Mapping the zebrafish brain methylome using reduced representation bisulfite sequencing

Aniruddha Chatterjee1,2,*, Yuichi Ozaki3, Peter A Stockwell4,5, Julia A Horsfield1,2, Ian M Morison1,2, and Shinichi Nakagawa2,3

1Department of Pathology; Dunedin School of Medicine; University of Otago; Dunedin, New Zealand; 2Gravida: National Centre for Growth and Development; Auckland, New Zealand; 3Department of Zoology; University of Otago; Dunedin, New Zealand; 4Department of Biochemistry; University of Otago; Dunedin, New Zealand; 5New Zealand Genomics Limited; Centre for Innovation; Dunedin, New Zealand

Keywords: zebrafish, brain, RRBS, DNA methylation, CpG site, MspI, CpG

Introduction

DNA methylation provides a stable mechanism of epigenetic regulation of gene expression that is prevalent in all vertebrates.1,2 DNA methylation has a crucial role in gene silencing, tissue differentiation, genomic imprinting, X chromosome inactivation, phenotypic plasticity, and disease susceptibility.3,7 In addition, aberrant DNA methylation is a well-established role in tumorigenesis.8

With the divergence of vertebrates, the mechanism of DNA methylation is conserved. In zebrafish, brain, RRBS, DNA methylation, CpG site, MspI, CpG

Promising to be an excellent model for studying trans-generational epigenetic inheritance and reprogramming.

Recently, whole genome bisulfite sequencing of zebrafish has been reported.19,20 These studies have described the distribution of DNA methylation, especially in gametes and during early development. The other methylation studies on zebrafish have used methylated DNA immunoprecipitation (MeDIP) on promoter arrays or sequencing,21-24 and a major limitation of this antibody-based method is that it does not allow investigation of CpG sites at base-pair resolution.25 Reduced representation bisulfite sequencing (RRBS) is an effective alternative approach to whole genome methylation sequencing that generates multiple base-pair resolution methylomes at a reduced cost.26-29 However, the RRBS technique has been primarily used for human and mouse genomes, and its effectiveness has yet been demonstrated in zebrafish. Here, we evaluate the technique of RRBS in zebrafish brain. Brain is predominantly used for human and mouse genomes, and its effectiveness has not been demonstrated in zebrafish brain. Here, we evaluate the technique of RRBS in DNA from zebrafish brain. Brain is probably the most complex vertebrate organ, and epigenetic events have roles in memory formation and learning,30,31 brain development,32 early life stress,30 neurodegeneration,33 neurological and neuropsychiatric disorders,34,35 and establishment of neuronal identity.36

*Correspondence to: Aniruddha Chatterjee; Email: aniruddha.chatterjee@otago.ac.nz
Submitted: 05/06/13; Revised: 07/01/13; Accepted: 07/18/13
http://dx.doi.org/10.4161/epi.25797

http://www.landesbioscience.com
We first performed a simulation of the distribution of MspI fragments: by taking a 200 bp rectangular window and a 20 bp gradient, the proportion of MspI fragments contained in each bin was determined (Fig. 1). The simulation showed that the reduced representation (RR) genome of 40–220 bp contained 20.9% of the total MspI fragments. Selection of fragments shorter than 40 bp included an additional 12.5% fragments. However, we selected 40–220 bp fragments to generate the RRBS library for two reasons. First, very short fragments are likely to map to multiple locations in the genome. Second, selection of 40–220 bp fragments provided the opportunity for a frame-by-frame comparison of zebrafish RRBS with previously published human and mouse RRBS data on the same size selection.37

CpG enrichment in the zebrafish RR genome. For zebrafish, the total size of the in silico RR genome was 31 Mb comprising 264598 MspI fragments of 40–220 bp lengths. It contained 1.43 million CpG sites representing 5.3% of the genomic total, indicating a 2.4-fold enrichment of CpG sites in the RR genome (Table 1). By comparison, the human RR genome has a size of 74 Mb consisting of 647626 MspI fragments. The human RR genome is more than double the size of the zebrafish RR genome, and both are proportional to the size of the whole genome of these organisms. However, the human RR genome, which represents 2.3% of the whole genome, contained 4.1 million CpG sites (13.5% of the genomic total), corresponding to a 5.7-fold enrichment of CpG sites. In mice, the RR genome represents 1.4% of the genome and contains 7.0% of the CpG sites in the genome, corresponding to a 5-fold enrichment. These comparisons demonstrate that CpG dinucleotides are more enriched in the human and mice MspI-generated RR genome of 40–220 bp than that of zebrafish.

