%	43%	99 bp
SA24	117080367	91048904	26031463	22.20%	44%	98 bp
SA48	161098154	120665498	40432656	25.10%	45%	98 bp
SA49	93122357	73140800	19981557	21.50%	45%	100 bp
SA29	96804073	74243932	22560141	23.30%	46%	99 bp
SA30	134819817	102086328	32733489	24.30%	45%	99 bp
SA25	136212868	102925591	33287277	24.40%	44%	100 bp
SA6	139037031	104722774	34314257	24.70%	45%	98 bp
SA13	112787606	82804291	29983315	26.60%	47%	100 bp
SA39	130645654	97610523	33035131	25.30%	44%	100 bp
SA36	120694644	89533026	31161618	25.80%	44%	99 bp
SA11	88422401	69332107	19090294	21.60%	45%	100 bp
SA15	85740633	67108764	18631869	21.70%	46%	100 bp
SA16	112665077	83213946	29451131	26.10%	46%	100 bp
SA40	79092487	62847217	16245270	20.50%	45%	100 bp
SA23	124988157	91567011	33421146	26.70%	45%	99 bp
SA17	149713684	112124585	37589099	25.10%	45%	98 bp
SA7	96063448	73354131	22709317	23.60%	45%	98 bp
SA46	125126533	95985303	29141230	23.30%	45%	100 bp
SA35	100991568	77580034	23411534	23.20%	46%	98 bp
SA41	110473196	83183574	27289622	24.70%	46%	99 bp
SA45	134170321	104654746	29515575	22.00%	44%	100 bp
SA9	118293149	90765691	27527458	23.30%	45%	100 bp
SA37	105459676	81599570	23860106	22.60%	44%	100 bp
SA43	147255139	104759839	42495300	28.90%	46%	99 bp
SA44	116909630	85015892	31893738	27.30%	46%	100 bp
SA20	107448501	84055862	23392639	21.80%	44%	100 bp
SA2	111902679	88792719	23109960	20.70%	43%	99 bp
SA10	117810459	93590345	24220114	20.60%	44%	99 bp
SA34	150763887	111002765	39761122	26.40%	46%	98 bp
SA31	165879131	126719760	39159371	23.60%	44%	99 bp
SA47	114225225	91019451	23205774	20.30%	45%	97 bp
SA1	93461800	77818156	15643644	16.70%	43%	99 bp
SA18	124201261	96453308	27747953	22.30%	44%	98 bp
SA8	114580739	82745384	31835355	27.80%	46%	98 bp
SA3	99872775	78850462	21022313	21.00%	43%	98 bp
SA32	86475220	66697476	19777744	22.90%	45%	100 bp
SA12	100221964	74809595	25412369	25.40%	46%	100 bp
SA19	94298267	72216447	22081820	23.40%	45%	100 bp
SA21	113514500	82159070	31355430	27.60%	45%	100 bp
SA26	111552751	88766297	22786454	20.40%	44%	98 bp
SA4	106638588	78089244	28549344	26.80%	45%	100 bp
SA27	110235830	86837219	23398611	21.20%	44%	98 bp

Table S7 Notes: Raw reads were trimmed with the TrimGalore wrapper script around the sequence-grooming tool cutadapt (version 0.50) with the following quality trimming and filtering parameters (--length 50 -q 5 --stringency 1 -e 0.1`)[68]. The trimmed reads were mapped onto the hg19/GRCh37 reference genome downloaded from UCSC genome browser using bowtie2 (version 2.3.5.1) as previously described [27,61]. We excluded reads that exhibited more than 3 mismatches. SAMtools (version 1.9)[69] was then used to convert SAM files and index BAM files. BigWig coverage tracks and binding heatmaps were generated from the aligned reads using deepTools2 (version 3.2.0)[70]. The coverage was calculated as the number of reads extended to 150bp fragment size per 25 bp bin and normalized using Reads Per Kilobase per Million mapped reads (RPKM) not located on the X chromosome.

