Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Apoorwa Thati Date

Primer Optimization for DNA Methylation Analysis in 22q11.2 Deletion Target Genes

By

Apoorwa Thati
Master of Science

Biology

_____

Brad Pearce
Advisor

_____

Victor Corces
Committee Member

_____

William Kelly
Committee Member

_____

Alicia Smith
Committee Member

Accepted:

_____

Lisa A. Tedesco, Ph.D.

Dean of the James T. Laney School of Graduate Studies

_____

Date

Primer Optimization for DNA Methylation Analysis in 22q11.2 Deletion Target Genes

By

Apoorwa Thati
B.S., Emory University, 2012

Advisor: Brad Pearce Ph.D

An abstract of

A thesis submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science

in Biology

2012

Abstract

Primer Optimization for DNA Methylation Analysis in 22q11.2 Deletion Target Genes

By Apoorwa Thati

Methylation of cytosine residues of CpG (C-phosphate-G) sites is the most characterized of the epigenetic mechanisms and involves the addition of a methyl group onto the 5' position of a cytosine residue. About 60-90% of CpGs are found methylated throughout the genome, but unmethylated CpG sites can be found clustered in CpG islands most often associated with promoters of the gene. Methylation of these CpG islands is most often associated with transcriptional silencing and has been found to be a significant contributor to gene expression. The mechanism of methylation as a silencing signal is thought to occur by either recruitment of repressive transcriptional silencing machinery or by steric hindrance, preventing the binding of transcriptional factors. We are investigating the methylation modifications of genes *CLTCL1* and *DGCR8* in the typically deleted region of patients with 22q11 Deletion Syndrome, by designing and optimizing successful primers for bisulfite treated DNA. Future examination of the methylation patterns of these target genes in patients will be done to help clarify the mechanistic connection between 22q11 Deletion Syndrome and schizophrenia.

Primer Optimization for DNA Methylation Analysis in 22q11.2 Deletion Target Genes

By

Apoorwa Thati
B.S.

Advisor: Brad Pearce

A thesis submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science

in Biology

2012

Acknowledgements

This thesis would not have been possible without the help of several integral people. I'd like to thank Dr. Myfanwy Hopkins for her guidance, keeping me on track with this project, and for being a reliable mentor. I'd like to thank Dr. Benjamin Youngblood for his valuable input on my project, and his constant availability to any of my questions. Additionally, this thesis would not have been possible without the constant guidance and support given by Dr. Alicia Smith. Her instruction for the methylation protocol was essential and I would not be where I am with this work if it were not for her. I'd also like to thank my committee members, Dr. Bill Kelly and Dr. Victor Corces for their time and input throughout this process. I appreciate their patience with my project's progress.

This work would not have been possible without the enormous contribution of Jessica Rowell, both inside and outside the lab. Her intellectual curiosity is inspiring. She helped keep me motivated and ensured that this was a constant learning process. We would not be where we are with the progress of this work without Jessica. Thank you Jessica.

Finally, I'm truly grateful for the opportunity Dr. Brad Pearce gave me by allowing me to come into his lab and learn the scientific process. I appreciate his constant support, effort and patience with my project. The past 2 and half years has been a process of intellectual growth and I'm indebted to him and his lab.

Table of Contents

Introduction:

22q11.2 Deletion Syndrome (22q11DS) is characterized by a hemizygous interstitial microdeletion located on chromosome 22, spanning about 43 contiguous genes.[1] The syndrome affects an estimated 1 in 4000 live births, making it the most common microdeletion.[2] Symptoms of the 22q11DS include peripheral manifestations such as congenital heart disease, facial abnormalities, thymic hypoplasia, velopharyngeal abnormalities, cleft palate, learning difficulties, autoimmune disorders, and mental illness.[2 3] The 22q11.2 locus contains genes that encode proteins involved in synaptic pruning, an aspect of brain development that is associated with cognitive efficiency.[4] 22q11DS patients show a trend of delayed cortical thinning in pre-adolescence yet increased thinning in adolescence suggesting abnormal synaptic pruning mechanisms, which may be a risk factor for psychosis.[5] People with 22q11DS have a 25-fold increased risk of schizophrenia, suggesting that hemizygous deletion at this locus has functional consequences for neurodevelopment. The 22q11 region is polymorphic and carries a number of low copy repeats (LCRs), which are presumed to play a role in the high incidence of 22q11DS.[2] Candidate genes of great clinical interest in the 22q11.2 deletion region most notably include catechol-O-methyltransferase (*COMT).*[5]

We are particularly interested in investigating the relationship between the increased incidence of schizophrenia among people with 22q11DS and the abnormal copy number of certain genes in the deletion region. Schizophrenia is a neuropsychiatric illness characterized by psychotic symptoms, such as delusions and hallucinations, as well as negative symptoms like avolition, reduced emotional response, and abnormal affect.[5]

While the specific causes of schizophrenia are largely unknown, literature suggests that genetic and environmental factors contribute to etiology of this disease. The heritability of schizophrenia has been investigated extensively, but genome-wide association studies (GWAS) have failed to discover SNPs that can explain the high heritability estimates.[6] Evidence shows environmental factors including prenatal events significantly influence the incidence of schizophrenia and other neuropsychiatric disorders.[7] In this regard, the phenomenon of epigenetics may play a role. Epigenetics is the study of heritable changes in gene function that occur without a change in the sequence of the DNA.[8] One of the most telling aspects of epigenetics is apparent in twin studies. Monozygotic twins have essentially the same DNA sequence and often times, similar environmental exposures. This being the case, discordant phenotypes, as in the case of approximately 40% of monozygotic twin schizophrenia cases, cannot be attributed to genetic sequence variation.[9] Recent studies have indicated that epigenetics variation can help explain this disease discordance.[10] [11] The two most well studied epigenetic mechanisms are histone modifications and DNA methylation.