RRBS and alignment. We obtained an average of 1.2 Gb of sequence from each of the four zebrafish RRBS libraries. For each sample, 24.5 million sequenced reads of 49 bp in length were supplied. For all four samples, more than 70% of the reads mapped to the reference genome of Zv9 (Table 2). However, the percentage of reads that mapped to multiple locations of the genome was higher (range: 43.6% to 45.5%) than the percentage of the reads that uniquely mapped (range: 27% to 32.7%). The percentage of multiple mapping in zebrafish was 5-fold higher than in our previous Bismark alignment of the human RRBS library, where only 7.7% of the reads (75 bp length) showed multiple mapping against the whole human genome (GRCh37 build).37

High coverage CpG dinucleotides (CpG10). After alignment, we filtered the CpG dinucleotides based on coverage. Only CpG sites covered by 10 or more reads (CpG10) were retained for further analysis. For four of our samples, 429088, 404563, 303757 and 405903 CpG10 were obtained, with a mean coverage ranging...
from to 55 to 77 (Table 3). The distribution of sequenced read coverage of CpG sites in the Male1 sample is shown in Figure 2 as an example (see also Figs. S3–S5 for coverage histograms of other samples). The distribution demonstrates that although the filtered CpGs received high mean coverage, the libraries did not suffer from bias due to excessive amplification of a subset of fragments, should a spurious amplification of fragments exist, an extra peak would be visible on the right hand side of the histogram.

**Global CpG methylation profile of the zebrafish RR genome.** The global CpG methylation ranged from 69.6–75.0% (Table 2) in male and female zebrafish RRBS libraries. The distribution of methylation at each CpG site revealed heavy methylation (>95% methylation) of 40.7 to 42.8% of CpG10 (Fig. 3). By contrast, 10.9% to 12.2% of CpG10 were completely unmethylated bases (<5% methylation). Between 15.8% and 16.6% of the CpG10 showed intermediate methylation (>20% and <80%) (Fig. 3; Figs. S6–S8).

In mammalian genomes, the RRBS protocol has been shown to enrich for CG rich regions (CpG islands) and, as CpG islands remain largely unmethylated in mammalian genomes, the percent methylation of CpGs in RRBS libraries is expected to be lower than the average methylation of the genome. For example, RRBS on rat dorsal root ganglia (analyzed CpG sites = 2.8 million) showed more than half of the CpG sites captured by RRBS were hypomethylated (0–10% methylation) and a fifth of the sites demonstrated hypermethylation (90–100% methylation). However, for zebrafish brain RRBS, the trend was opposite, i.e., >50% of the CpG10 were hypermethylated (>90% methylation) and less than 15% of the CpG sites were hypomethylated (0–10% methylation). Meissner and colleagues performed RRBS for the first time in mouse embryonic stem cells (analyzed CpG sites = 543678 with coverage of ≥10) and reported a similar methylation pattern to rat RRBS, i.e., >40% analyzed CpG sites showed hypomethylation. In the same study, the median percentage CpG methylation for mice brain was shown to be 10 (analyzed CpG sites = 906010 with a median coverage of 14). As a result of much higher prevalence of hypermethylated CpG10, zebrafish brain RRBS showed higher median methylation ranging from 92.4 to 93.3 (Table 4).

Although RRBS and whole genome bisulfite sequencing (WGBS) methylation are not directly comparable (as RRBS covers 5–15% of the CpG sites in the genome enriching for CpG rich regions, whereas WGBS includes 80–100% of the CpG sites including those in repetitive elements), we extracted WGBS data for human and mouse brain and compared their CpG methylation with that from zebrafish RRBS. The results suggested that the proportion of methylated (80–100% methylation) CpG sites is higher in zebrafish RRBS brain methylene compared with the WGBS methylene of human and mice. Further, the percentage of unmethylated (0–20% methylation) bases was higher in zebrafish RRBS samples as well, reflecting the relative enrichment of RRBS for CpG features compared with WGBS (Table S1).