Figure S1

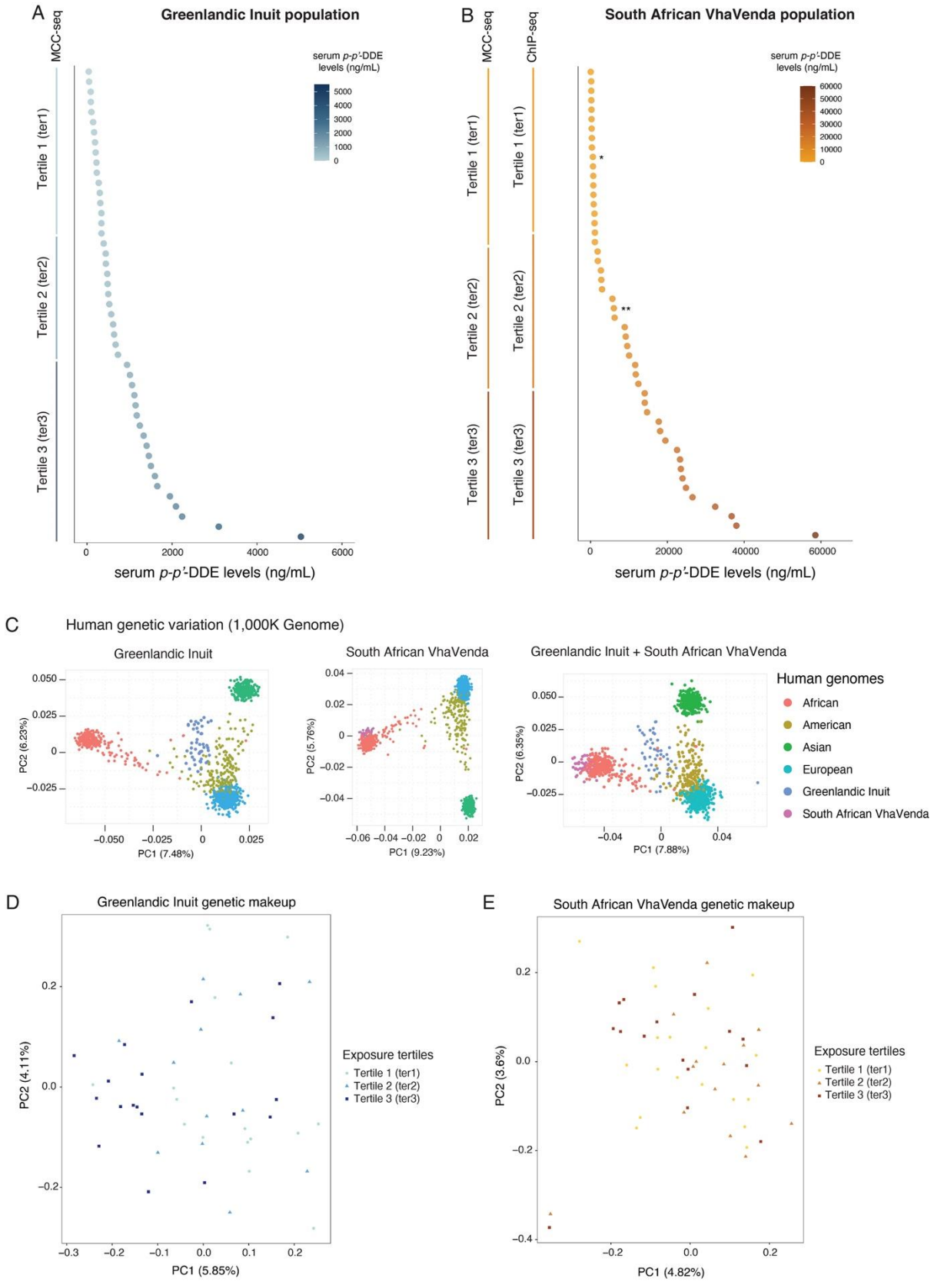

Figure S1: Characterization of Greenlandic and South African sperm relative to serum *p,p'*-DDE levels and single-nucleotide polymorphisms.

(A) Distribution of *p,p'*-DDE serum levels in Greenland men (n = 47) from this study.

(B) Distribution of *p,p'*-DDE serum levels in South African men (n = 50) from this study. *p,p'*-DDE serum level tertiles for H3K4me3 analysis are indicated on the y-axis. (*) denotes sample with only an MCC-seq dataset and (**) denotes sample with only a ChIP-seq dataset.

(C) Principal component analysis plot on genotype profiles of Greenland and South African populations from sperm MCC-seq. Chromosome 1 genotype data from 1,000K genomes (Imputed hapmap V3) were used as the reference genotype profile (the human population genetic background) to compare with Greenlandic (indigo) and South African cohorts (magenta).

(D) Principal component analysis plot on common SNPs on chromosome 1 and 1,000K genome dataset (Imputed hapmap V3) for Greenlandic population coloured by *p,p'*-DDE serum tertiles.

(E) Principal component analysis plot on common SNPs on chromosome 1 and 1,000K genome dataset (Imputed hapmap V3) for South African population coloured by *p,p'*-DDE serum tertiles.