Histone acetylation, a common histone modification, is catalyzed by histone acetyltransferases (HATs). This modification involves a transfer an acetyl group from acetyl-coenzyme A to lysine residues on the N-termini of histone proteins. Acetylation results in conformational relaxation of the chromatin complex and thus allowing access for transcription of genes.[12] On the other hand, histone deacetylases (HDACs) reverse the action by transfer of acetyl groups back to coenzyme A, resulting in chromatin complex condensation and promoting decreased gene transcription.[12]

A seemingly independent epigenetic mechanism, DNA methylation, is actually highly interconnected with histone modification.[12 13 14] The mechanism of DNA methylation involves a covalent transfer of a methyl group to cytosine nucleotides from the homocysteine-methionine cycle.[12] Specifically methylation occurs on CpG sites (C-phosphate-G), which describes any area of DNA containing a cytosine nucleotide upstream to a guanine nucleotide.[15] The process of DNA methylation occurs via DNA methyltransferases (DNMTs). In mammals, DNMT1 is primarily used for maintenance methylation of hemi-methylated DNA after DNA's semi-conservative replication process.[12] This ensures that methylcytosine marks persists through cell division.[16] DNMT3a and DNMT3b are involved with de novo methylation, which is the initial action of methylation on previously unmethylated sites.[16] Methylated cytosines bind to methyl-DNA-binding proteins, which have both methyl-DNA binding domains (MBDs) and transcription-regulatory domains (TRDs). These domains recruit other proteins, which in turn attract HDACs to the CpG site, deacetylating histone proteins and altering the chromatin complex to promote transcription suppression.[12 14 16] In general the result of DNA methylation is seen to be transcription silencing.

CpG sites characteristically are found in groups called "islands". CpG islands are defined as a region with at least 200 base pairs and with a cytosine guanine content of at least 50 percent.[17] Within the human genome, approximately 70% of CpG sites are normally methylated.[12 18] Methylation studies have tended to focus on CpG islands within promoter regions since they tend to be areas of hypermethylation and hence transcriptional

silencing. However, methylation does not exclusively occur in CpG islands. Recent studies have shown that methylation also occurs in regions 2 kilobase pairs upstream and downstream of CpG islands, termed CpG shores. CpG shores have been associated with transcriptional repression and tissue-specific DNA methylation.[19 20]

Various studies have investigated associations between differential methylation and disease. In cancer, notable findings have implicated global hypomethylation in chromatin instability and promoter-specific CpG island hypermethylation in transcriptional gene inactivation.[17 21] Global hypomethylation is associated with DNA instability, partially because it normally occurs in areas of the genome that should be less activated, such as repetitive sequences, transposons, and endoparasitic sequences.[17] Hypermethylation at specific CpG islands can affect genes involved with cell cycle, apoptosis and other cancer regulatory genes.[17 22] Additionally, the colon cancer and breast cancer methylomes have shown tissue-specific differential methylation in areas upstream and downstream of CpG islands (CpG shores) suggesting CpG shore regulation of specific genes.[19 20]

In relation to neuropsychiatric disease, studies have suggested differential methylation is associated with specific phenotypes. One study found global hypomethylation in leukocytes of schizophrenic patients when compared to controls. However, patients that took haloperidol showed higher methylation percentages, which suggest a molecular mechanism or side effect linked to antipsychotic medications.[23] Additionally, the same study showed hypermethylation in the soluble catechol-O-methyltransferase (S-COMT) region of schizophrenic patients. A similar methylation study showed hypomethylation in

the promoter region of membrane-bound catechol-O-methyltransferase (MB-COMT), which corresponded with higher gene-expression of MB-COMT transcripts. The study suggested that the over-expression due to promoter hypomethylation might increase dopamine degradation in the frontal lobe potentially providing a mechanism for schizophrenia and bipolar disorder.[24] Similarly, in relation to 22q11DS and schizophrenia, DNA methylation is a proposed mechanism that could modulate expression of the remaining copy of genes in the 22q11 region. However, there are few studies of gene methylation patterns for genes in this region, other than *COMT*.

The methylation pattern of DNA in the 22q11.2 deletion region is of considerable interest due to the combination of genetic and environmental effects in the etiology of schizophrenia. Our long-term hypothesis is that in patients with 22q11DS there is a different methylation pattern in *DGCR8* (and *CTLCL1*) compared to control subjects without a known chromosomal abnormality. The methylation pattern may correlate and/or compensate for gene expression. We postulate that unusually low levels of DNA methylation for a subset of genes enhances expression levels that would normally be diminished in patients with only one copy of such genes. Variation between individuals in methylation may be a factor in determining which 22q11DS patients develop schizophrenia.

Determination of DNA methylation patterns requires implementation of protocols to quantitate and differentiate methylated cytosines from unmethylated cytosines. Sodium bisulfite conversion protocol is a well-established technique in methylation literature, and

is often considered more definitive than other techniques.[25 26 27 28] The protocol involves using sodium bisulfite to convert unmethylated cytosine residues to uracil residues in single-stranded DNA, but leave 5-methylcytosine residues unaffected. After successful conversion, all unmethylated cytosines from the original sequence will have been converted to uracils, which are later converted to thymine, and all methylated cytosines will still remain as cytosines. To analyze methylation patterns in a specific region, specific primers must be designed to bind to the newly converted DNA. Downstream sequencing of the amplified region will display this converted pattern allowing researchers to pinpoint locations in the genome, which have methylated cytosines, and also identify the proportions of methylated cytosines.[25 26 27 28]

However well established, this protocol still faces a number of challenges. Due to the acidic conditions of the reaction, DNA often fragments during bisulfate conversion. Additionally, complete conversion of all unmethylated cytosines is difficult to guarantee.[28] Furthermore, there are a number of recognized limitations to designing successful primers for bisulfite converted DNA. The difficulty of finding appropriate primers hinders the scope of many candidate gene methylation analyses. For example, a number of studies have examined limited numbers of CpG sites in a given candidate gene.[23 29] While still informative, the validity and applicability of methylation studies would be improved with examination of greater number of CpG sites.