To test whether the striking difference in global CpG methylation in zebrafish RRBS compared with the other mammalian genomes is due to the tissue under investigation (i.e., brain) we performed RRBS of adult zebrafish liver. The global CpG methylation for liver was 68.8% and the median methylation of the CpG10 was 90.9 (number of CpG10 = 219947 with a mean coverage of 21.4) similar to that observed for zebrafish brain samples (Table 4). The distribution of methylation at each CpG site revealed that 63.4% of the CpG10 in zebrafish liver have a methylation percentage ranging from 80–100%; whereas, for 23.8% of the CpG10, the methylation percentage ranged from 0–20% (Fig. 4), with 12.8% CpG10 showing intermediate methylation. Overall, the comparison of the CpG10 methylation distribution between brain and liver (Table 5) indicates the zebrafish methylome, as captured by RRBS, is different than the mammalian genome. In fact, our results are concordant with recent WGBS studies on zebrafish sperm, oocytes, mid-blastula embryo, and muscle (Table 5). In all these tissues, WGBS analysis showed a high proportion of methylated CpGs (80–100%) compared with unmethylated (0–20%).

**Non-CpG methylation in the zebrafish genome.** For all samples, both Bismark and methylKit results suggested successful bisulfite conversion of the genome. We found that the percentage of non-CpG methylation was low in adult zebrafish brain, ranging from 1.7% to 2.9% (Table 2). Previously, heavy non-CpG methylation in zebrafish was reported; however, other bisulfite sequencing experiments have demonstrated very low levels of non-CpG methylation. RRBS studies on rat brain suggested less than 1% non-CpG methylation. On the contrary, whole genome methylation analysis of mice brain suggested 35% of DNA methylation could occur in a non-CpG context (CHH and CHG). Our results, however, show that non-CpG methylation in adult zebrafish brain is minimal. Further analysis in zebrafish liver revealed 3.7% non-CpG methylation, supporting the idea that globally non-CpG methylation in zebrafish genome is negligible and are concordant with recent WGBS studies on zebrafish.

**Reproducibility of zebrafish RRBS methylomes.** One reason for performing RRBS experiments with four pools of DNA (each containing six individuals) was to minimize the effects of biological variability. Figure 3 and Figures S5–S8 show that

---

**Table 2. Details of output and mapping of zebrafish RRBS libraries**

| Sample ID | Male1 | Male2 | Female1 | Female2 | ZF liver |
|-----------|-------|-------|---------|---------|----------|
| Number of reads (millions) | 24.48 | 24.49 | 24.49 | 24.49 | 9.0 |
| Mapping (%) | 78.2 | 76.2 | 71.0 | 74.5 | 71.9 |
| Unique mapping (%) | 32.7 | 32.4 | 27.0 | 30.9 | 40.4 |
| Multiple mapping (%) | 45.5 | 43.8 | 44.0 | 43.6 | 31.5 |
| CpG methylation (%) | 75.0 | 71.5 | 69.6 | 70.0 | 68.8 |
| Non-CpG methylation* | 1.7 | 2.9 | 1.8 | 2.2 | 3.7 |

*As indicated by Bismark, this percentage is the sum of two factors: the actual non-CpG methylation in the genome + possible incomplete bisulfite conversion. However, methylKit analysis showed consistent bisulfite conversion (99%) for all the samples. Male 1–2, Female 1–2 are from zebrafish brain.
sought to determine whether the mapped CpG sites correspond to CpG islands. The list of CpG islands from the SeqMonk feature table was used for this analysis. Regions 2 kb either side of CpG islands were defined as CpG island shores, and regions a further 2 kb from the CpG shores were defined as CpG island shelves. This analysis revealed CpG sites are highly enriched in CpG island shores (46.8% of the total CpG sites), while only 9.25% of the total CpG sites are in the core CpG islands (Table 6). Furthermore, 42.7% of CpG sites are distant from any CpG features, i.e., more than 4 kb distant from either side of a core CpG island. The CpG sites were exported to the UCSC genome browser as a custom track and visualized in comparison with the CpG island track of the browser for Zv9 assembly. The whole genome chromosome-wide view enabled genome-wide visualization of CpG sites and showed several regions where core CpG islands are absent but CpG sites were densely mapped.

In contrast, a recent RRBS study on humans (MspI digested and size selection of 40–220 bp fragments) showed 47.5% and 19.5% of the investigated CpG dinucleotides were in core CpG islands and CpG island shores respectively. Our data indicate that the mapped CpG dinucleotides are more prevalent in CpG island shores of the zebrafish RR genome and less frequent in core CpG islands compared with the equivalent analysis in human.