Figure S2

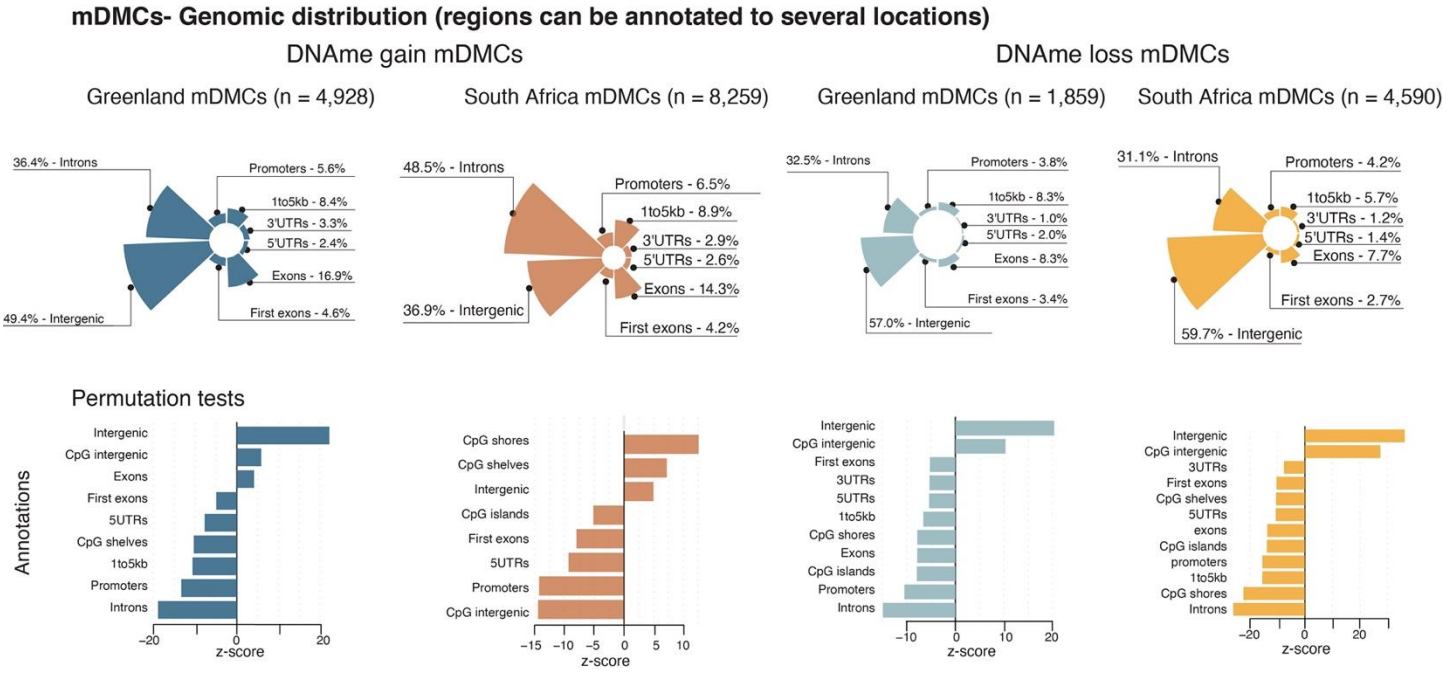

Figure S2: Genic and transposable element characterization of differentially methylated regions.

Genic distributions and genic / CpG enrichments at DNAme gain or loss mDMCs in Greenland or South Africa sperm. Positive enrichments were determined by Z scores using the Bioconductor package regioneR. For all annotations displayed, $p < 0.0001$ and n = 10,000 permutations of random regions (of the same size) resampled from the targeted MCC-seq regions.

Figure S3

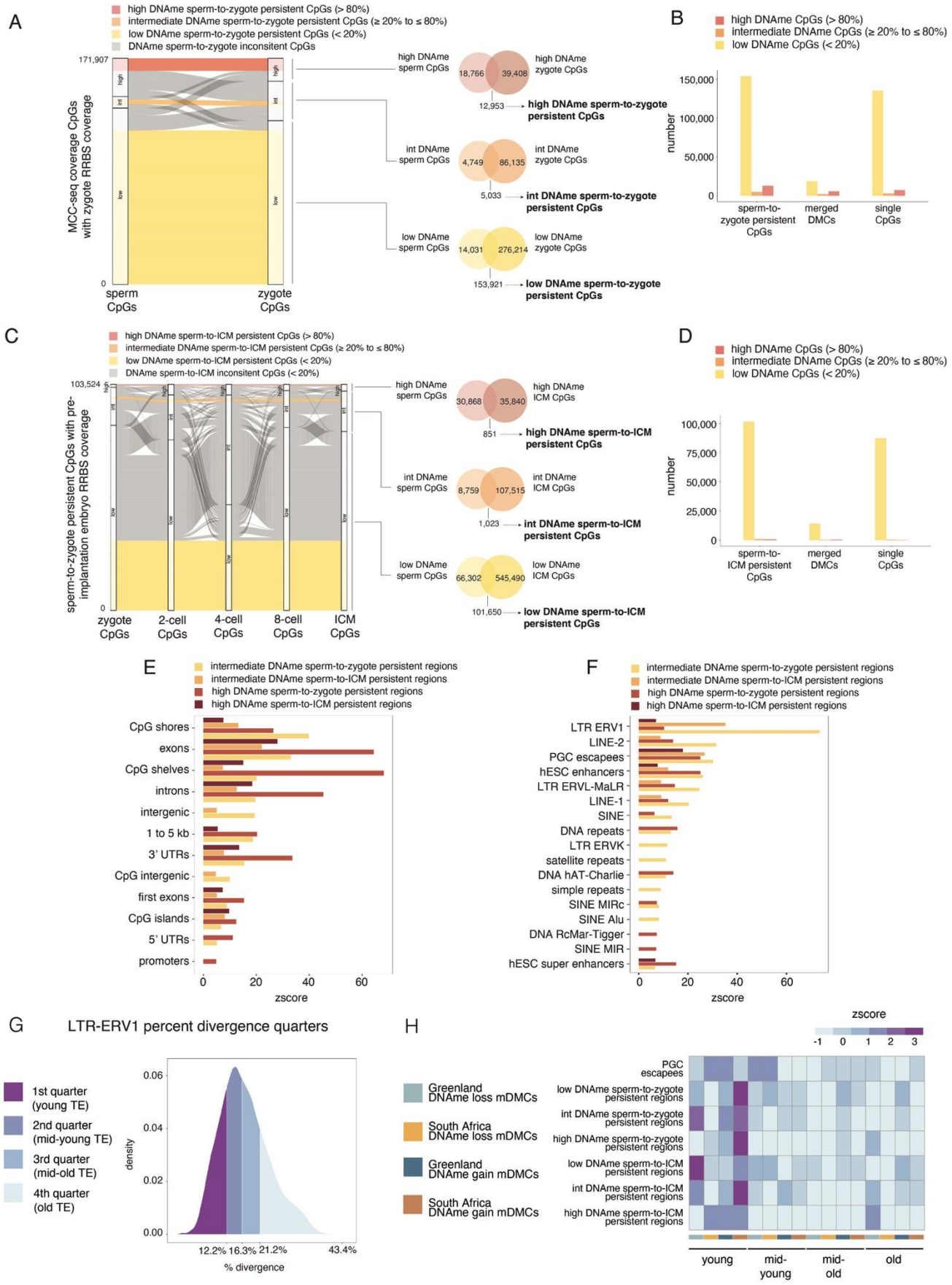

Figure S3: Identification of predicted persistent DNAme regions from sperm to the pre-implantation embryo.

