Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by Aid deficiency

Christian Popp[1,*], **Wendy Dean**[1,*], **Suhua Feng**[2,*], **Shawn J. Cokus**[2], **Simon Andrews**[3], **Matteo Pellegrini**[2], **Steven E. Jacobsen**[2,4], and **Wolf Reik**[1]

[1]Laboratory of Developmental Genetics and Imprinting, The Babraham Institute, Cambridge CB22 3AT, and Centre for Trophoblast Research, University of Cambridge, Cambridge CB2 3EG, UK

R3me2s together with H2AZ, as well as loss of the active histone mark H3K9ac[9,20]. Thus an epigenetic ground state which is depleted both of key activating and repressive epigenetic marks appears to be uniquely characteristic of reprogrammed PGCs. We carried out a biological replicate of the BS-Seq experiment on E13.5 PGCs isolated by a different method (see Supplementary Methods) which indeed replicated qualitatively all conclusions with the *Oct4-Gfp* FACS sorted cells, while the base line was shifted to higher levels of methylation, reflecting either a greater contamination with somatic gonadal cells or epigenetic heterogeneity within germ cells that are not *Oct4-Gfp* positive (Supplementary Fig. 2).

We next examined the effects of Aid deficiency on erasure of methylation in PGCs. We

(based on methylation) across generations, which if it occurred would be of potential evolutionary significance as well as affecting disease risk in humans[22]. Indeed in the mouse the only two well docum

Methods Summary

Mouse tissues, including male and female PGCs at E13.5, were isolated from C57BL/6J, C57BL/6J mice transgenic for *Oct4-Gfp*, or *Aid*-/- knockout mice bred into a C57BL/6J background for 7 generations prior to this study. The *Oct4-Gfp* transgene was subsequently bred into the *Aid*-/- knockout. PGCs were isolated on a FACS-Aria cell sorter by sorting for green fluorescence; the isolated cell populations were > 98% pure. DNA was isolated, bisulfite converted, and prepared for Illumina Solexa libraries. Each Illumina Solexa library was sequenced in a single end read, except for *Oct4-Gfp* isolated PGC libraries which were sequenced in two single end reads each; subsequently, a published highly customized software package was used to carry out Gaussian basecalling and sequence alignment for bisulfite converted reads against the mouse genome[11]. On average, around 1.5 million aligned 27 bp reads (5.4 million 50 bp reads for the PGC libraries) were obtained for each library. For methylation analysis, bases 6 to 22 in the 27 bp reads (bases 15 to 41 in the 50 bp reads) were used, and CpGs were base called as methylated or unmethylated, respectively. Genome-wide averages of DNA methylation of individual samples, or averages of methylation in promoters, exons, introns, remainder of the genome, and different classes of transposons, were bioinformatically determined. For Sequenom MassArray, bisulfite converted DNA was amplified and subjected to quantitative analysis of methylation by masspectrometry.

Methods

Mice and isolation of tissue and DNA samples

Mice deficient for Aid have been described previously[21] and were kindly provided by Dr T. Honjo. These were backcrossed for 7 generations into the C57BL/6J strain during the course of this study. C57BL/6J mice or C57BL/6J mice carrying an *Oct4-Gfp* transgene were used as controls throughout. The *Oct4-Gfp* transgene was bred into the *Aid*-/- knockout following backcrossing into the C57BL/6J background. Epididymal sperm was collected from fertile adult males. PGCs were isolated on a FACS-Aria cell sorter by sorting for green fluorescence; the isolated cell populations were > 98%. Placentas and carcasses were taken from fetuses used for PGC collection. Genomic DNA from wild-type (E14) and *Np95*-/- mouse ES cells was kindly provided by Amander Clarke.