**Distribution of CpG sites in relation with genes.** We investigated the distribution of the CpG sites relative to the location of genes in the zebrafish genome (gene location information was taken from the SeqMonk feature table information). Interestingly, we found that 45% of CpG sites mapped to gene bodies (Fig. 6A) and a much smaller percentage (7%) mapped to gene promoters (promoters were defined as regions up to ~5 kb from the transcription start site (TSS) of the gene). These results contrasted with human RRBS data where 32% of the investigated CpG sites were in promoters. Fifty one percent of the CpG sites were distant, i.e., further than 5 kb upstream of the gene. Table 7 describes the detailed distribution of CpG sites in relation to TSS and gene bodies. Analysis of gene body-associated CpG sites revealed that 59% reside within introns and 41% within exons (Fig. 6B).
The zebrafish RR genome (40–220 bp) was proportional (1.43 million CpG sites compared with almost 4 million in humans). The overall frequency of CpG sites in the zebrafish genome is much higher than that in human and mice (1.77, 1.04 and 1.15 CpGs per 100 bp respectively). However, the fold enrichment of CpG sites obtained for the zebrafish RR genome was lower than human and mice (2.4-fold compared with 5.7-fold in humans and 5-fold in mice). Although the zebrafish genome has a higher density of CpG sites, CCGG (MspI site) motifs are not as prevalent as in humans. Therefore, in future studies it may be desirable to explore the use of other restriction enzymes that might cut more frequently at the CpG motif. Recently, a double digest of MspI with another methylation-insensitive enzyme was shown to improve the CpG coverage in RRBS experiments in human and mice. A similar approach could be applied to the zebrafish genome for enhanced CpG coverage.

**Discussion**

Zebrafish are an ideal animal model for studying developmental biology since their embryos are transparent, easily accessed from the 1-cell stage, and available in large numbers. Therefore, it is of interest to determine how global methylation patterns in zebrafish compare with those in humans. For this comparison, methylation analysis must be applied in the same detail to zebrafish as it is in humans. RRBS is a tractable method for generating base-pair resolution methylation profile to generate data that could inform biological processes in development and disease.

Here, we have evaluated RRBS in zebrafish and, to our knowledge, we have provided the first single-base resolution DNA methylation map for brain tissue in this organism. The zebrafish genome is almost half the size of the human genome, and considering the relative size of RR genome of both organisms (31 Mb vs. 74 Mb), the number of CpG sites covered in the zebrafish RR genome (40–220 bp) was proportional (1.43 million CpG sites compared with almost 4 million in humans). The overall frequency of CpG sites in the zebrafish genome is much higher than that in human and mice (1.77, 1.04 and 1.15 CpGs per 100 bp respectively). However, the fold enrichment of CpG sites obtained for the zebrafish RR genome was lower than human and mice (2.4-fold compared with 5.7-fold in humans and 5-fold in mice). Although the zebrafish genome has a higher density of CpG sites, CCGG (MspI site) motifs are not as prevalent as in humans. Therefore, in future studies it may be desirable to explore the use of other restriction enzymes that might cut more frequently at the CpG motif. Recently, a double digest of MspI with another methylation-insensitive enzyme was shown to improve the CpG coverage in RRBS experiments in human and mice. A similar approach could be applied to the zebrafish genome for enhanced CpG coverage. For human and mice, as MspI already provides good coverage of core CpG islands, a non

**Table 4.** Per quartile methylation distribution of CpG, in zebrafish RRBS samples

| Sample    | Minimum | 1st Quartile | Median | Mean  | 3rd Quartile | Maximum |
|-----------|---------|--------------|--------|-------|--------------|---------|
| Male1     | 0.0     | 65.9         | 93.2   | 75.0  | 97.8         | 100.0   |
| Male2     | 0.0     | 59.1         | 93.3   | 74.0  | 98.1         | 100.0   |
| Female1   | 0.0     | 50.0         | 92.4   | 71.4  | 97.8         | 100.0   |
| Female2   | 0.0     | 56.3         | 92.7   | 73.1  | 97.9         | 100.0   |
| ZF liver  | 0.0     | 27.3         | 90.9   | 68.7  | 100.0        | 100.0   |

**Figure 3.** CpG methylation distribution in zebrafish brain. The X-axis shows percent methylation for each CpG. The numbers on the bars denote the percentage of CpGs contained in the respective bins.
The zebrafish genome is substantially polymorphic in nature and the quality of the reference genome is not optimal (Sanger Institute release notes for Zv9). In addition, the recently published zebrafish genome indicates that zebrafish repetitive DNA constitutes 52.2% of the genome, the highest repeat content so far recorded for a vertebrate. Increased levels of repetitive DNA for CpG island shore and distant CpGs compared with the core CpG island).