(A) Alluvial plot identifying CpGs that retain the same level of DNAme from sperm to the zygote. High DNAme sperm to zygote persistent CpGs are shown by red ribbon, intermediate DNAme sperm to zygote persistent CpGs are denoted by orange ribbon, low DNAme sperm-to-zygote persistent CpGs are characterized by yellow ribbon. Grey ribbon corresponds to non-persistent DNAme CpGs. Each node indicates the DNAme level for the specific cell type. Y-axis corresponds to MCC-seq coverage CpGs that have zygote RRBS coverage. Venn diagrams show sperm-to-ICM persistent CpGs for high, intermediate, or low DNAme levels. See Excel Table S9 – S13.

(B) Number of total DNAme sperm-to-zygote persistent CpGs, DNAme sperm-to-zygote persistent regions, and single DNAme sperm-to-zygote persistent CpGs, for low DNAme (yellow), intermediate DNAme (orange), and high DNAme (red). DNAme sperm-to-zygote persistent regions were called by merging DNAme sperm-to-zygote persistent CpGs separated by a maximum distance of 500 bp.

(C) Alluvial plot identifying CpGs that retain the same level of DNAme across all stages of pre-implantation embryogenesis. High DNAme sperm to zygote persistent CpGs are shown by red ribbon, intermediate DNAme sperm to zygote persistent CpGs are denoted by orange ribbon, low DNAme sperm-to-zygote persistent CpGs are characterized by yellow ribbon. Grey ribbon corresponds to non-persistent DNAme CpGs. Each node indicates the DNAme level for the specific cell type. Y-axis corresponds to sperm-to-zygote persistent CpGs that have RRBS coverage across the studied stages of pre-implantation embryogenesis. Venn diagrams show sperm-to-ICM persistent CpGs for high, intermediate, or low DNAme levels.

(D) Number of total DNAme sperm-to-ICM persistent CpGs, DNAme sperm-to-ICM persistent regions, and single DNAme sperm-to-ICM persistent CpGs, for low DNAme (yellow), intermediate DNAme (orange), and high DNAme (red). DNAme sperm-to-ICM persistent regions were called by merging DNAme sperm-to-ICM persistent CpGs separated by a maximum distance of 500 bp.

(E) Enrichment for characterized DNAme persistent regions at genic annotations. Positive enrichments are determined by Z scores using the Bioconductor package regioneR. For all annotations displayed, $p < 0.0001$ and $n = 10,000$ permutations of random regions (of the same size) resampled from the targeted MCC-seq regions.

(F) Enrichment for characterized DNAme persistent regions at transposable elements (RepeatMasker hg19 library 20140131). Positive enrichments are determined by Z scores using the Bioconductor package regioneR. For all annotations displayed, $p < 0.0001$ and $n = 10,000$ permutations of random regions (of the same size) resampled from the targeted MCC-seq regions.

(G) Density plot of percent divergences for LTR-ERV1 transposable elements in the hg19 genome. Color shading corresponds to quantile cutoffs and associated age categories.

(H) Enrichment of young, mid-young, mid-old, and old LTR-ERV1 transposable elements that overlap a PGC escapee and/or characterized sperm-to-pre-implantation-embryo persistent region (see Figure S2) at given mDMC.