Shotgun bisulfite sequencing and computational analysis of data

Genomic DNA extracted from various mouse tissues with the Qiagen blood and tissue kit were treated with sodium bisulfite and then used to generate Illumina/Solexa sequencing libraries as described previously[11], except that fewer cycles of PCR amplification were used (15 cycles instead of 18 cycles) in order to optimize the base composition of the libraries. For PGC samples, due to the limited sources of tissue, the input DNA amount for libraries had to be reduced to as low as 150 ng. Therefore, the enzymatic reaction steps used for library construction (including reagents and adapters for PCR) were scaled down to accommodate the reduced input amount. On the other hand, more DNA template (in volume) was used in each PCR reaction and more duplicate PCR reactions were performed in parallel in order to obtain equivalent amounts of product as for the other libraries. The libraries were sequenced on an Ultra-high-throughput Illumina/Solexa 1G Genome Analyzer following manufacturer's instructions. Initial sequencing data analysis was performed using version 0.3 of the Illumina/Solexa Analysis Pipeline; subsequently, a previously published highly customized software package was used to carry out Gaussian basecalling and sequence alignment for bisulfite converted reads against the mouse genome[11]. Around 5.4 million aligned 50 bp reads were

coverage of around 5.8% and 1% respectively. Methylated cytosines were identified as cytosines (or guanines as appropriate) in sequencing reads aligned to genomic cytosines, while unmethylated cytosines were identified as thymines (or adenines as appropriate) in sequencing reads aligned to genomic cytosines. Bisulfite conversion efficiency was always above 95% as judged by conversion of cytosines in CHG and CHH contexts (data not shown). The mapped bisulfite sequences were split into three groups. Sequences not spanning a CpG were discarded, and separate lists were made for sequences showing complete methylation or complete demethylation. In the very small number of cases where the same sequence showed both methylation and demethylation it was added to both lists. Where there were multiple datasets for the same sample the methylated and unmethylated lists were merged. Analysis of the data was performed using SeqMonk (www.bioinformatics.bbsrc.ac.uk/projects/seqmonk). The methylated and unmethylated lists were merged together with the methylation status being encoded in the strand of the read (methylated=forward, unmethylated=reverse). A tile of 250 kilobase regions was overlaid on the genome and the methylation status of each tile was calculated. Tiles containing less than 10 reads were discarded, as were tiles where there were 5 or more reads with exactly the same mapped position. The methylation status was calculated as the log2 ratio of the methylated:unmethylated counts. The distribution of values showed a normal distribution and a comparison between tissues was made using a boxwhisker plot which showed the median, upper and lower quartiles and extremities (median +/- 2 × interquartile range). Any values outside this range were plotted individually as outliers. To calculate the methylation levels in specific genomic regions (promoters, genes, introns, exons, transposon families) SeqMonk was used to generate probe regions using the Ensembl features from the annotated NCBIM37 genome as a template. Total counts of overlapping reads in all of these regions across the genome were made and a single methylated:unmethylated ratio was produced. The positions of all repeats in the NCBIM37 mouse genome were extracted from Ensembl and classified into families based on their annotation. A count was made of reads which overlapped with all of these repeat regions and these counts were combined across all members of each family. A single measure per family was then made of the log2 ratio of methylated:unmethylated reads. All repeat families shown are represented by more than 1000 CpG containing reads in each dataset.

Methylation analysis by Sequenom MassArray

DNA from FACS-sorted PGCs was extracted using the AllPrep DNA/RNA Micro Kit (Qiagen). The DNA was then treated with bisulfite reagent using the two-step modification procedure outlined in the Imprint DNA Modification kit (Sigma). Primer pairs were designed using the MethPrimer program (http://www.urogene.org/methprimer/index1.html). A complete list of primers used for analysis is available on request (primers for IAPs were based on the consensus sequence of IAPLTR1a repeats which represent approximately 1.5 % of the ERV-K family). Amplification of the bisulfite converted DNA and preparation of PCR products for quantitative analysis of methylation as detected by the MassArray system was according to the protocol provided by the manufacturer.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Hugh Morgan for his contributions to some of the early analysis of *Aid*-/- mice, Anne Segonds-Pichon for help with statistical evaluation and Jonathan Hetzel for assisting in preparing the Illumina Solexa libraries and their sequencing. We also thank Svend Petersen-Mahrt, Cristina Rada, and Fatima Santos for advice and discussions. C.P. was a Boehringer-Ingelheim predoctoral Fellow. S.F. is a Howard Hughes Medical Institute Fellow of the Life Science Research Foundation. S.E.J. is an investigator of the Howard Hughes Medical Institute. This work

was supported by BBSRC, MRC, EU NoE The Epigenome, and TSB (to W.R.) and by HHMI, NSF Plant Genome Research Programme, and NIH (to S.E.J.).

