Comparison of Methylation at Imprinted and Non-Imprinted Genes

Clementine Payne and Dr. Tamara Davis
Biology Department, Bryn Mawr College, Bryn Mawr PA 19010-2899

Introduction to Genomic Imprinting and Methylation

During development, both maternal and paternal DNA contributes to an embryo's genome. Much of the development of an organism is regulated by both the maternal and paternal copies of these genes. However, when a gene is imprinted, only one of these two parental copies is expressed (Figure 1). Imprinted expression, a mammalian-specific phenomenon, is regulated through epigenetic modification or DNA methylation, the addition of a methyl group (-CH_3) to cytosine in a CpG dinucleotide (**Figure 2A**). This change in DNA chemistry affects the expression of the gene (**Figure 2B**). Differential methylation of the parental alleles is established during gametogenesis and is maintained throughout development by DNA methyltransferase 1. The methylated copy is typically silent while the unmethylated copy is expressed (**Figure 2B**). Methylation patterns acquired during gametogenesis, referred to as primary differentially methylated regions (DMRs), have proven to be stable across development. In contrast, research suggests that methylation patterns are less stable at secondary DMRs, which acquire methylation during embryonic development, suggesting that the epigenetic modifications that are acquired during post-implantation development are regulated differently.

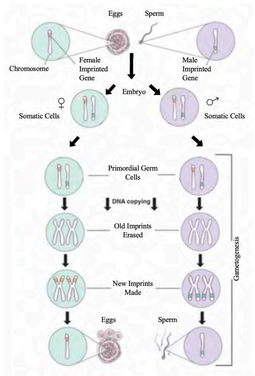


Figure 1: An overview of how imprinting marks are established in gametes and maintained in somatic cells. Figure adapted from Jirtle & Weidman (2007) *Am Sci* 95:143-149.

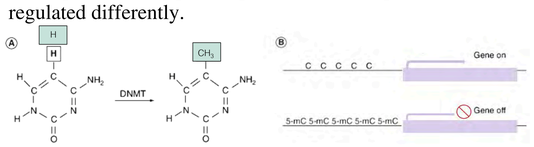


Figure 2: (A) Cytosine is converted to 5-methycytosine through the methylation of the 5-carbon in the cytosine ring.
(B) The presence of 5-methycytosine frequently results in gene silencing.
Figure adapted from Lorincz, Attila. (2014). Expert review of molecular diagnostics. 14. 293-305. 10.1586/14737159.2014.897610.

Experimental Design

Through observation, high levels of hemimethylation were detected at secondary DMRs, but not at primary DMRs (Figure 3). This observation is inconsistent with the mechanistic activity of DNA methyltransferase 1, which works to maintain methylation. This study aims to determine if methylation instability and high levels of hemimethylation are unique features distinctive to secondary DMRs. This will be achieved through a comparison of the methylation stability and symmetry at non-imprinted, tissue-specific DMRs (tDMRs) to that of the variability and asymmetry of the secondary DMRs. In order to investigate methylation patterns, the complementary strands of tDMRs were ligated together using a hairpin loop (**Figure 4**). By using a hairpin loop with a barcode embedded within the sequence, an analysis of both the coding and complementary strands of the DNA was made. The barcode, a segment of random nucleotides, allowed us to identify unique sequences following PCR.

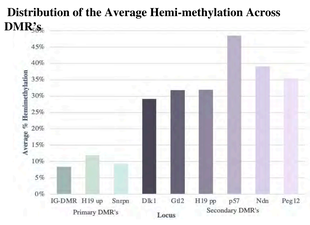


Figure 3: Visualizes the average range of hemimethylation distributed across differently methylated regions (DMRs).This figure illustrates that primary DMRs (IG-DMR, *H19* up, and *Snrpn*) had an average of about 10% hemimethylation. In contrast, the secondary DMRs (*Dlk1*, *Gtl2*, *H19* pp, *P57, Ndn*, *Peg12*) were documented as having hemimethylation averages ranging from 29 to 49 percent. These data points echo observations made within the field that secondary DMRs have a higher rate of hemimethylation. Adapted from Nechin et al. (2019).

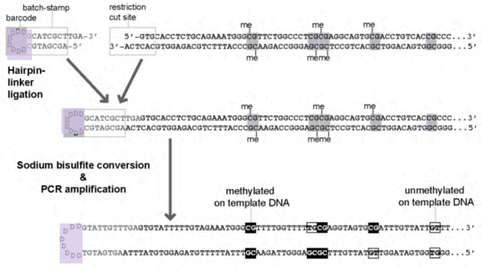


Figure 4: The steps required to digest and ligate DNA before performing bisulfite mutagenesis and PCR. The linkers and hairpin in this image are not the ones used during this experiment. The barcode is a random sequence that allows us to eliminate duplicate clones from our analysis.
Figure adapted from: Stöger, Genereux, Hagerman, Hagerman, Tassone, & Laird (2011) PloS one. 6. e23648.