In our study, we chose to optimize 2 regions in the genes *CLTCL1* and *DGCR8* containing CpG islands and shores of the 22q11.2 deletion for future DNA methylation

analysis. The two target genes from the deletion region were picked based on their potential involvement in the abnormal psychological phenotype that occurs frequently in 22q11DS patients. *CLTCL1* is of high interest for its role in synaptic vesicle formation, an integral part of the neurobiology of psychiatric disorders.[30] *DGCR8* is a double-stranded RNA binding protein that has a functional role in the Dicer pathway. Dicer is involved in RNA interference and post-transcriptional silencing.[30 31 32] Dicer is an endoribonuclease in the RNase III family that cleaves double-stranded RNA (dsRNA) and precursor microRNA (pre-miRNA) into mature 22-nucleotide micro-RNAs (miRNAs). Micro-RNAs are small non-coding RNAs that act in regulating translation via cleavage or transcriptional repression of target transcripts.[31 32] Dicer catalyzes a key step in the RNA interference pathway and initiates formation of RNA-induced silencing complex (RISC). This is capable of degrading mRNA transcripts.[30] It is a novel mechanism for epigenetic silencing. *DGCR8*, along with RNase III enzyme Drosha, acts in the nucleus to prepare the precursor-microRNA for Dicer-mediated processing in the cytoplasm.[31 32 33] Thus levels of *DGCR8* are proposed to play a critical role in the effectiveness of the Dicer pathway.[30 31 32] The lack of one copy of *DGCR8* in 22q11 DS is hypothesized to have a large impact on this pathway, and DNA methylation of CpGs in *DGCR8* may influence the ability of a cell to compensate for this loss.[34] The importance of micro-RNA mechanisms to schizophrenia has received growing attention in the literature, yet little is known concerning *DGCR8*, despite its potential connection with psychiatric symptoms in the 22q11DS.

My goal was to optimize multiple primers in order to thoroughly span the CpG islands in these genes and to further examine CpG shores 2-kilobase pairs upstream and downstream of the island for a total of a 5-kilobase pair region around each island.[19] Successful primers can then be used in future studies as an invaluable source for interrogating DNA methylation analysis of this region.

Materials and Methods

1. ***Note:*** *All names and labels for figures/tables throughout this thesis are placed below each figure/table itself.*
2. ***Note:*** *All primer pair shorthand throughout this thesis includes the first letter of the gene name, followed by the island area letter (if applicable), and concluded with primer pair number. For example, for gene CTCL1, primer pair 3, the shorthand would be C3. For gene DGCR8, island area B, primer pair 4, the shorthand would be DB4*

DNA Methylation Analysis of Target Genes

In order to focus on the best targets, forty-three genes from the common deletion area on 22q11.2 were prioritized. These are the 22q11.2 deletion target genes of interest also identified by Weksberg et al., 2007. Prioritization of genes was based on abundance of supporting evidence and relevance to neuropsychiatric disease. The list of forty-three genes from the 22q11.2 deletion region is below. A subset of these genes was prioritized over the other in the region and is the basis of this study.

AIFM3	*SEPT5*	*ARVCF*	*C22orf25*
C22orf29	*C22orf39*	*CDC45L*	*CLDN5*
CLTCL1	*COMT*	*CRKL*	*DGCR14*
DGCR2	*DGCR6*	*DGCR6L*	*DGCR8*
GNB1L	*GP1BB*	*GSC2*	*HIRA*
TRMT2A	*KLHL22*	*LZTR1*	*MED15*
MRPL40	*P2RX6*	*PI4KA*	*PRODH*
RANBP1	*RIMBP3*	*RTN4R*	*SCARF2*
SERPIND1	*SLC25A1*	*SLC7A4*	*SNAP29*
TBX1	*THAP7*	*TSSK2*	*TXNRD2*
UFD1L	*ZDHHC8*	*ZNF74*	

Obtaining patient samples

Blood samples are provided from the Atlanta Veterans Affairs Medical Center (VAMC, Dr. E. Duncan) and Emory Autism Center (Dr. O. Ousley), as well as Dr. E. Walker in the Department of Psychology, Emory Hospital. Patients were defined as controls lacking any known chromosomal or medical diagnosis, those with verified 22q11.2 deletion, patients with prodromal psychosis, and patients diagnosed with schizophrenia. We obtained a particular subset of B cells, CD 19 cells, from all participants isolated from a peripheral blood draw. Cells where chosen due to their relative accessibility and because the majority have some level of expression of 22q11 genes of interest. Once the B cells were isolated they were cultured for 10 days in the presence of IL-2 and SAC to stimulate growth. Then cell bullets were spun down and frozen at -80°C until implementation of DNA isolation and sodium bisulfite conversion protocol. Primer designs, sodium bisulfite treatment of the DNA and polymerase chain reaction optimization are the key components of the protocol. For initial optimizations and troubleshooting of the protocol, we used transformed B-cells since these yield inexhaustible DNA.

Prioritization of target genes

Genes chosen from the deletion range were based on potential significance in 22q11 DS psychiatric phenotypes. The genes chosen were *CLTCL1* and *DGCR8*.

Choosing cells to implement protocol

Bisulfite conversion and PCR optimization was done using transformed B-lymphocytes obtained from Coriell Cell Repositories. We chose an Apparently Healthy individual as a

representative control for methylation analysis and a DiGeorge Syndrome individual, as a representative case. Cells were cultured using appropriate B-lymphocyte culturing technique (RPMI 1640, 2mM l-glutamine, 15% fetal bovine serum). Pellets were frozen dry at 5 million cells/pellet and stored at -80° C until needed for DNA extraction.

Overview of Bisulfite Conversion Sequencing Approach

Bisulfite sequencing involves treating DNA with bisulfite to convert unmethylated cytosine residues to uracil. Hence, the treatment leaves 5-methylcytosine residues intact and unaffected.[25 26 27 28] Because the bisulfite treatment induces specific changes to the DNA sequence that depends on the natural methylation status of the organism, it allows determination of methylation pattern. Downstream analyses, such as polymerase chain reaction (PCR), can be implemented to retrieve the methylation sequence.[26] The objective of bisulfite sequencing is therefore to induce a site-specific change. In this case, the nucleotide would shift from a cytosine residue to a thymine residue.

After successful bisulfite conversion, the DNA product was used for polymerase chain reactions (PCRs). The primers designed specifically for CpG sites were used for the PCR, allowing the regions of methylation to be amplified. Gel electrophoresis was performed to separate the amplified regions and confirm expected amplicons. Confirmed amplicon products will then be sequenced for analysis of methylation patterns. We considered 2 types of sequencing protocols: first being a ligation/bacterial transformation and second direct sequencing from PCR or gel product via high-throughput sequencing (Illumina V3, Emory Genomics Core).