References

1. Reik W, Dean W, Walter J. Epigenetic reprogramming in mammalian development. Science 2001;293:1089–1093. [PubMed: 11498579]
2. Sasaki H, Matsui Y. Epigenetic events in mammalian germ-cell development: reprogramming and beyond. Nature Rev Genet 2008;9:129–140. [PubMed: 18197165]
3. Oswald J, et al. Active demethylation of the paternal genome in the mouse zygote. Curr Biol 2000;10:475–478. [PubMed: 10801417]
4. Mayer W, Niveleau A, Walter J, Fundele R, Haaf T. Demethylation of the zygotic paternal genome. Nature 2000;403:501–501. [PubMed: 10676950]
5. Dean W, et al. Conservation of methylation reprogramming in mammalian development: aberrant reprogramming in cloned embryos. Proc Natl Acad Soc USA 2001;98:13734–13734.
6. Hajkova P, et al. Epigenetic reprogramming in mouse primordial germ cells. Mech Dev 2002;117:15–23. [PubMed: 12204247]
7. Lee J, et al. Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. Development 2002;129:1807–1817. [PubMed: 11934847]
8. Yamazaki Y, et al. Reprogramming of primordial germ cells begins before migration into the genital ridge, making these cells inadequate donors for reproductive cloning. Proc Natl Acad Sci USA 2003;100:12207–12212. [PubMed: 14506296]
9. Hajkova P, et al. Chromatin dynamics during epigenetic reprogramming in the mouse germ line. Nature 2008;452:877–881. [PubMed: 18354397]
10. Morgan HD, Dean W, Coker HA, Reik W, Petersen-Mahrt SK. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications

24. Teixeira FK, et al. A role for RNAi in the selective correction of DNA methylation defects. Science 2009;323:1600–1604. [PubMed: 19179494]
25. Bhutani N, et al. Reprogramming towards pluripotency requires Aid-dependent DNA demethylation. Nature. 2010 in press.
26. Robbiani DF, et al. Aid produces DNA double-strand breaks in non-Ig genes and mature B cell

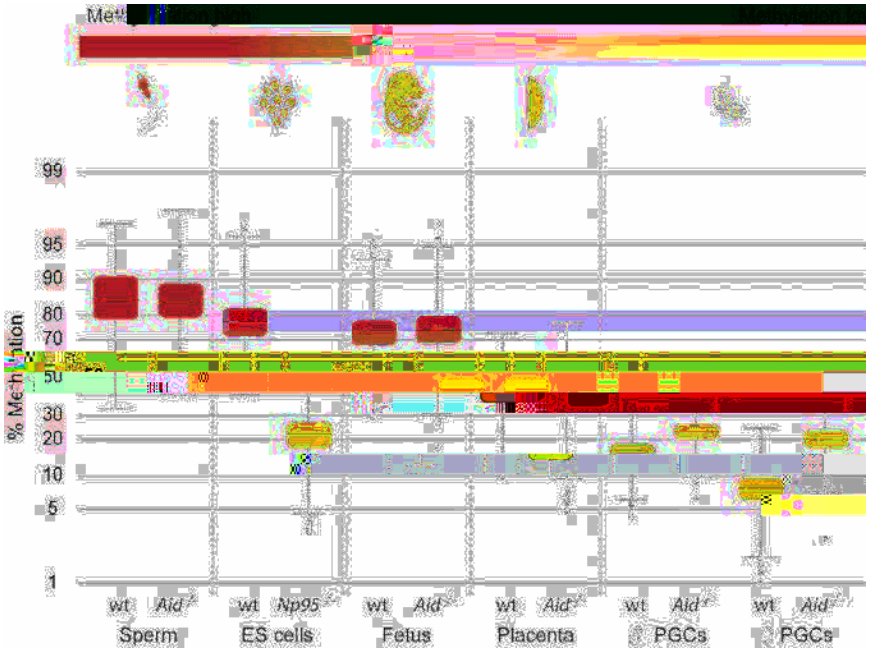

Figure 1. Genome-wide BS-Seq reveals global hypomethylation in PGCs dependent on Aid