Acknowledgements

Support for this research was provided by Bryn Mawr College Summer Science Research program. Additional support for this research was provided by awards to TLD from the National Science Foundation (NSF Grand number 1514600) and the Bryn Mawr College Data Science Incitive.

Methodology

1. Bisulfite Mutagenesis: Converts unmethylated Cytosine into Uracil
2. Polymerase Chain Reaction: Amplifies DNA
3. Gel Electrophoresis: Confirms amplification of the DNA
4. Ligation: Ligates PCR product into bacterial vector
5. Transformation: DNA introduced into a cellular environment to replicate.
6. Mini Prep: Purify plasmid DNA
7. Restriction Digest: Confirms the presence of inserted DNA through visualization of the enzymatically cut product
8. Sequencing: Generates and analyzes sequences to find CpG dinucleotides; determines methylation patterns

Figure 5: Illustrates the molecular processes that were performed to complete this project.

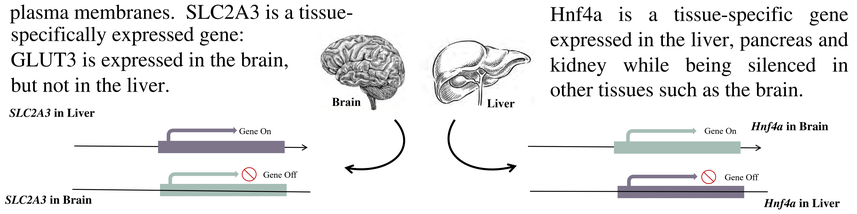
Analysis of non-imprinted genes *Hnf4a* & *SLC2A3*

To assess whether the variability in DNA methylation patterns at secondary DMRs are a unique feature of these loci, it was important to assess the methylation at both imprinted and non-imprinted loci. In order to do this, two genes were chosen that have been reported as having tissue-specific differential methylation correlating with tissue-specific expression. The two genes that we selected are as follows.

SLC2A3: Codes for the GLUT3 which facilitates the transport of glucose across the plasma membranes. SLC2A3 is a tissue-specifically expressed gene: GLUT3 is expressed in the brain, but not in the liver.

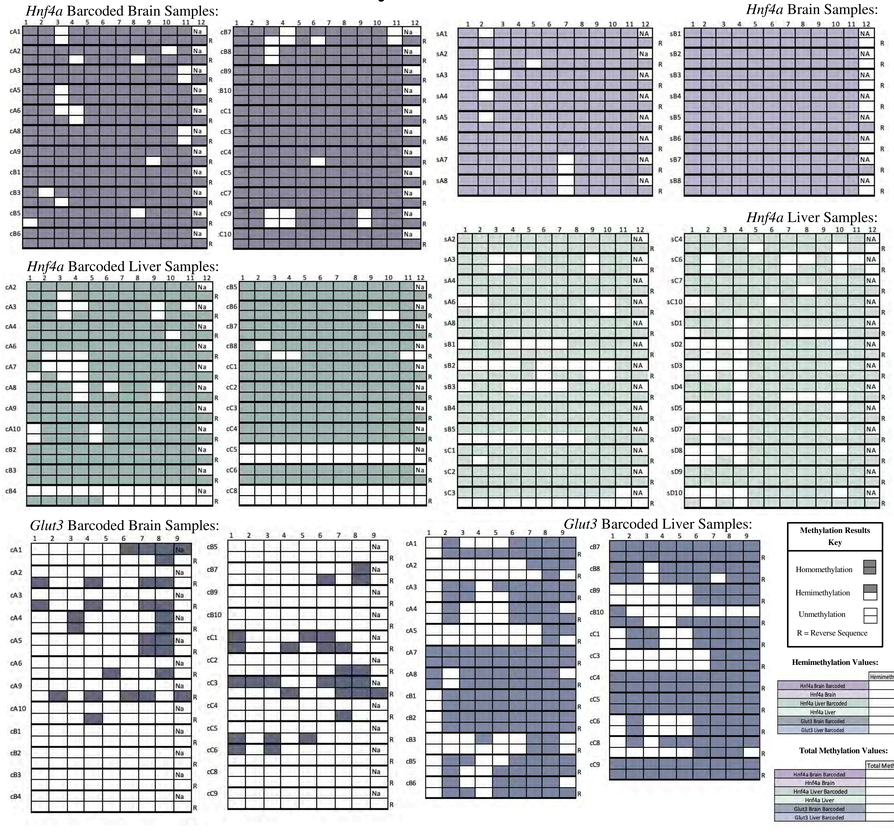
Hnf4a: Codes for hepatocyte nuclear factor- 4 alpha that acts as a transcription factor.