Figure S4

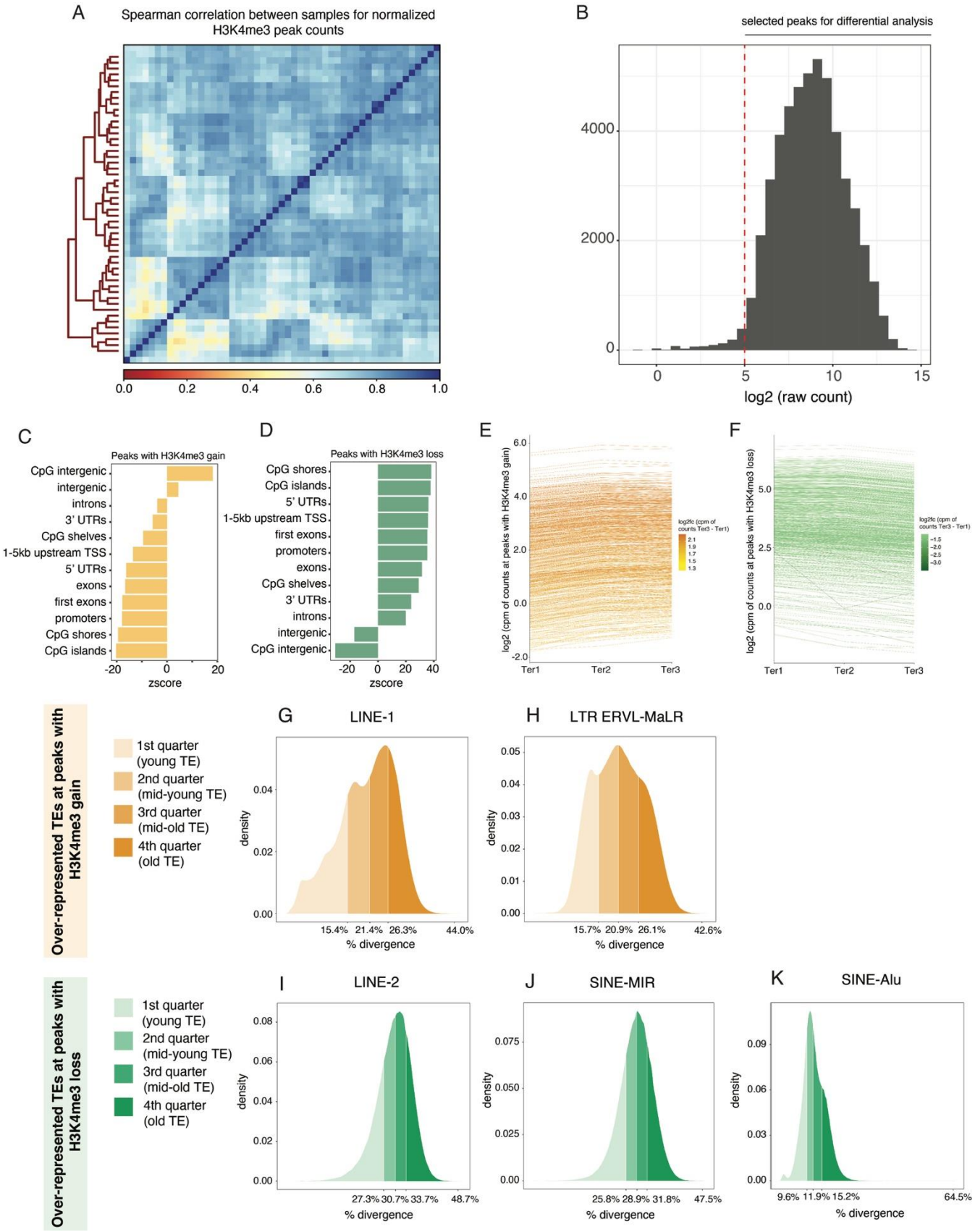

Figure S4: Genic and transposable element characterization of differentially enriched H3K4me3 peaks.

(A) Heatmap depicting spearman correlations between samples based on normalized counts of H3K4me3 peaks identified in our reference human dataset (50,117 peaks)[27]. Normalized counts are reads per kilobase per million mapped reads.
(B) Distribution of the log-transformed median raw counts of H3K4me3 reference peaks across samples in this dataset. We selected peaks with log2 median counts above 5 for downstream analyses (n = 48,499 peaks).
(C) Enrichment of peaks with H3K4me3 gain at genic annotations. Positive and negative enrichments are determined by Z scores. For all annotations displayed, $p < 0.0001$ and n = 10,000 permutations of random regions (of the same size) resampled from sperm H3K4me3 peaks.
(D) Enrichment of peaks with H3K4me3 loss at genic annotations. Positive and negative enrichments are determined by Z scores. For all annotations displayed, $p < 0.0001$ and n = 10,000 permutations of random regions (of the same size) resampled from sperm H3K4me3 peaks.
(E) Line diagram for regions with H3K4me3 gain in VhaVenda sperm where each line corresponds to the log2 cpm of H3K4me3 counts at individual regions with H3K4me3 gain in sperm across men from ter1, ter2, and ter3 *p,p'*-DDE exposure levels.
(F) Line diagram for regions with H3K4me3 loss in VhaVenda sperm where each line corresponds to the log2 cpm of H3K4me3 counts at individual regions with H3K4me3 loss in sperm across men from ter1, ter2, and ter3 *p,p'*-DDE exposure levels.
(G - H) Density plots of percent divergences for LINE-1 (E) and LTR ERVL-MaLR (F) transposable elements in the hg19 genome. Color shading corresponds to quantile cutoffs and associated age categories.
(I - K) Density plots of percent divergences for LINE-2 (G) and SINE-MIR (H), and SINE-Alu (I) transposable elements in the hg19 genome. Color shading corresponds to quantile cutoffs and associated age categories.