Tissues and cells analysed by BS-Seq are shown in a gradient from red to yellow illustrating methylation levels from high to low. BS-Seq reads were analysed using windows of 250 kilobases across the whole genome. Yellow boxes show the range of the 25-75th quartiles of the data, the line in the middle the median value. Whiskers show either highest and lowest values (if there are no outliers) or upper and lower confidence intervals. Outliers are shown as circles. Placenta, fetal carcass and PGCs were collected at E13.5. Note the global hypomethylation in PGCs, and the substantially higher levels of methylation in *Aid*-/- PGCs.

Figure 2. Erasure of DNA methylation in different genomic elements in PGCs

Methylation levels in promoters, exons, introns and intergenic regions in ES cells, *Np95*-/- ES cells, and various tissues of C57BL/6J and *Aid*-/- knockout mice are shown based on ratios of methylated to unmethylated BS-Seq reads. Placenta, fetal carcass and PGCs were all collected at E13.5. Note the particularly pronounced effect of Aid deficiency on methylation of introns and intergenic regions.

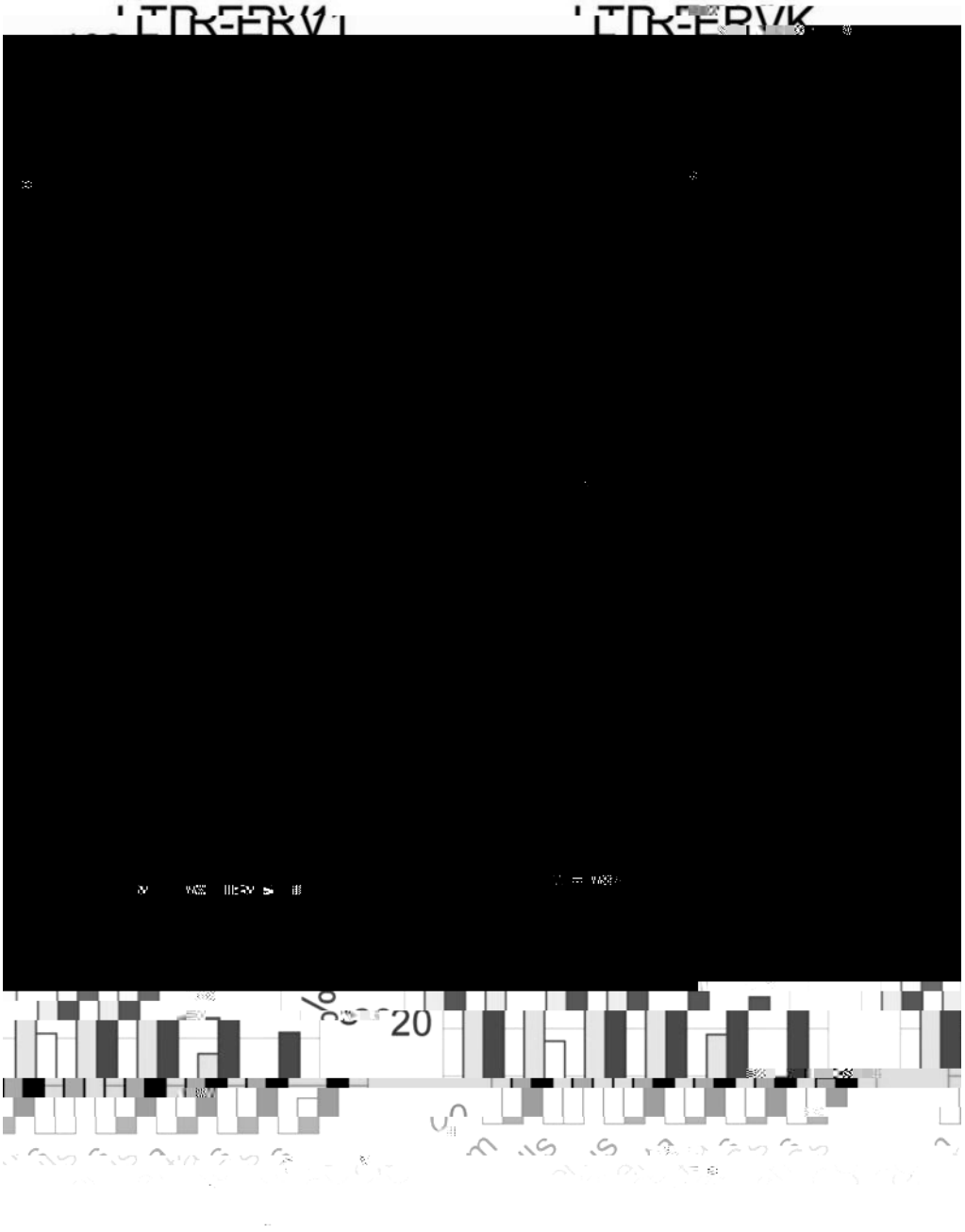

Figure 3. Erasure of DNA methylation in different classes of transposable elements in PGCs