CG containing enzyme (ApeKI, recognition site G^CWGC) was used in combination with MspI to improve coverage in CpG poor regions. However, to apply the double restriction digest method for zebrafish, consideration could be given to the use of another CG motif containing enzyme (for example TaqαI, recognition site T^CGA or BssSI, recognition site C^ACGAG) with MspI as it is likely to improve the coverage within the core CpG islands (we already showed that MspI provides higher coverage for CpG island shore and distant CpGs compared with the core CpG island).

The zebrafish genome is substantially polymorphic in nature and the quality of the reference genome is not optimal (Sanger Institute release notes for Zv9). In addition, the recently published zebrafish genome indicates that zebrafish repetitive DNA constitutes 52.2% of the genome, the highest repeat content so far recorded for a vertebrate. Increased levels of repetitive DNA

![Figure 4. CpG methylation distribution in zebrafish liver. The X-axis shows percent methylation for each CpG site. The numbers on the bars denote the percentage of CpGs contained in the respective bins.](image)

Table 5. Percentage distribution of methylated and unmethylated CpG zebrafish RRBS samples and comparison with zebrafish WGBS*

| Sample          | Unmethylated (0–20% methylation) CpG | Intermediate methylation (20.1–79.9%) | Methylated (80–100% methylation) |
|-----------------|--------------------------------------|---------------------------------------|----------------------------------|
| Male1 (RRBS)    | 15.5                                 | 15.6                                  | 68.9                             |
| Male2 (RRBS)    | 16.4                                 | 16.9                                  | 66.7                             |
| Female1 (RRBS)  | 19.1                                 | 14.5                                  | 66.4                             |
| Female2 (RRBS)  | 17.0                                 | 18.6                                  | 64.4                             |
| Liver (RRBS)    | 23.8                                 | 12.8                                  | 63.4                             |
| Sperm (WGBS)    | 5.0                                  | 1.2                                   | 93.8                             |
| Egg (WGBS)      | 6.3                                  | 29.7                                  | 64.0                             |
| Muscle (WGBS)   | 4.1                                  | 26.7                                  | 69.2                             |
| Sphere† (WGBS)  | 5.7                                  | 0.80                                  | 93.5                             |

*The methylation percentages for zebrafish brain (Male 1–2 and Female 1–2) and liver RRBS samples were calculated on CpG1 as part of the current study. Percentages for zebrafish sperm, egg, muscle and sphere were derived from recently published whole genome bisulfite sequencing.†Four hours post-fertilization.
almost certainly contribute to the observation of increased global methylation in zebrafish relative to human. Further, repetitive sequences and the incomplete genome (the zebrafish genome sequence is 83% complete) provide an explanation for increased multiple mapping in zebrafish compared with human (almost 5-fold higher than in humans). We observed that increasing the sequenced read length to 100 bp improved the unique mapping efficiency to >50% in zebrafish (unpublished data).

The true reflection of CpG distribution in the RR genome should be obtained after alignment and filtering by coverage and not on the initial output from the sequencing, since for subsequent analysis and interpretation only high coverage CpGs will be included. We obtained 0.3 to 0.42 million CpG sites with high coverage in our libraries (CpG10). Considering the higher CpG density of the zebrafish genome, these numbers are low compared with the human RRBS data. Nevertheless, RRBS allowed us to generate a genome-wide nucleotide resolution methylation profile of 0.39 million high coverage CpG sites (on average) in the zebrafish samples. Increased read length or deeper sequencing would significantly increase the number of high coverage CpG sites.