Figure S5

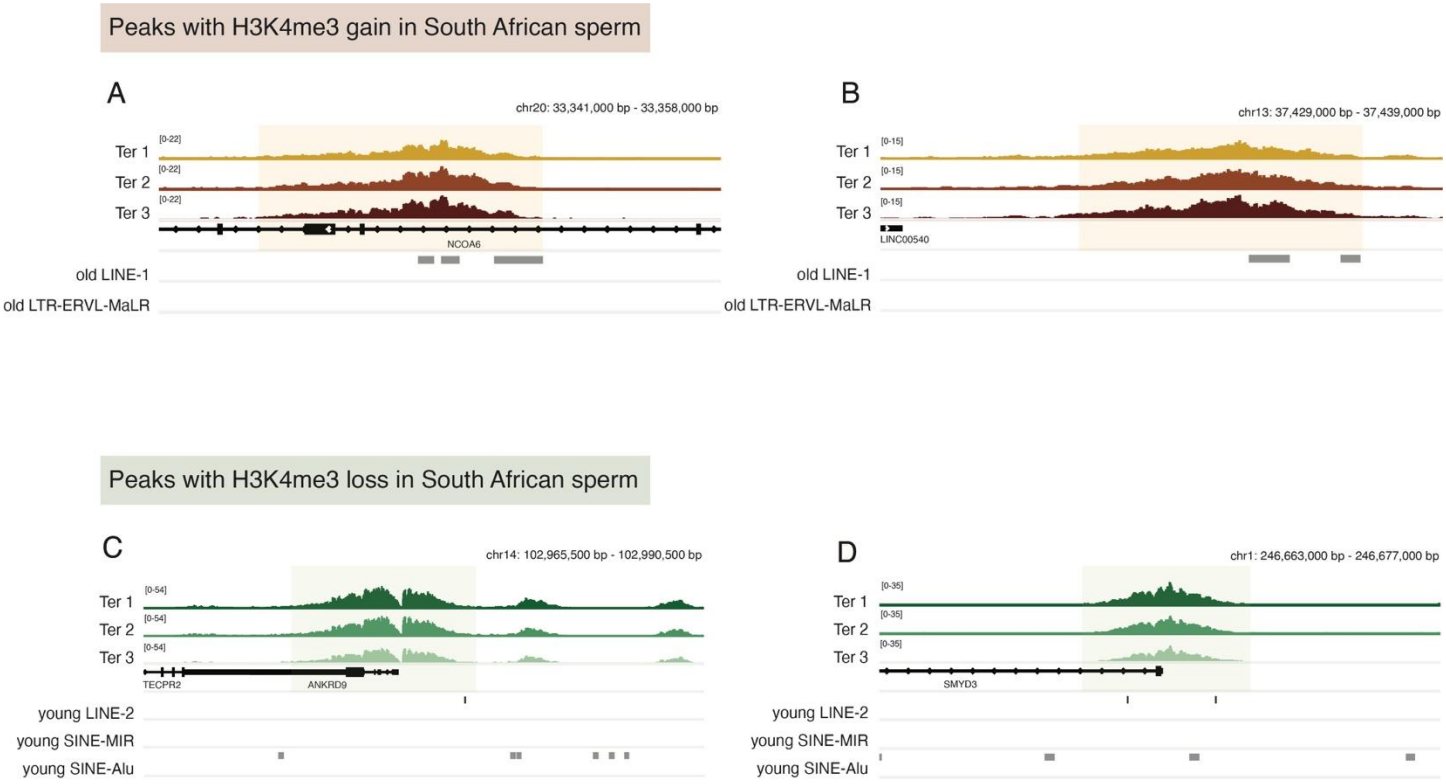

Figure S5: Examples of peaks with deH3K4m3 that overlap promoters, transposable elements and / or putative enhancers.

(A – B) Representative IGV tracks of peaks with H3K4me3 gain at (A) genic NCOA6 (important for hormone-dependent coactivation of steroid receptors, implicated in Kabuki Syndrome 1) which overlaps a putative sperm enhancer and multiple old LINE-1 elements, and (B) intergenic space in proximity to LINC00540 which overlaps a putative sperm enhancer and multiple old LINE-1 elements.

(C – D) Representative IGV tracks of peaks with H3K4me3 loss at (C) the ANKRD9 gene (involved in metabolism of proteins and Class I MHC mediated antigen processing and presentation) which overlaps a putative fetal brain enhancer and a young LINE-2 element, and (D) the SMYD3 gene (a methyltransferase implicated in hepatocellular carcinoma) which overlaps a putative fetal brain enhancer and young LINE-2 and SINE-Alu elements.

Figure S6

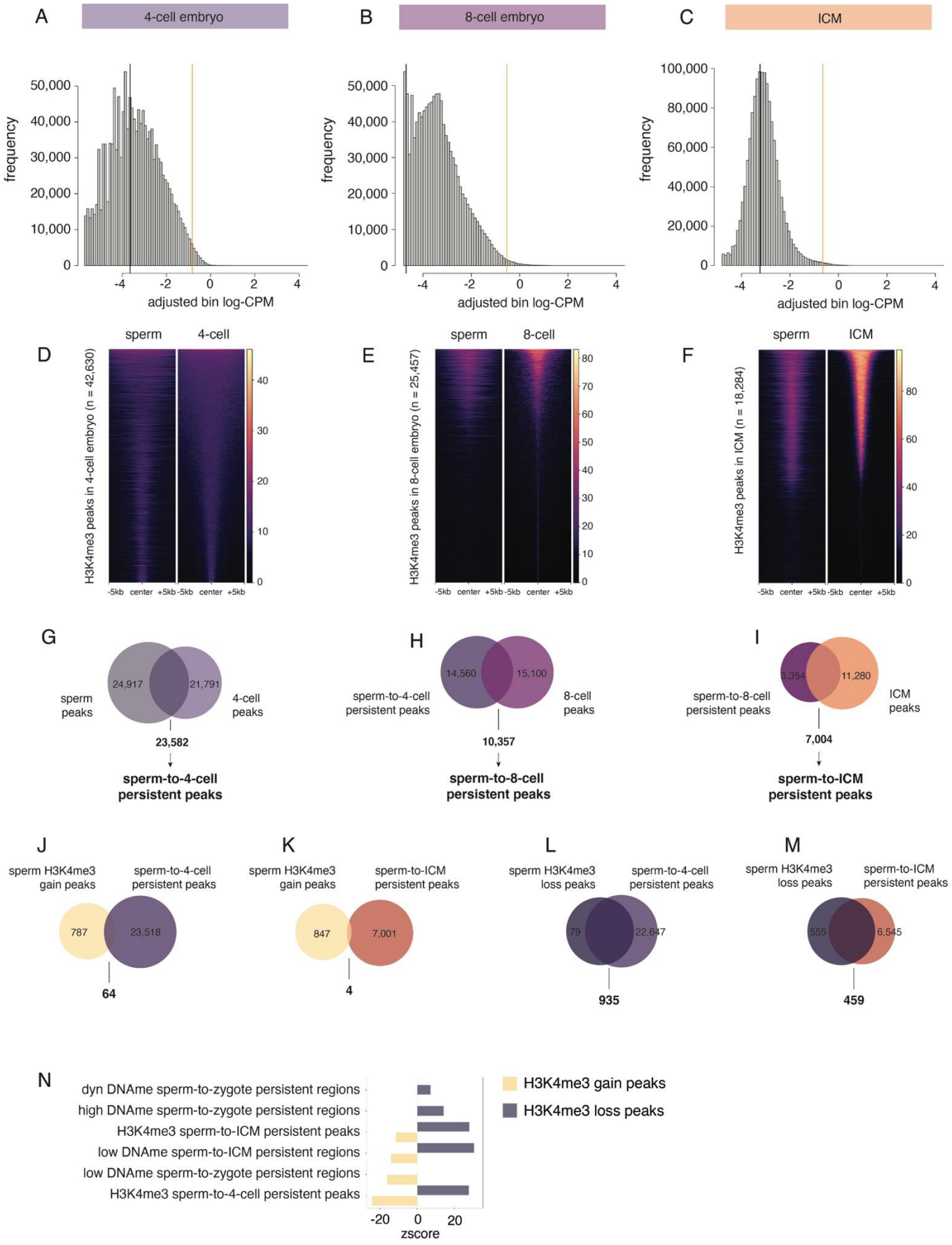
A
4-cell embryo
B
8-cell embryo
C
ICM
frequency
adjusted bin log-CPM
D
E
F
sperm
4-cell
8-cell
ICM
H3K4me3 peaks in 4-cell embryo (n = 42,630)
H3K4me3 peaks in 8-cell embryo (n = 25,457)
H3K4me3 peaks in ICM (n = 18,284)
-5kb center +5kb
G
H
I
sperm peaks
24,917
21,791
4-cell peaks
23,582
sperm-to-4-cell persistent peaks
sperm-to-8-cell persistent peaks
14,560
15,100
8-cell peaks
10,357
3,354
11,280
ICM peaks
7,004
J
K
L
M
sperm H3K4me3 gain peaks
sperm-to-4-cell persistent peaks
787
23,518
64
sperm-to-ICM persistent peaks
847
7,001
4
sperm H3K4me3 loss peaks
79
22,647
935
555
6,545
459
N
dyn DNAme sperm-to-zygote persistent regions
high DNAme sperm-to-zygote persistent regions
H3K4me3 sperm-to-ICM persistent peaks
low DNAme sperm-to-ICM persistent regions
low DNAme sperm-to-zygote persistent regions
H3K4me3 sperm-to-4-cell persistent peaks
zscore
H3K4me3 gain peaks
H3K4me3 loss peaks

