DNA methylation analysis in *Crassostrea gigas* using Nanostring nCounter Technology

Background

In collaboration with Nanostring, assays were developed to detect DNA methylation in oyster samples. This technology would allow for methylation detection in multiple genomic regions at a time in order to assess if DNA methylation patterns were dynamic between different developmental stages, tissue types or in response to different environmental exposures (see Table 1. for a description of samples). The assays combined methylation specific restriction enzymes MspI (methylation insensitive) and HpaII (methylation sensitive) with Nanostring's nCounter technology (http://www.nanostring.com). Fifty-six regions were targeted for analysis based on, 1) the presence of a restriction site (CCGG), 2) suitability of sequence surrounding the site for probe design (performed by Nanostring), and 3) the biological interest in the probe. The following report summarizes the results from this collaboration. Supporting documents, including the collaboration document and a summary of the codeset design, can be found in Gavery & Roberts 2014.

Assay Date	Description of Samples
11/05/2011	Gill A (sample prepped for Phase I repeatability assay)
	Gill C (sample prepped for Phase I repeatability assay)
	Gill F (sample prepped for Phase I repeatability assay)
	Gill H (sample prepped for Phase I repeatability assay)
11/14/11	OA gill (no additional description)
	EE2 gill (96hr exposure to ethinylestradiol 10/27/11)
	Larvae (no additional description)
	Male gamete (no additional description)
12/19/11	Female gonad (vinclozolin exposed parent ID: 17)
	Male gonad (control parent ID: 19)
	5-azacytidine treated larvae (DNA isolated by SJW 6/8/10)
	Larvae (V. tubiashii exposed (ETS 8/23/10)
3/21/12	Female gonad (control parent ID: 28)
	Male gonad (control parent ID: 10)
	Larvae 155 (ETS 4/19/11)
	Larvae 159 (ETC 4/19/11)
4/23/13	Male gonad of control offspring (vinclozolin experiment) ID: 44
	Male gonad of control offspring (vinclozolin experiment) ID: 45
	Male gonad of control offspring (vinclozolin experiment) ID: 57
	Male gonad of control offspring (vinclozolin experiment) ID: 58
6/5/13	Male gonad of vinclozolin exposed offspring ID: 3
	Male gonad of vinclozolin exposed offspring ID: 4
	Male gonad of vinclozolin exposed offspring ID: 20
	Male gonad of vinclozolin exposed offspring ID: 21

Table 1. Description of samples assayed using the Nanostring codeset.

Summary of Results

- A majority of the loci are unmethylated across all samples tested regardless of tissue type, developmental stage or treatment. See Figure 1 below (error bars are +/- 1SD).

- Four of the loci show methylation (above 20%). One locus appears to be methylated across all samples (EW779247_392). One locus is polymorphic for methylation depending on sample type (EU342886_1129). The remaining two loci are not strong performers across all samples, but appear to be hemi-methylated in samples where the probe was valid. Each of the 4 probes with methylation will be described in more detail below.

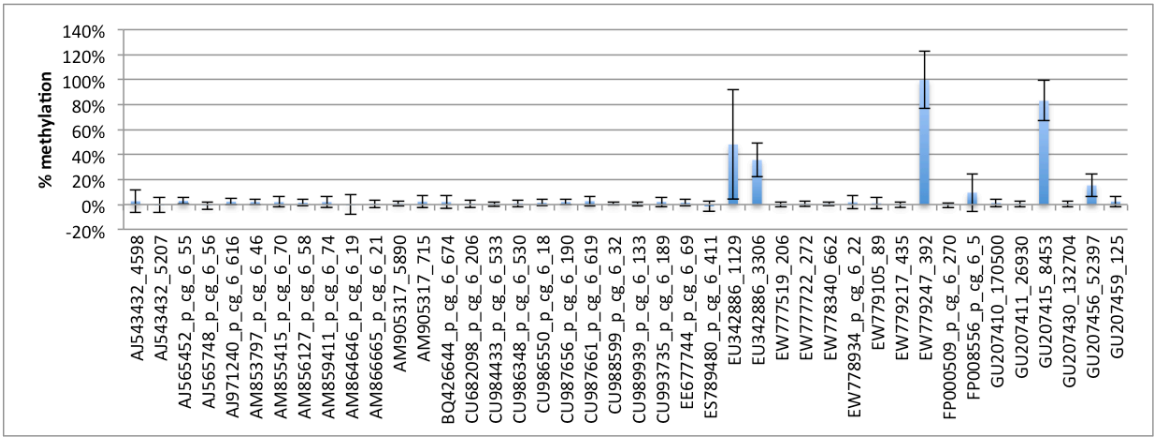


Figure 1. Average % methylation for each probes across all samples (n=23). A majority of the probes are unmethylated regardless of sample type. Four probes show methylation (>20% 5mC) in one or more of the samples.