Methylation levels of different classes of transposons in ES cells, *Np95*-/- ES cells, and various tissues of C57BL/6J and *Aid*-/- knockout mice are shown based on ratios of methylated to unmethylated BS-Seq reads. Placenta, fetal carcass and PGCs were all collected at E13.5. Note that LTR-ERV1 and LTR-ERVK elements retain more methylation in PGCs than any other repeat family.

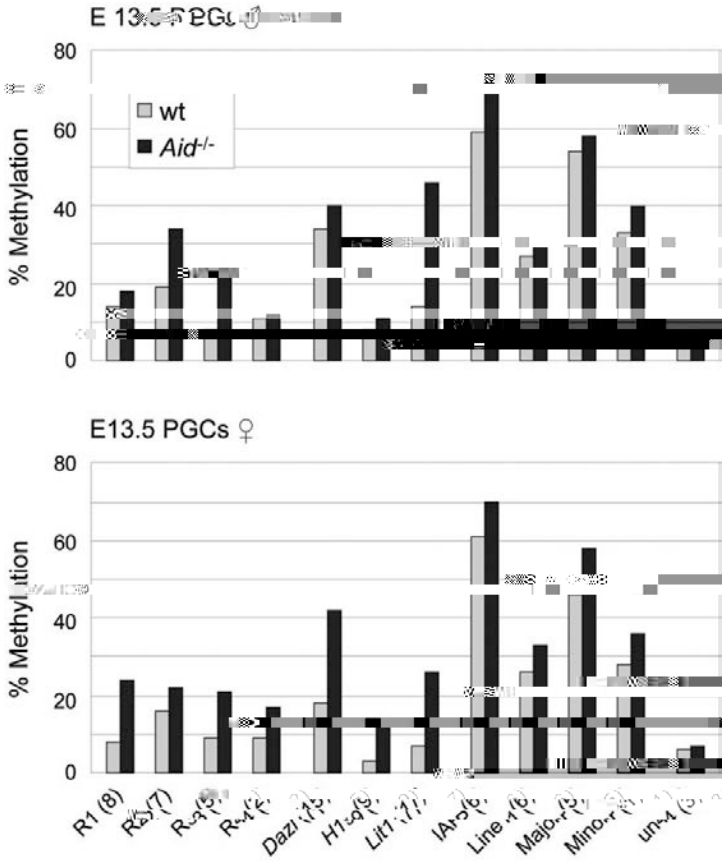

Figure 4. Analysis of DNA methylation of individual genomic loci in E13.5 PGCs by Sequenom MassArray

Methylation levels of individual genomic loci (R1-4, randomly selected sequences, located in intergenic regions, the first intron of *Foxo1* and the seventh exon of *Xirp2*, respectively; the *Dazl* amplicon is located in the promoter region and the *H19* and *Lit1* amplicons are located in the differentially methylated regions), and of transposons and satellite repeats in wildtype and *Aid*-/- E13.5 PGCs are shown based on Sequenom Massarray analysis. The number of CpGs analysed for each region is stated in brackets. Un4, unmethylated control located in the *Hoxc* cluster. Note substantially increased levels of methylation in many genomic loci in *Aid*-/- PGCs.