Primer design for bisulfite converted DNA

Part I

Once the target genes were chosen, primers were hand-designed based around CpG islands, usually in the promoter region of the gene (Dr. Benjamin Youngblood). Areas with high levels of CpG repeats are considered to be sites, or islands, when they extend for 200 base pairs or more. The UCSC Human Genome Browser assisted in mapping the exact locations of CpG islands for the target genes. We located 1 island for target gene *CLTCL1* and 2 islands for target gene *DGCR8* as seen in ***Figures 1 and 2***.

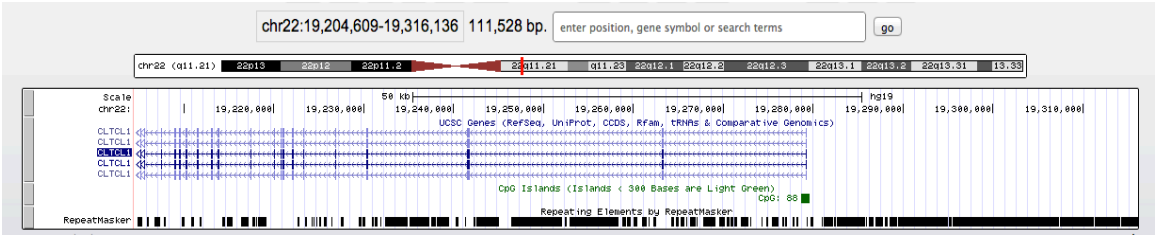


Figure 1: Map of *CLTCL1* from UCSC genome browser. Green areas indicate CpG islands

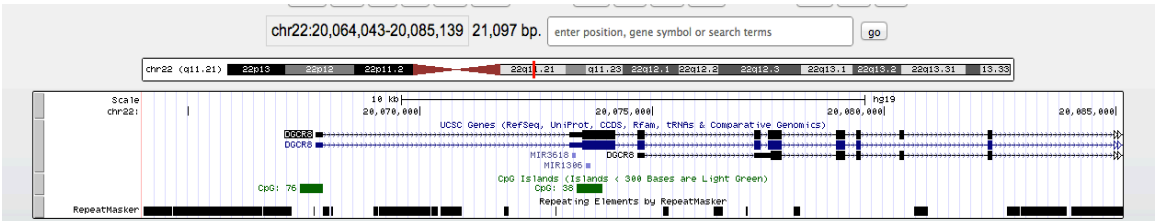


Figure 2: Map of *DGCR8* from UCSC genome browser. Green area indicated CpG islands

Once located, a set of guidelines was followed to aid in successful design of primers.

Guidelines

1. Primers should not contain CpG sites in sequence
2. Primers should start/end in guanine residue.
3. Primers should be between 20 to 30 base pairs.
4. Primer amplicon range should be under 500 base pairs.
5. Once chosen, all cytosines residues within the primer sequence must be replaced with thymine residues for final primer sequence.

The primers chosen for each target gene were the following:

CLTCL1

1. BSfor 5’GAGTTTTTGTTTAAAGTAGGGGGTTAG

1. BSrev 5’CTAAAAAATAATAAAAATAAAAACC

2. BSfor 5’GTTTTAGTTTAGGTGGGAGG

2. BSrev 5’CAAACCCCATCAAAATTAATCCC

3. BSfor 5’GAGTTTAGGGTTATTTTTTATTTAG

3. BSrev 5’CCTTAACTAAACCCTCACCAAAAAAC

DGCR8

1. BSfor 5’gttttggaaattttgtattagtaaagggg

1. BSrev 5’caaccatctaaaaaacctatccaaactctc

2. BSfor 5’gagagtttggataggttttttagatggttg

2. BSrev 5’ccatactcattattacaataaaac

3. BSfor 5’gttaatagtgtttggtttttaatttgg

3. BSrev 5’CTATCCATCACCACCAAAACC

4. BSfor 5’GTTTTTTTTTGATTTTAAGTTGTTTAAGG

4. BSrev 5’ctaaaaaccctctaaaaaaaaaactacc

Implementation of handpicked primers yielded poor PCR amplification results and unsuccessful ligation and bacterial transformation. Because of these results, we decided to implement an alternative protocol to accelerate progress that involved direct sequencing.

Part II

By using high-throughput sequencing instead of Sanger sequencing, we were able to interrogate a greater region of our genes of interest within our constraints of time and manpower. Although the handpicked primer design method was instructional, many of the designed primers failed to amplify the expected region or failed to amplify any region at all. We decided to use a bioinformatics online program. We chose an online tool called Sequenom EpiDesigner, which designed primers specifically for bisulfate-converted DNA based on our parameters.

In this round of primer design our goal was to interrogate not only the CpG islands of interest but also the 2kb-regions upstream and downstream known as CpG islands shores.[19 20] For each CpG island region, a total sequence length of approximately 5 kilobases (2 kb upstream, ~1 kb island, 2 kb downstream) was entered into Sequenom's parameters for primer design.

The parameters we chose for the Sequenom EpiDesigner program were the following: Primer Tm: minimum 52° C, optimal 60° C, maximum 68° C; Primer size: minimum 18 base pairs, optimal 25 base pairs, maximum 32 base pairs; Product size: minimum 200 base pairs, optimal 325 base pairs, maximum 450 base pairs.

We selected primers for 3 areas of interest (areas of interest included upstream shore, island, and downstream shore). Gene *CLTCL1* had 1 area of interest and *DGCR8* had 2 areas of interest. ***Figures 3 and 4*** display Sequenom EpiDesigner's output of various

primer regions for each of our areas of interest. We chose a subset of EpiDesigner's output.