- The most interesting locus is in EU342886_1129 (annotated as hexokinase (in intron)). Figure 2 shows % methylation for each sample. This probe was not valid for a few of the poor performing samples (EE2 gill, female gonad and larvae 155). All of the gill samples show low methylation, while the other tissue types: gonad and larvae show high methylation. Initial verification using bisulfite sequencing (Figure 5) corroborate these results. For the bisulfite sequencing, 7 CpG sites were covered in the region of the probe. There are a very limited number of clones that were sequenced (The uppermost numbers in the legend of Figure 3 indicate the number of clones analyzed for each CpG), but the trend is similar to what is observed from the Nanostring assay. This corroboration would be strengthened by sequencing additional clones.

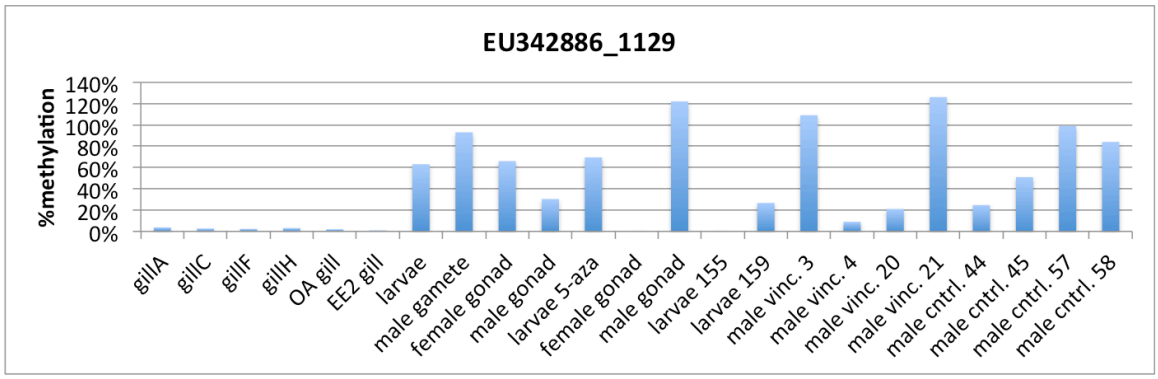
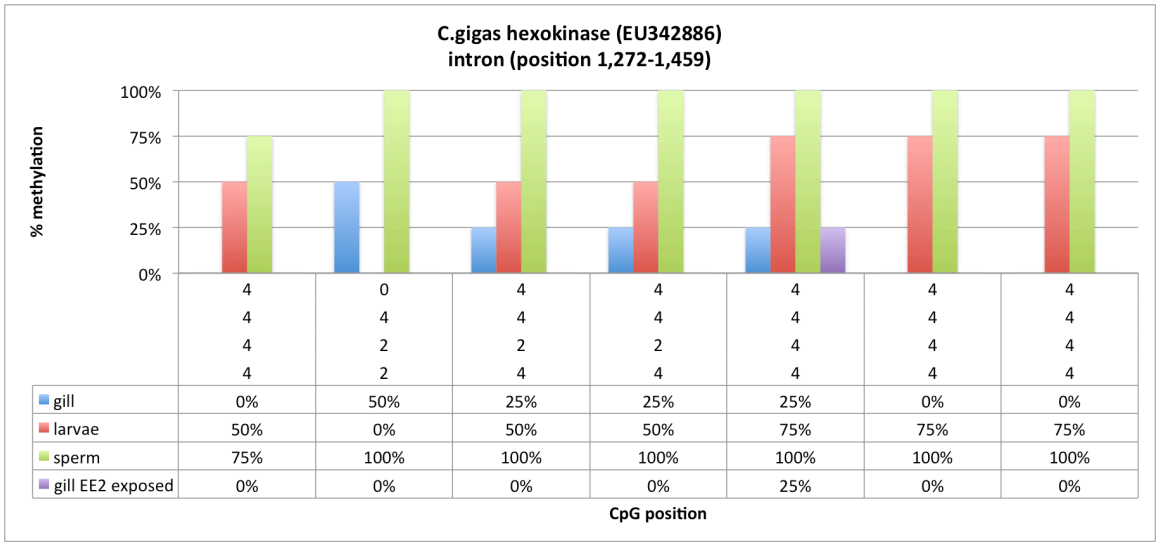