Investigation of CpG10 distribution lead to the interesting observation of higher enrichment of CpG island shores in zebrafish compared with core CpG islands, unlike similar human RRBS libraries. There are multiple methods for defining CpG islands and, as noted by Saxanov et al., definitions of CpG islands are based on “ad hoc thresholds.” For humans, Takai and Jones’s method of predicting CpG island, which is widely used, states the minimum length of CpG island should be 500 bp, whereas for example Gardiner and Frommer defined a minimum 200-bp stretch of CpG rich DNA. Within the SeqMonk feature table (based on Ensembl annotation), the length of the shortest CpG island in zebrafish is 399 bp. Therefore, the definition of CpG islands affects the distribution of mapped CpGs within different CpG features, i.e., if CpG islands are defined as short regions, then more sequenced CpGs will fall outside the defined CpG islands. As the definition of CpG islands is arbitrary, we propose that inclusion of different CpG features (including shore and shelf) provides a more comprehensive approach for methylation profiling in enrichment-based methylation analysis (e.g., RRBS).

For decades, the methylation status of CpG islands received much attention from investigators. However, recent studies showed CpG island shores as crucial elements where DNA methylation status was highly variable between diseased and matched normal tissues. Furthermore, differential methylation of CpG

---

**Figure 5.** Scatter plot and correlation of CpG methylation between zebrafish RRBS methylomes. Scatter plots of percentage methylation values for each pair in four zebrafish libraries (Male1, Male2, Female1 and Female2). Numbers on the upper right corner denote pairwise Pearson’s correlation scores. The histograms on the diagonal are methylation distribution of CpG sites for each sample.

**Table 6.** Distribution of CpG10 in CpG feature context*

| Features          | Number of CpG10 bases | Percentage of total CpG10 |
|-------------------|-----------------------|---------------------------|
| Inside CpG island | 39695                 | 9.25                      |
| CpG island shore  | 200822                | 46.8                      |
| CpG island Shelf  | 4216                  | 0.98                      |
| Outside CpG features | 184289               | 42.97                     |

*These data are generated from Male1 sample, other samples showed similar trend.

**Figure 6.** Distribution of CpG10 in genome and in gene body. (A) Shows the overall distribution of CpG10 of Male1 sample in the genome. (B) Shows the distribution of gene body associated CpG10.
island shores is often tissue specific, and the methylation status of shores correlates strongly with gene expression. In this respect, our results on zebrafish RRBS are promising, as the CpG 10 were indicative. Therefore, high levels of methylation present in the zebrafish brain likely reflect global methylation differences between teleosts and mammals.

Recent studies detailing WGBS analysis of methylation in oocyte, sperm, mid-blastula embryo and muscle in zebrafish have provided detailed information about global methylation of the zebrafish genome. It is notable that the RRBS genomes generated here broadly reflect the global methylation profiles observed using the WGBS method, in that both indicate a much higher percentage methylation in zebrafish than in mammalian genomes. The similarity between RRBS and WGBS generated data indicates that the RRBS method provides a reliable snapshot of methylation in the zebrafish genome. The RRBS method can therefore be used as an economical substitute for WGBS where.

Another feature of interest is the high proportion of mapped CpGs in the gene bodies compared with the promoter regions. The functional role of gene body methylation remains unclear. Recent studies showed sharp transitions in methylation status at exon-intron boundaries. It has been suggested that site-specific occupancy of CTCF pauses RNA polymerase II activity, and that the interaction of CTCF and RNA polymerase II plays role in alternate splicing of genes. It has also been suggested that specific histone modifications mark gene bodies of different classes of genes based on CpG density and methylation status. Furthermore, it is possible that DNA methylation levels are generally higher in gene bodies than in non-genic regions and that gene-body methylation facilitates transcriptional elongation by blocking non-specific or intragenic transcription. However, a recent large-scale meta-analysis called in question the notion that gene body methylation is associated with transcriptional activity, and instead proposed that gene body methylation is determined by accessibility to DNA methyltransferases. The high CpG coverage in gene bodies provided by zebrafish RRBS will allow future exploration of the function and consequences of gene body methylation.

It is likely that, similar to other species, variations in zebrafish methylation regulate tissue-specific gene expression. Modifications of chromatin proteins (such as histones) at enhancers were shown to be associated with tissue-specific cell differentiation in zebrafish. However, little is known about the role of methylation at such enhancers. Further, DNA methylation can determine binding of chromatin architecture proteins such as CTCF, to give cell type-specific chromatin configurations that may instruct gene expression. Zebrafish represent a powerful model for determining the in vivo function of non-coding DNA elements.