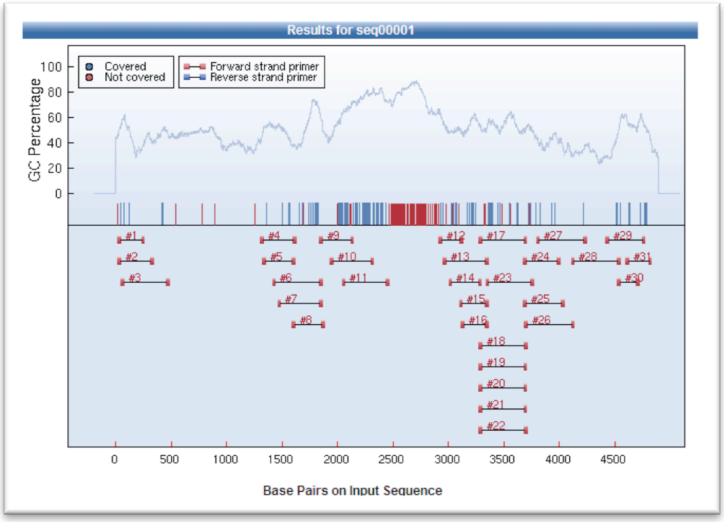
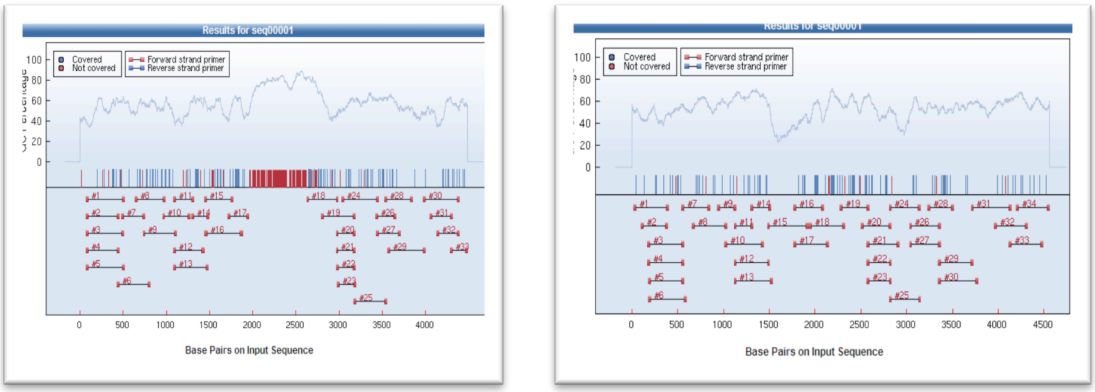


Figure 3: Primer map of 5-kilobase region for *CLTCL1*. Note that this figure includes all possible primers chosen by Sequenom EpiDesigner, not restricted to primers that were ultimately chosen. The horizontal axis indicates the base pair location in the input sequence and the vertical axis indicates the GC percentage. The blue lines are representative of CpG sites that could be interrogated by the program while the red lines are representative of CpG sites that could not be interrogated.



(4A) ***(4B)***

Figure 4: (4A) Primer map of 5-kilobase region for *DGCR8* Island A (located in promoter region of gene). ***(4B)*** Primer map of 5-kilobase region for *DGCR8* Island B (not located in promoter region of gene). Note that this figure includes all possible primers chosen by Sequenom EpiDesigner, not restricted to primers that were ultimately picked. The horizontal axis indicates the base pair location in the input sequence and the vertical axis indicates the GC percentage. The blue lines are representative of CpG sites that could be interrogated by the program while the red lines are representative of CpG sites that could not be interrogated.

Eight primers were chosen for *CLTCL1* that thoroughly spanned the 5-kilobase region (***Table 1)***. Ten primers each was chosen for *DGCR8* island A and island B for a total of twenty primers for the *DGCR8* gene ***(Tables 2 and 3)***.

Name	Left Sequence	Right Sequence	Amp
C1	**GGTTTTATTATGTTGGTTAGGTTGG**	**AACCTACAATCTAACTTTTATCCCCA**	**324**
C2	**AGTGATTTTGGTAGGATTTTTTGGA**	**CCCCCTACTTTAAACAAAAACTCAA**	**446**
C3	**TGAGTTTTTGTTTAAAGTAGGGGGT**	**CTACAAACAAAACAACCCAAAAACC**	**307**
C4	**GGTTGAGTTTAGGGTTATTTTTTATTT**	**CCCATACATCCATACATTCTAAAACA**	**409**
C5	**AGTGGTTGTGTTATTTGAGTTTGTG**	**TAAATACCAAAAACTTCCCTTTCCT**	**435**
C6	**ATTTTGAATAAGATAAGGATGGTGT**	**AAAACTTAATCCCCACTTAAAATCC**	**449**
C7	**GGGATTTTAATATTTTTATTATTTGTAGG**	**AACTCACACCTATAATCCCAACACT**	**357**
C8	**GATTGGTGTTGGTTTAGTTTAGTGG**	**ATCTCAAAAATATTAAAAAACTCATCTTCA**	**335**

Table 1: Summary of chosen *CLTCL1* primers to span the 5 kb region. Name indicates our labeled name of the primer. L-Seq and R-Seq indicate the left and right primer sequences respectively. Amp indicates the size of the target amplicon in base pairs

Name	Left Sequence	Right Sequence	Amp
DA1	**TATTTTGGGGGAGAAATTTTGTAAT**	**CAACCACTACACTCCAACATAAACA**	**440**
DA2	**GGTTTTATTATGTTGGTTAGGTTGG**	**ATCACCTAAACTCAAAAATTCCACA**	**349**
DA3	**TATTATTGGGGTGGTTATTTAGTGG**	**AAAATCACCTAAATCCAAAAATTCC**	**234**
DA4	**GGGTAGATTAATTGAGGTTAGGAGTT**	**TCAAAAAAATTACTTAAACCCAAAA**	**327**
DA5	**GAAAGGTGTGGATAGGGTTAGTGTT**	**TTTAAACCACTCCCACAACTAAAAA**	**392**
DA6	**GGTAGGAGTTTGATTTGTTTTTTTG**	**ACTTAAACAACACCCTTCCTCCTAA**	**426**
DA7	**TTTTGTTGGTAAGGTAGGGTTTGTA**	**AAAAAAAATAACCAAAAACACCAAA**	**329**
DA8	**AATATTTTGGTTTAGTTTATTTGGTG**	**CCTACAATCCCAACTACTTAAAAAA**	**267**
DA9	**TTTTTTAAGTAGTTGGGATTGTAGG**	**AAAAAAATCACTCTATAACCAAACACAA**	**200**
DA10	**TGGAAATTTTGTATTAGTAAAGGGGAT**	**CAACCATCTAAAAAACCTATCCAAAC**	**568**

Table 2: Summary of chosen *DGCR8* island A primers to span the 5 kb region. Island A is in the promoter region of *DGCR8*. Name indicates our labeled name of the primer. L-Seq and R-Seq indicate the left and right primer sequences respectively. Amp indicates the length of target amplicon in base pairs.