Figure 2. Percent methylation for the region of probe EU342886_1129.



	4	0	4	4	4	4	4
	4	4	4	4	4	4	4
	4	2	2	2	4	4	4
	4	2	4	4	4	4	4
gill	0%	50%	25%	25%	25%	0%	0%
larvae	50%	0%	50%	50%	75%	75%	75%
sperm	75%	100%	100%	100%	100%	100%	100%
gill EE2 exposed	0%	0%	0%	0%	25%	0%	0%

Figure 3. Bisulfite sequencing results for 7 CpGs in C. gigas hexokinase gene. The uppermost numbers in the legend represent the number of clones sequenced for each sample. The lower number represent the average % methylation at each CpG position.

- There is a second probe in the hexokinase gene (EU342886_3306) the probe also covers the first intron (this regions was not covered by the bisulfite sequencing). This region shows methylation between 15 - 50% for 5 of the samples across all assays (Figure 4). The issue is that this probe is a poor performer and was not a valid probe for the other 18 samples. I am less confident about these results from a quantitative perspective considering the irreproducibility between samples. The assay may not be working in the other samples due to a SNP in the restriction site, but methylation would need to be confirmed by bisulfite sequencing before reporting this site as methylated in these 5 samples. What is interesting about this region is that the results are indicating a hemimethylated state in these samples. This could be confounded in the larvae samples, which are pools (i.e. half of larvae could be fully methylated other half unmethylated), but is more interesting for the female gonad sample and the male gonad vinclozolin treated sample.

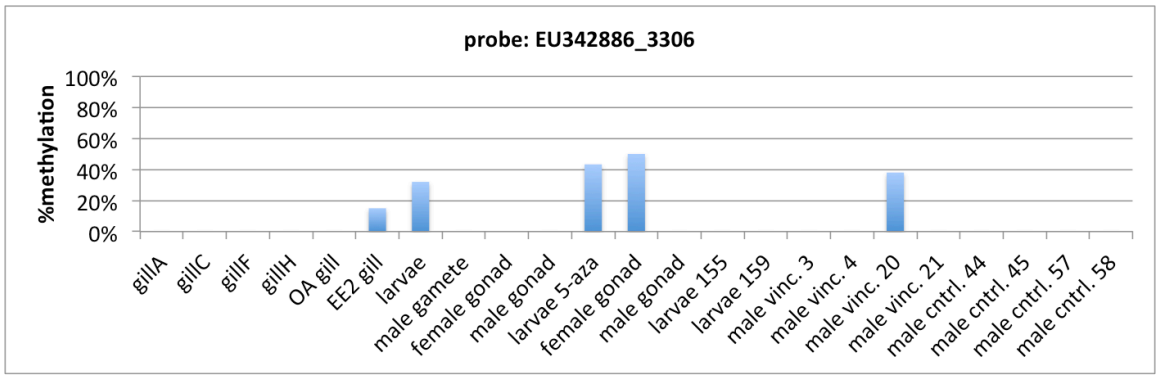


Figure 4 Percent methylation for the region of probe EU342886_3306.

- The probe EW779247_392 (annotated as phosducin, in exon) shows heavy methylation across all samples with valid probes (this probe was not valid for samples without visible bars). Figure 5 shows the methylation for each sample. This region appears to be constitutively methylated regardless of tissue type or developmental stage.