This study is limited by its use of a heterogeneous tissue type (brain) and by the poorer quality of the zebrafish genome compared with the human genome. Further, a known limitation of bisulfite sequencing based methods is that the protocol is unable to distinguish difference between 5-methylcytosine and 5-hydroxymethylcytosine. Recent studies indicated presence of 5-hydroxymethylcytosine in human and mouse brain; however, determining the status of potential demethylating modifications such as 5-hydroxymethylcytosine in zebrafish is warranted. Our analysis demonstrates differences in the global methylation signatures and distributions of CpG sites between zebrafish and other mammalian genomes. The base-pair resolution reference methylome of zebrafish provides a resource for future studies to document the functional role of DNA methylation in this organism.

### Materials and Methods

**Ethics statement.** Animal handling and manipulations were conducted in accordance with the guidelines of the University of Otago Animal Ethics Committee under protocol 48-11.

**Sample preparation.** Adult zebrafish AB strains used for this study were maintained at the Otago Zebrafish Facility, Department of Pathology, University of Otago. Brains were dissected from 12 males and 12 females, and then each brain was halved through the sagittal plane. Six halved male brains were combined into a pooled sample referred to as Male1. Similarly, Male2, Female1 and Female2 were comprised of a pool of six halved brain. Livers (n = 10) were harvested by dissection from wild type male and female fish, and pooled. Genomic DNA was extracted from each sample using the PureLink Genomic DNA Mini Kit (Invitrogen) according to the manufacturer’s instructions.

**RRBS library preparation and sequencing.** Bisulfite-converted genomic DNA libraries were prepared according to

| Features          | Number of CpG10 | Percentage of total aligned CpG10 |
|-------------------|-----------------|----------------------------------|
| Gene body         | 192,026         | 44.75                            |
| TSS200 (0–200 bp) | 5040            | 1.17                             |
| TSS500 (0–500 bp) | 7327            | 1.70                             |
| TSS1000 (0–1000 bp)| 10,511          | 2.44                             |
| TSS5000 (0–5000 bp)| 32,531          | 7.58                             |
| TSS10000 (0–10,000 bp) | 54,425       | 12.92                            |
| TSS > 10,000      | 182,571         | 43.37                            |

*The number indicated in each bin of TSS is not exclusive, rather additive, i.e., while calculating the number for TSS500, the numbers from TSS200 is also added to show the total number in that bin.*
scripts and commands were used to describe the distribution of fragments in the genome. Methylation analysis was performed using the R package of methylKit. Briefly, after alignment by Bismark, the SAM files containing uniquely aligned reads were numerically sorted and then processed in R studio (version 0.97.312) using the methylKit package. CpG sites covered by at least 10 sequenced reads (termed as CpG10) were retained to generate the reference methyome. Each sequenced and filtered CpG site was assigned a percentage methylation score. Coverage and correlation plots were generated by methylKit using sorted SAM file for the samples. Human and mouse brain whole genome bisulfite sequencing data for control samples were downloaded from MethylomeDB64 and processed with UNIX and awk scripts (see Table S1).

To investigate CpG10 positions, in relation to the gene and CpG features, the SeqMonk feature table information for Zv9 was used. SeqMonk (freely distributed from Babraham Institute) provide DAT files containing information on CpG islands and genes in zebrafish. These files were parsed by a purpose-written program (identgeneloc), which then identified proximal genes and CpG islands for the CpG10 sites. The resulting information was further processed with awk scripts to generate the distribution of CpG10 positions.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Acknowledgments
We gratefully acknowledge the contribution of Dr Gwenn Le Méé. of the Department of Pathology, University of Otago in Zebrafish liver DNA sample preparation and for helpful discussions. We thank Dr Euan Rodger for assistance and discussions during the preparation of the manuscript. This work is supported by University of Otago research grant and Gravida: National Centre for Growth and Development research grant. AC is supported by a scholarship from Gravida. SN is supported by the Marsden Fund and the Rutherford Discovery Fellowship.

Supplemental Materials
Supplemental materials may be found here:
www.landesbioscience.com/journals/epigenetics/article/25797