Figure S6: Identification of predicted persistent H3K4me3 peaks from sperm to the pre-implantation embryo.

(A - C) Histograms of background read abundance as determined by the number of H3K4me3 ChIP-seq reads in 2000 bp windows tiled across the hg19 genome for 4-cell embryos (A), 8-cell embryos (B), and ICM (C). An abundance threshold was set at ≥ log2(7) fold over background for 4-cell embryos (A), ≥ log2(18) fold over background for 8-cell embryos (B), and ≥ log2(6) fold over background for ICM (C). Windows below this threshold were filtered out for downstream analysis. Remaining windows less than 5,000 bp (A), 6,000 bp (B), or 2,000 bp (C) apart were merged to generate H3K4me3 peaks with a maximum width of 20,000 bp. Peaks identified as enriched for H3K4me3 in the pre-implantation embryo were then inspected and confirmed in IGV.

(D - F) H3K4me3 signal intensity enrichment heatmaps at +/- 5kb center of 4-cell (D), 8-cell (E), or ICM (F) peaks relative to sperm or pre-implantation embryo signal, and sorted by pre-implantation embryo H3K4me3 signal intensity.

(G - I) Overlap between sperm H3K4me3 peaks and 4-cell H3K4me3 peaks (= 23,582; G), sperm-to-4-cell persistent H3K4me3 peaks and 8-cell H3K4me3 peaks (= 10,357; H), sperm-to-8-cell persistent H3K4me3 peaks and ICM H3K4me3 peaks (= 7,004; I). See Excel Table S21 – 23.

(J - K) Overlap between sperm H3K4me3 gain peaks and 4-cell H3K4me3 peaks (= 64; J), sperm H3K4me3 gain peaks and sperm-to-ICM persistent H3K4me3 peaks (= 4; K).

(L - M) Overlap between sperm H3K4me3 loss peaks and 4-cell H3K4me3 peaks (= 935; L), sperm H3K4me3 loss peaks and sperm-to-ICM persistent H3K4me3 peaks (= 459; M).

(N) Enrichment for peaks with H3K4me3 gain or loss at characterized DNAme / H3K4me3 persistent regions (see Figure S2 and S5). Positive enrichments are determined by Z scores using the Bioconductor package regioneR. For all annotations displayed, $p < 0.0001$ and n = 10,000 permutations of random regions (of the same size) resampled from sperm H3K4me3 peaks.