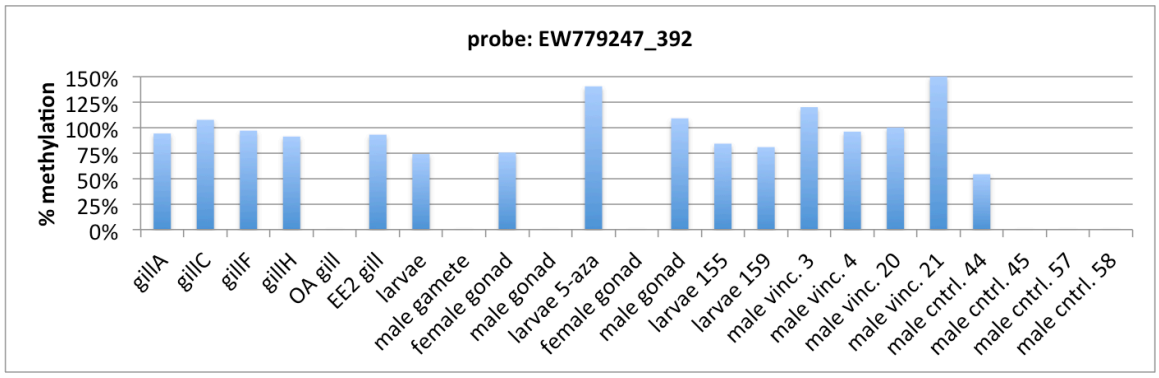


Figure 5. Percent methylation for the region of probe EW779247_392.

- The last probe that showed some indication of methylation is GU207415_8453 (annotated as tubulin tyrosine ligase-like family, coding region). Again, similar to EU342886_3306 this probe was not very robust across assays, see Figure 6 (all samples without visible bars were invalid), so bisulfite sequence validation would be required.

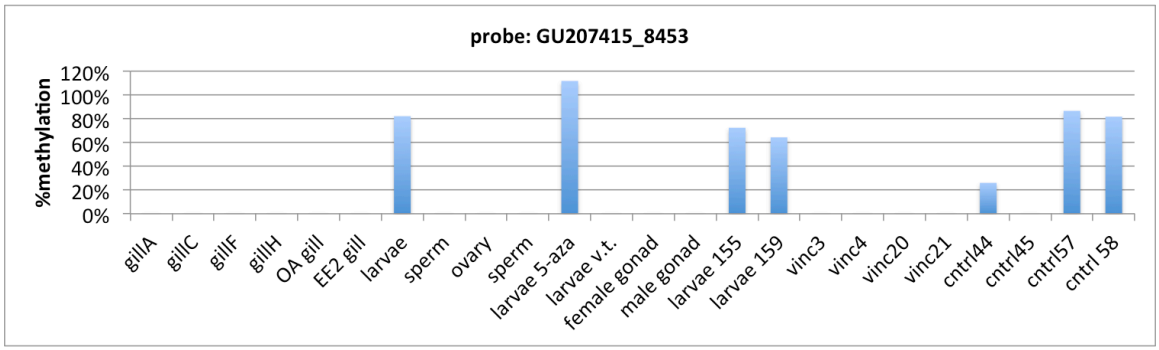


Figure 6. Percent methylation for the region of probe GU207415_8453.

Conclusions:
This collaboration was an effort to apply the Nanostring nCounter technology to assess DNA methylation at targeted loci. I believe this effort was limited by the lack of a genome at the time of assay development. Probe selection was challenging, as very little gDNA sequence was available. Instead, we used mostly expressed sequence tag (EST) resources, meaning some probes may cross exon/intron boundaries. The lack of a genome also made it impossible to test for specificity of the probes. As a result, a number of the probes did not provide reliable results. The other issue is that very little was known about where methylation was in the oyster genome so probes were selected somewhat randomly (although we did use CpG observed to expected ratios to try to select some loci that were predicted to be methylated).

There were some interesting results based on this collaboration. First, we identified the first differentially methylated CpG site between oyster tissues/life stages (probe: EU342886_1129). We also identified a number of consistently unmethylated CpG loci as well as one consistently methylated CpG.

Beyond validation of the methylated probes, I don't think this codeset would be useful going forward due to the lack of 'informative' (i.e. methylated) sites. However, now that the *C. gigas* genome is available, this technology could be used to develop a new codeset if probes were designed based on predefined regions of interest.