Name	Left Sequence	Right Sequence	Amp
DB1	**TTGGTTGTTGTTTATAGTTTTTTGAG**	**TACAAACATCACACATTTAACAAAA**	**384**
DB2	**GTGTTTTTTTTGTTTTGATGTAGGG**	**AAAAACACCCAAATTCCAACATAC**	**308**
DB3	**TTTTGGTTTATTGTTAGGGTTTTTTT**	**AAAACCTAAATACCACCTACAACCC**	**209**
DB4	**TGTTATTTTGTGGGTTTAGGAGAGT**	**CAAACATACAATAAAAAACACCCTTT**	**222**
DB5	**TGGTTTTTAATTTGGTATTAGGGAA**	**ACAAAACCAAAATTTTACAAACCCT**	**327**
DB6	**GGTTTTGGTGGTGATGGATAGTT**	**CCCCACTAACAAAACTAAAAAACAA**	**319**
DB7	**GGATGAAGAGGTTTTGAATTTTTTT**	**ACCAACTAACACCAACACACCTAAT**	**344**
DB8	**TGTTTGTTTTTAGGAATGTTGTTGA**	**TAAAACACCCACTACTTCTAACCCA**	**281**
DB9	**TAGGTATTGTGTAGAGGAATGGGAG**	**AAAACCCTCATAAATACCTCCAAAC**	**439**
DB10	**ATAGGTGAGATTTTAGGTTGAGGGT**	**TCCAAACCAAAATAACTACTAACCAA**	**364**

Table 3: Summary of chosen *DGCR8* island B primers to span the 5-kilobase region. Island B is NOT in the promoter region of *DGCR8*. Name indicates our labeled name of the primer. L-Seq and R-Seq indicate the left and right primer sequences respectively. Amp indicates the length of target amplicon in base pairs.

Optimization of PCR input DNA

DNA extraction

Due to limited quantity of patient B-lymphocytes, optimization of the PCR protocol with the lowest possible working quantity of input DNA for polymerase chain reaction was essential. To do this, we extracted DNA (Qiagen DNeasy Blood & Tissue kit, Valencia, CA) from our previously cultured B-lymphocyte cell lines (Coriell Cell Repositories – Camden, NJ). These cell pellets were from cell lines of an apparently healthy individual and one from a DiGeorge Syndrome individual. Pellets were previously frozen down to 5 million cells/pellet. Double-stranded DNA yield was quantified using spectrophotometer (BioTek, Take3 – Winooski, VT)

Sodium bisulfite conversion

After quantification, samples are ready for sodium bisulfite conversion (Zymo EZ DNA Methylation-Gold kit). 1 μg of each DNA sample is diluted for a final volume of 20 uL. Next, 130 μL of CT Conversion Reagent is added to each DNA sample in a PCR tube. Samples are mixed and placed in a thermal cycler. Thermal cycler steps are as follows:

1. 98° C for 10 minutes
2. 64° C for 2.5 hours
3. 4° C storage up to 20 hours

Next, 600 uL of M-Binding Buffer is added to Zymo-Spin IC Column and placed in a provided collection tube. Samples are added to the spin column containing the M-Binding Buffer and mixed by inverting. Next, samples are centrifuged at 10,000 x g (full speed) for 30 seconds. Flow-through is discarded. Then, 100 μL of M-Wash Buffer is

added to each column and centrifuged at full speed for 30 seconds. Next, 200 µL of M-Desulphonation Buffer is added to each column. Samples are let to stand at room temperature (20-30°C) for 15-20 minutes. After incubation, samples are centrifuged at full speed for 30 seconds. Next, 200 µL of M-Wash Buffer is added to each column and centrifuged at full speed for 30 seconds. Another 200 µL of M-Wash Buffer is added and additional centrifuge is done for 30 seconds. Next, the columns are placed into a 1.5 ml microcentrifuge tube and 10 µL of M-Elution Buffer is directly loaded into the column matrix. Samples are then centrifuged for 30 seconds at full speed to elute DNA. DNA yield was quantified using spectrophotometer (BioTek, Take3 – Winooski, VT), with parameters set to single-stranded DNA

Polymerase chain reaction optimization

After quantification, bisulfite converted DNA (BS DNA) samples were amplified by polymerase chain reaction (PCR) using JumpStart *Taq* ReadyMix, (Sigma-Aldrich, St. Louis, MO) which contains all components needed excluding primers and input DNA. Primers used in this optimization are described in *Part I* of primer design step. Specifically, primer *DGCR8*-4 was used. Two groups of reactions were prepared: 1 for GM22296 (Apparently Healthy BS DNA) and 1 for GM17942A (DiGeorge Syndrome BS DNA). For each group, 6 reactions were prepped with DNA input as the variable for a total of 12 reactions. DNA input varied from 2 ng, 5 ng, 10 ng, 15 ng, 20 ng, and 25 ng for each type of DNA. For each of the 12 reactions the remaining reagents were constant. Primers were used at a concentration of 10uM. Thermal cycler was set to the following parameters:

1.	Initial denaturation	94.0° C	2 min
2.	45 cycles of …		
	a. Denaturation	94.0° C	30 sec
	b. Annealing	48.6° C	30 sec
	c. Extension	72.0° C	2 min
3.	Final extension	72.0° C	5 min
4.	Hold	4.0° C	

Gel Electrophoresis

Next, a 2 % agarose gel was made to run the performed PCR products. All gels in this study were run with a 100 base pair ladder, which indicates a brighter band at the 500 base pair size (Thermo Scientific GeneRuler 100 bp DNA Ladder – Waltham, MA). Gels were imaged using Bio-Rad gel imager (Philadelphia, PA).

1st round of PCR optimization

All 28 primers from both gene targets were tested via PCR using previously isolated and converted DNA from B-lymphocyte cell lines (Case – DiGeorge Syndrome individual and Control – Apparently Healthy individual) at both 2 ng and 5 ng DNA input for a total of 56 PCR reactions. ***Table 4*** displays the PCR setup for the 56 reactions. Annealing temperature was set at 51° C. Gel electrophoresis was performed to obtain results.