Hnf4a is a tissue-specific gene expressed in the liver, pancreas and kidney while being silenced in other tissues such as the brain.



We chose to look at tissue-specific DMRs so we would have datasets containing methylated and unmethylated sequences depending upon which tissue was being analyzed.

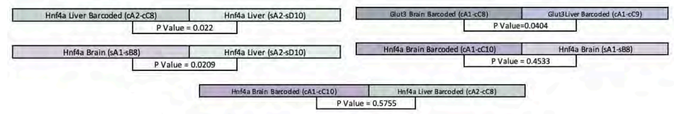
Methylation Results



Statistical Analysis Results

Overall Methylation Patterns: The methylation patterns observed in the *Glut3* data were as hypothesized: low levels of methylation were observed in the brain where *Glut3* is expressed and high levels of methylation were observed in the liver where the gene is not expressed. Unexpectedly *Hnf4a* had high levels of methylation in both the brain and the liver .

Mann-Whitney U test Two Tailed Analysis: Allows for a comparison of the distribution of methylation patterns to be made between samples.



When comparing the distribution patterns of methylation statistically, significant p-values were produced for the *Glut3* samples as expected. However, we did not observe consistent significant differences at *Hnf4a,* probably due to the high level of methylation in tissues.

Chi Squared Analysis: Allows for an evaluation of the amount of hemi-methylation detected in samples to be compared to one another. Through conducting the Chi Squared analysis we will be able to test our hypothesis that the amount of hemimethylation at nonimprinted genes will look like primary, not secondary, DMRs.

	H19-up	IG-DMR	Snrpn
Hnf4a Brain Barcoded (cA1-cC10)	0.019039392	0.117373813	0.075732777
Hnf4a Liver Barcoded (cA2-cC8)	0.82014166	0.3951449	0.5860896
Glut3 Brain Barcoded (cA1-cC8)	5.41E-34	1.84E-54	2.08E-47
Glut3 Liver Barcoded (cA1-cC9)	5.27E-02	9.10E-04	3.59E-03

Figure 7: P-Values of Primary DMRs vs Methylation Data

When looking at the data in **Figure 7** we did see a partial confirmation of our hypothesis. The p-values produced from the *Hnf4a* data suggests that its hemimethylation pattern is like primary DMRs. However, *Glut3* shows more hemimethylation then expected and as a result does not look like the hemimethylation patterns observed at primary DMRs, differing from our hypothesis.

	H19-pp	Gtl2	Dlk1	p57	Ndn	Peg12
Hnf4a Brain Barcoded (cA1-cC10)	2.55 E-13	1.83 E-14	1.39 E-13	1.61754 E-29	4.51208 E-20	8.47253 E-18
Hnf4a Liver Barcoded (cA2-cC8)	3.11E-08	3.01E-09	1.90E-08	9.18E-22	6.30E-14	6.00E-12
Glut3 Brain Barcoded (cA1-cC8)	3.73E-13	1.85E-10	1.39E-13	1.62E-29	4.51E-20	8.47E-18
Glut3 Liver Barcoded (cA1-cC9)	1.83E-03	3.05E-04	1.35E-03	5.74E-12	3.12E-07	8.63E-06

Figure 8: P-Values of Secondary DMRs vs Methylation Data

Previously we hypothesized that hemimethylation in our samples would be significantly different from secondary DMRs, due to the unexpected asymmetric methylation observed at those loci. **Figure 8** illustrates that *Hnf4a* is, as predicted, significantly different from the secondary DMRs studied. Similarly, when analyzing the *Glut3* samples, p-values indicated that its hemimethylation levels were also different than secondary DMRs.

Discussion

Neither of our genes were methylated as expected:

-***Hnf4a*:** While the samples produced from our *Hnf4a* data supported our hypothesis, the gene was methylated in both brain and liver tissue.

-***Glut3:*** While *Glut3* was methylated as expected (it was methylated in liver tissue and unmethylated in brain tissue), its hemimethylation levels were not similar to either the primary or the secondary DMRs. In liver the hemimethylation levels were intermediate while in the hypomethylated brain DNA hemimethylation levels were extremely high.

These results indicate that a comparison between genes which are regulated through tissue specific expression is not an adequate model for comparison with methylation at imprinted primary and secondary DMRs.

Future Research

We are currently initiating an analysis of genes that have strain specific differential methylation in order to better model the same kind of allele specific distribution of methylation that we see at imprinted genes. In F_1 hybrid mice, one allele of these genes will be methylated while the other is unmethylated based on genetic strain (BL/6 vs. *castaneus*). We will continue to focus our analysis on methylated alleles to test our hypothesis that the methylation at these non-imprinted loci will be stable like those observed within the primary DMRs.