Additional information

Methods

DNA is isolated from genomic DNA and ~1ug is sent to NanoString. NanoString performs the enzyme digests. All digests include the AluI rescriction enzyme, which fragments DNA to sizes appropriate for the hybridization (determined *in silico* during probe design). There are three digests that are assayed: AluI alone, MspI + AluI and HpaII + AluI. The fragmented DNA is then used as input into the nCounter assay using the codeset designed for this project (see useful links above for a summary of the codeset). Nanostring provides raw and normalized (see Description of controls below for normalization process) count data, which, if validity criteria are met, are used to calculate percent methylation of a particular probe: HpaII +AluI counts/AluI counts x 100%.

Description of Controls:
Positive controls: A dilution series of positive spike-in RNA hybridization controls (not oyster specific) is run in each assay. Positive control counts are used to normalize all platform-associated sources of variation (e.g. automated purification, hybridization conditions, etc.). To generate the normalization factor (lane-specific), the sum of positive control counts are calculated for each lane and divided by the average of these sums (n=12 lanes per assay). This lane-specific scaling factor is then applied to all of the counts generated for that lane. There is variability in the total counts of the positive controls between assays (columns below are dates of each assay), but in general the positive control scaling factors were all within the expected range (0.3 – 3). The exception is the 06/05/13 assay, where the positive controls failed to produce counts.

Nanostring felt the data were still usable but the counts are not normalized. Regardless, it's important to note this difference between the 06/05/13 assay and all of the other assays.

Negative controls: No template, negative controls are run in each assay. Nanostring generated a 'background' value that was average negative control counts +2 standard deviations (SD). For the final analysis, I did not report % methylation values for probes where the counts of the AluI digested samples were less than 100, which is conservative compared to the 'background' value calculated by Nanostring. Again, there is much higher background for the 06/05/13 assay, so for the final analysis AluI only counts needed to be >200 before % methylation was reported.

Probe validity criteria:

There were a number probes that performed very poorly across all assays. Of the 56 targets, 15 probes consistently had very low counts. Therefore, these probes were removed from further analysis and a total of 41 target loci were evaluated for methylation status.

In order to report percent methylation (%5mC) for a certain probe in an assay, two criteria needed to be met. First, the AluI digest had to be >100 counts (or >200 for the 06/05/13 assay). Second, the background had to be less than 25%. Percent background is calculated as the ratio of the MspI counts/AluI counts. This is designed to control for the presence of the CCGG restriction site. If the counts of the MspI digested sample are not significantly lower than the undigested sample, it indicates that there was no restriction site present (or alternatively the outer cytosine is methylated and therefore MspI can't cut, there is no way to tell which is happening). The graph below shows the proportion of assays that were valid for a given probe (total samples = 23) Total number of probes on the x-axis is 41. This graph shows that not all probes performed well across all samples. This is likely due to polymorphisms (e.g. no restriction site present or poor hybridization of probes), but sample quality could also affect the probes. For example, the larvae + V.t. sample had low counts over all. See the next section for additional information about sample performance.

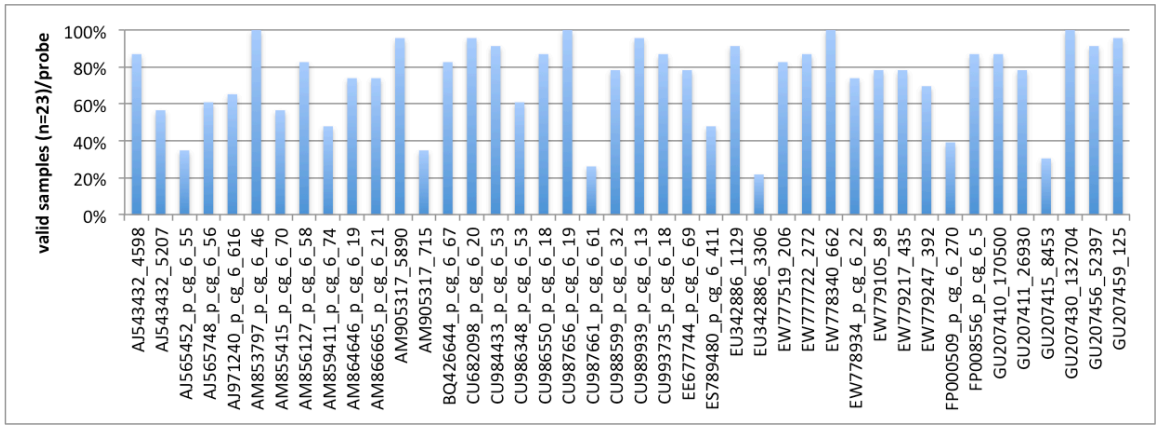


Figure 1. Probe validity summary (as percent valid samples). Probes showed variable performance across all 23 samples.