C1	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**C8**	**DA1**	**DA2**	**DA3**	**DA4**
C1	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**C8**	**DA1**	**DA2**	**DA3**	**DA4**
DA5	**DA6**	**DA7**	**DA8**	**DA9**	**DA10**	**DB1**	**DB2**	**DB3**	**DB4**	**DB5**	**DB6**

DA5	**DA6**	**DA7**	**DA8**	**DA9**	**DA10**	**DB1**	**DB2**	**DB3**	**DB4**	**DB5**	**DB6**
DB7	**DB8**	**DB9**	**DB 10**								
DB7	**DB8**	**DB9**	**DB 10**								

Table 4: PCR setup of 96-well plate. White rows represent reactions of input DNA of 2 ng. Gray rows represent reactions with input DNA of 5 ng. PCR reactions were performed at an annealing temperature of 51° C

Obtaining control DNA – 100% and 0 % methylated DNA

Control DNA samples of 100% methylated DNA and 0 % methylated DNA was obtained (Zymo Reseach – Human Methylated and Non-Methylated Set, Irvine, CA – given from Dr. Alicia Smith). Sodium bisulfite conversion protocol was performed on these control DNAs. Samples were then quantified using spectrophotometer (BioTek, Take3 – Winooski, VT).

2nd round of PCR optimization

We performed PCR on successful primers from 1st round of optimization. Successful primers from round 1 must have correct amplicon size band and little to no byproduct such as primer dimers or alternate product amplicons (8 successful primers). For this round, DNA for PCR was from control DNAs (100% and 0% methylated DNA) at 20 ng and 2 ng input for a total of 36 reactions. ***Tables 5 and 6*** display the PCR setup for the 36 reactions. Gel electrophoresis was performed to obtain result

DA7	**DB2**	**DB4**	**DB5**	**DB6**	**DB7**	**DB8**	**C3**
DA7	**DB2**	**DB4**	**DB5**	**DB6**	**DB7**	**DB8**	**C3**

Table 5: PCR setup of 96-well plate. White row represents 100 % methylated DNA input and pink row represent 0% methylated DNA input. All inputs at 20 ng of DNA. PCR reactions were performed at an annealing temperature of 51° C.

DA7	**DB2**	**DB4**	**DB5**	**DB6**	**DB7**	**DB8**	**C3**
DA7	**DB2**	**DB4**	**DB5**	**DB6**	**DB7**	**DB8**	**C3**

Table 6: PCR setup of 96-well plate. White row represents 100 % methylated DNA input and pink row represent 0% methylated DNA input. All inputs at 2 ng of DNA. PCR reactions were performed at an annealing temperature of 51° C.

Preparation of patient samples

We chose five 22q11 patients and 5 controls that were age-matched as described in ***Table 7***. Patient B-cell aliquots were pooled and DNA extraction was performed as described earlier. Samples were quantified using spectrophotometer; however DNA concentrations were quite dilute. Samples were then speed vacuumed for 90 minutes and resuspended in 20 uL of AE buffer (Qiagen, elution buffer) to yield a more concentrated solution. Sample concentrations were confirmed via spectrophotometer (BioTek, Take3 – Winooski, VT). For bisulfite conversion protocol, an optimal input of 800 ng of DNA was used. After conversion, samples were quantified via spectrophotometer (BioTek, Take3 – Winooski, VT) and were ready for 3rd round of PCR.

	Case	T	Case	T	Case	T	Case	T	Case	T
P#	1	2	3	4	5	6	7	8	9	10
Age	12	14	12	14	27	24	27	25	41	43

Table 7: 22q11.2 DS patient cases and controls, age-matched. Cases are indicated as "case" and controls as "T". Cases have the 22q11.2 deletion while controls do not have the deletion.

3rd round of PCR

Successful primers from round 2 were used for this round of PCR using patient bisulfite converted DNA. With an input of 2 ng per patient sample, and 10 patients total (5 cases 22q11 and 5 controls) we had 40 total PCR reactions. ***Table 8*** displays the PCR setup for the 40 reactions. Gel electrophoresis was performed to confirm a correct product.

#	1	2	3	4	5	6	7	8	9	10
P	*DB2*	*DB2*	*DB2*	*DB2*	*DB2*	*DB2*	*DB2*	*DB2*	*DB2*	*DB2*
P	*DB4*	*DB4*	*DB4*	*DB4*	*DB4*	*DB4*	*DB4*	*DB4*	*DB4*	*DB4*
P	*DB7*	*DB7*	*DB7*	*DB7*	*DB7*	*DB7*	*DB7*	*DB7*	*DB7*	*DB7*
P	*DB8*	*DB8*	*DB8*	*DB8*	*DB8*	*DB8*	*DB8*	*DB8*	*DB8*	*DB8*

Table 8: Gray and white area shown in italics represents the PCR setup of a 96-well plate. Red row labeled "#" represents patients numbers. Patients are in case and control pairs – (1 and 2, 3 and 4, 5 and 6, 7 and 8, 9 and 10). Red column labeled "P" represent primer pair used for corresponding row. PCR reactions were performed at an annealing temperature of 51° C.

Repeated PCRs

In order to ensure that null results were not caused by pipetting errors, all PCR reactions that showed no band in gel electrophoresis were repeated. Six PCRs were redone with appropriate patient/primer combination. Gel electrophoresis was performed to confirm a correct product.

PCR purification

PCR product purification (Qiagen QIAquick PCR Purification kit, Valencia, CA) was performed for all patient and primer combinations that displayed correct band on gel electrophoresis (40 reactions). Eluted DNA is quantified using spectrophotometer (BioTek, Take3 – Winooski, VT).

4th round of PCR Optimization – Temperature gradient

Due to limited resources, manpower and time, primers in gene *DGCR8* were chosen to be further optimized. In this 4th round of PCR optimization, a temperature gradient was setup for the annealing step of the PCR. The thermal cycler was setup to run 12 different annealing temperatures, one temperature for each column of a 96-well plate. The temperature gradient ranged from 46° C to 66° C, with a 1.5 C interval between each column. The remaining 16 *DGCR8* primers were prepared for PCR with control DNA

inputs of both 100 % methylated and 0% methylated DNA. Therefore, each of the 16 primer pairs was run at 12 different temperatures and 2 types of control DNA (100% and 0%) yielding 384 PCR reactions. PCR protocol was almost identical as before, using a 2 ng amount of input DNAs and only differing in that forward and reverse primers were aliquoted into one tube, mixed and frozen until use. ***Tables 9, 10, 11, and 12*** display the PCR setups for the 384 temperature gradient PCR reactions. Gel electrophoresis was performed to obtain results.