Figure S7

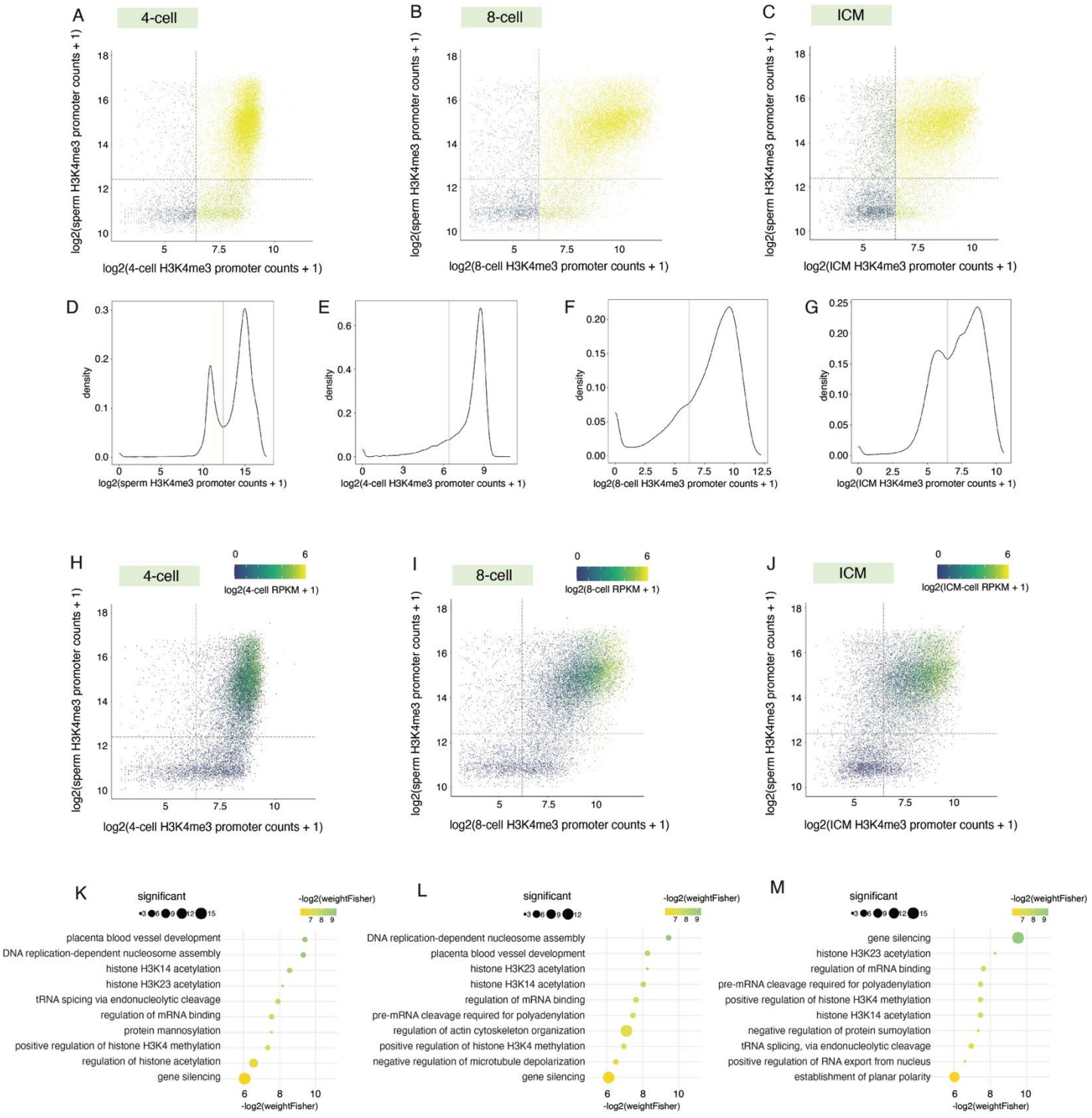

Figure S7: Promoters marked by H3K4me3 in sperm and pre-implantation embryos correspond to expressed genes in the embryo.

(A - C) Scatterplots where the x axis corresponds to the log2 (pre-implantation embryo H3K4me3 promoter counts + 1) and the y axis corresponds to the log2 (sperm H3K4me3 promoter counts + 1) at +/- 1 kb TSS of the hg19 genome. Colour of the scatter points corresponds to H3K4me3 enrichment categories determined by density cutoffs (grey dotted lines, see Figure S7D-G): yellow points = H3K4me3 enrichment in both sperm and pre-implantation embryos; light green points =

H3K4me3 enrichment in only pre-implantation embryos; dark green points = H3K4me3 in sperm; blue points = absence of H3K4me3 enrichment in sperm and pre-implantation embryos. Represented pre-implantation embryo stages are 4-cell embryos (A), 8-cell embryos (B), and ICM (C).

(D - G) Distribution of log2 (H3K4me3 counts + 1) at +/- 1 kb TSS of the hg19 genome in sperm (D), 4-cell embryos (E), 8-cell embryos (F), ICM (G). The local minimum was identified on the density plots (brown line) and used as the cutoff threshold value to identify promoters enriched for H3K4me3 in sperm and pre-implantation embryos (see Figure S7A-C).

(H - J) Scatterplots where the x axis corresponds to the log2 (pre-implantation embryo H3K4me3 promoter counts + 1) and the y axis corresponds to the log2 (sperm H3K4me3 promoter counts + 1) at +/- 1 kb TSS of the hg19 genome. Color of the scatter points correspond to the log2 pre-implantation embryo RPKM gene expression + 1. Dashed lines correspond to H3K4me3 promoter density cutoffs for pre-implantation embryo (x axis) or sperm (y axis). Represented pre-implantation embryo stages are 4-cell embryos (H), 8-cell embryos (I), and ICM (J).

(K - M) Selected significant pathways from gene ontology analysis on promoters with H3K4me3 loss in South African sperm that retain H3K4me3 in the pre-implantation embryo and are expressed at the associated pre-implantation embryo stage (RPKM > 1; weighed Fisher $p < 0.05$). Size of dots corresponds to the number of genes from a significant pathway that overlap a peak with H3K4me3 gain. Color of the dots indicates -log2(weightFisher) value of significant pathway. Represented pre-implantation embryo stages are 4-cell embryos (K), 8-cell embryos (L), and ICM (M).