Sample performance: For a few of the samples, the number of valid probes was very low. For example, 0% of the probes were valid for the larvae + V.t sample (121911). This could be due to the quality of the DNA or the accuracy of the spec reading (e.g. perhaps not enough DNA was digested or run on the assay). This sample is not included in further analysis. Other samples that did not perform well were the female gonad sample P28 (03/12/13) or the OA gill (11/14/11). The results for the probes that passed validity criteria may still be accurate, but the overall quality of the sample should be noted. The graph below shows the number of valid probes per sample. The poorly performing samples stand out, but it should also be noted that all of the samples had probes that were not valid either due to low AluI counts or inability of MspI to cut (high background). The proportion of valid probes is out of 41 total probes (i.e. the consistently poor performing probes have already been removed).

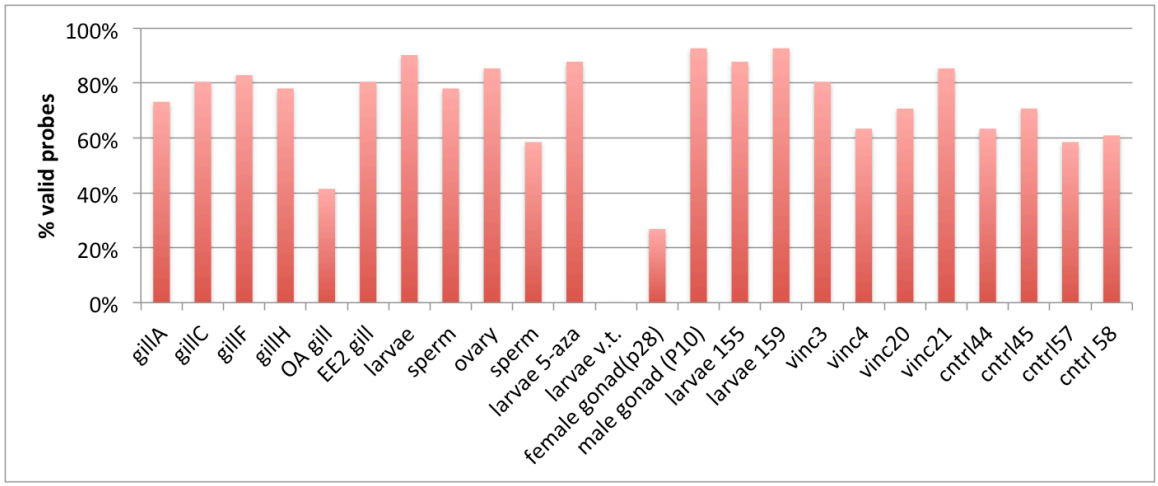


Figure 2. Summary of the proportion of valid probes per sample (as percent valid probes). Some samples performed poorly across a majority of the probes (i.e. larvae V.t., female gonad (p28) and OA gill).

Phase I (Assay development) summary of results:

Pilot () : The purpose of this assay was as a first look at overall counts and to see if the probes from 1st codeset that were included in the 2nd codset were performing as expected. Many of the probes from the 2nd codeset showed high counts (definitely a larger proportion than the 1st codeset) and the probes that had been previously tested performed as expected.

Titration (08/05/2011): Triplicates of 250, 500 and 837 ng of Alu digest and a no digest control were assayed and the counts increased linearly with the amount of input DNA (as expected). Even at the highest concentration the sum of the counts is well below saturation of the assay with this probe set so it was recommended that ~1ug of DNA could be used in the assay going forward.

Repeatability (10/05/2011): This assay showed high repeatability between biological replicates in gill tissue. The standard deviation between biological replicates is <10%

5mC for all probes. Prior this assay the variation in methylation status between biological replicates was unknown.

Works Cited

Gavery & Roberts 2014. Figshare…