References
1. Elango N, Yi SV. DNA methylation and structural and functional bisimilarity of vertebrate promoters. Mol Biol Evol 2008; 25:1602-8; PMID:18469331; http://dx.doi.org/10.1093/molevol/mns140
2. Bock C, Beerman I, Lien WH, Smith ZD, Gu H, Yi SV. DNA methylation and structural and functional bisimilarity of vertebrate promoters. Mol Biol Evol 2008; 25:1602-8; PMID:18469331; http://dx.doi.org/10.1093/molevol/mns140
3. Rollins RA, Haghighi F, Edwards JR, Das R, Zhang MQ, Ju J, et al. Large-scale structure of genomic methylation patterns. Genome Res 2006; 16:157-63; PMID:16365381; http://dx.doi.org/10.1101/gr.436206
4. Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet 2008; 9:465-76; PMID:18463664; http://dx.doi.org/10.1038/nrg2341
5. Igarashi J, Muroi S, Kawashima H, Wang X, Shinojima Y, Kitamura E, et al. Quantitative analysis of human tissue-specific differences in methylation. Biochem Biophys Res Commun 2008; 376:658-64; PMID:18805397; http://dx.doi.org/10.1016/j.bbrc.2008.09.044
6. Charterjee A, Morison DM. Monomeric twins: genes are not the destiny? Bioinformation 2011; 7:369-70; PMID:22352329; http://dx.doi.org/10.2144/05392RV02
7. Layton C, Waddy PC, Odom DT, Bell GW, Yuan B, Danford TW, Wiellette EL, et al. Zebrafish promoter microarrays identify actively transcribed embryonic genes. Genome Biol 2006; 7:871; PMID:16889661; http://dx.doi.org/10.1186/gb-2006-7-8-71
8. Baylin SB, Henriksson M. Altered methylation patterns in cancer genomes: cause or consequence? Cancer Cell 2002; 1:299-305; PMID:12068641; http://dx.doi.org/10.1016/S1535-6108(02)00061-2
9. Michibata L, Williams TD, Zhan H, Gong Z, Chipman JK. Comprehensive profiling of zebrafish hepatic proximal promoter CpG island methylation and its modification during chemical carcinogenesis. BMC Genomics 2011; 12:5; PMID:21205313; http://dx.doi.org/10.1186/1471-2164-12-5
10. Bailey GS, Williams DE, Hendricks JD. Fish models for environmental carcinogenesis: the rainbow trout. Environ Health Perspect 1996; 104(suppl 1):5-21; PMID:8722107
11. Hildebrandt F, Richter A, Flierl S, Deckert K, Grueneberg C, et al. Zebrafish liver DNA sample preparation and for helpful discussions. We thank Dr Euan Rodger for assistance and discussions during the preparation of the manuscript. This work is supported by University of Otago research grant and Gravida: National Centre for Growth and Development research grant. AC is supported by a scholarship from Gravida. SN is supported by the Marsden Fund and the Rutherford Discovery Fellowship.

Supplemental Materials
Supplemental materials may be found here:
www.landesbioscience.com/journals/epigenetics/article/25797
59. Marsman J, Horsfield JA. Long distance relationships: enhancer-promoter communication and dynamic gene transcription. Biochim Biophys Acta 2012; 1819:1217-27; PMID:23124110; http://dx.doi.org/10.1016/j.bbamcr.2012.10.008

60. Kruijswijk E, Liukkonen Z, Klimasauskas S. 5-Hydroxymethylcytosine—the elusive epigenetic mark in mammalian DNA. Chem Soc Rev 2012; 41:6916-30; PMID:22842880; http://dx.doi.org/10.1039/c2cs35104h

61. Chatterjee A, Rodgers EJ, Stockwell PA, Weeks RJ, Morison IM. Technical considerations for reduced representation bisulfite sequencing with multiplexed libraries. J Biomed Biotechnol 2012; 2012:741542; PMID:23193365; http://dx.doi.org/10.1155/2012/741542

62. Krueger F, Andrews SR. Bismark: a flexible aligner and methylator caller for Bisulfite-Seq applications. Bioinformatics 2011; 27:1571-2; PMID:21493656; http://dx.doi.org/10.1093/bioinformatics/btr167

63. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biol 2012; 13:R87; PMID:23034086; http://dx.doi.org/10.1186/gb-2012-13-10-r87

64. Xin Y, Charron B, O’Donnell AH, Milekic M, Costa R, Ge Y, et al. MethylomeDB: a database of DNA methylation profiles of the brain. Nucleic Acids Res 2012; 40(Database issue):D1245-9; PMID:22140101; http://dx.doi.org/10.1093/nar/gkr1193

65. Smith ZD, Gu H, Bock C, Ginishe A, Meissner A. High-throughput bisulfite sequencing in mammalian genomes. Methods 2009; 48:226-32; PMID:19442738; http://dx.doi.org/10.1016/j.ymeth.2009.05.003