°C	48	49.6	51.3	52.9	54.5	56.2	57.8	59.5	61.1	62.7	64.4	66
P	*DA1*	*DA1*	*DA1*	*DA1*	*DA1*	*DA1*	*DA1*	*DA1*	*DA1*	*DA1*	*DA1*	*DA1*
P	*DA1*	*DA1*	*DA1*	*DA1*	*DA1*	*DA1*	*DA1*	*DA1*	*DA1*	*DA1*	*DA1*	*DA1*
P	*DA2*	*DA2*	*DA2*	*DA2*	*DA2*	*DA2*	*DA2*	*DA2*	*DA2*	*DA2*	*DA2*	*DA2*
P	*DA2*	*DA2*	*DA2*	*DA2*	*DA2*	*DA2*	*DA2*	*DA2*	*DA2*	*DA2*	*DA2*	*DA2*
P	*DA3*	*DA3*	*DA3*	*DA3*	*DA3*	*DA3*	*DA3*	*DA3*	*DA3*	*DA3*	*DA3*	*DA3*
P	*DA3*	*DA3*	*DA3*	*DA3*	*DA3*	*DA3*	*DA3*	*DA3*	*DA3*	*DA3*	*DA3*	*DA3*
P	*DA4*	*DA4*	*DA4*	*DA4*	*DA4*	*DA4*	*DA4*	*DA4*	*DA4*	*DA4*	*DA4*	*DA4*
P	*DA4*	*DA4*	*DA4*	*DA4*	*DA4*	*DA4*	*DA4*	*DA4*	*DA4*	*DA4*	*DA4*	*DA4*

Table 9: Gray and white area shown in italics represents PCR setup of a 96-well plate. Gray rows represent PCR reactions with input of 100 % methylated DNA and white rows represent PCR reactions with input of 0% methylated DNA. Blue row labeled "°C" represents the temperature for each column. Blue row labeled "P" represents the primers used for each row.

°C	48	49.6	51.3	52.9	54.5	56.2	57.8	59.5	61.1	62.7	64.4	66
P	*DA5*	*DA5*	*DA5*	*DA5*	*DA5*	*DA5*	*DA5*	*DA5*	*DA5*	*DA5*	*DA5*	*DA5*
P	*DA5*	*DA5*	*DA5*	*DA5*	*DA5*	*DA5*	*DA5*	*DA5*	*DA5*	*DA5*	*DA5*	*DA5*
P	*DA6*	*DA6*	*DA6*	*DA6*	*DA6*	*DA6*	*DA6*	*DA6*	*DA6*	*DA6*	*DA6*	*DA6*
P	*DA6*	*DA6*	*DA6*	*DA6*	*DA6*	*DA6*	*DA6*	*DA6*	*DA6*	*DA6*	*DA6*	*DA6*
P	*DA7*	*DA7*	*DA7*	*DA7*	*DA7*	*DA7*	*DA7*	*DA7*	*DA7*	*DA7*	*DA7*	*DA7*
P	*DA7*	*DA7*	*DA7*	*DA7*	*DA7*	*DA7*	*DA7*	*DA7*	*DA7*	*DA7*	*DA7*	*DA7*
P	*DA8*	*DA8*	*DA8*	*DA8*	*DA8*	*DA8*	*DA8*	*DA8*	*DA8*	*DA8*	*DA8*	*DA8*
P	*DA8*	*DA8*	*DA8*	*DA8*	*DA8*	*DA8*	*DA8*	*DA8*	*DA8*	*DA8*	*DA8*	*DA8*

Table 10: Gray and white area shown in italics represents PCR setup of a 96-well plate. Gray rows represent PCR reactions with input of 100 % methylated DNA and white rows represent PCR reactions with input of 0% methylated DNA. Blue row labeled "°C" represents the temperature for each column. Blue row labeled "P" represents the primers used for each row.

°C	48	49.6	51.3	52.9	54.5	56.2	57.8	59.5	61.1	62.7	64.4	66
P	*DA9*	*DA9*	*DA9*	*DA9*	*DA9*	*DA9*	*DA9*	*DA9*	*DA9*	*DA9*	*DA9*	*DA9*
P	*DA9*	*DA9*	*DA9*	*DA9*	*DA9*	*DA9*	*DA9*	*DA9*	*DA9*	*DA9*	*DA9*	*DA9*
P	*DA10*	*DA10*	*DA10*	*DA10*	*DA10*	*DA10*	*DA10*	*DA10*	*DA10*	*DA10*	*DA10*	*DA10*
P	*DA10*	*DA10*	*DA10*	*DA10*	*DA10*	*DA10*	*DA10*	*DA10*	*DA10*	*DA10*	*DA10*	*DA10*
P	*DB1*	*DB1*	*DB1*	*DB1*	*DB1*	*DB1*	*DB1*	*DB1*	*DB1*	*DB1*	*DB1*	*DB1*
P	*DB1*	*DB1*	*DB1*	*DB1*	*DB1*	*DB1*	*DB1*	*DB1*	*DB1*	*DB1*	*DB1*	*DB1*
P	*DB3*	*DB3*	*DB3*	*DB3*	*DB3*	*DB3*	*DB3*	*DB3*	*DB3*	*DB3*	*DB3*	*DB3*
P	*DB3*	*DB3*	*DB3*	*DB3*	*DB3*	*DB3*	*DB3*	*DB3*	*DB3*	*DB3*	*DB3*	*DB3*

Table 11: Gray and white area shown in italics represents PCR setup of a 96-well plate. Gray rows represent PCR reactions with input of 100 % methylated DNA and white rows represent PCR reactions with input of 0% methylated DNA. Blue row labeled "°C" represents the temperature for each column. Blue row labeled "P" represents the primers used for each row.

°C	48	49.6	51.3	52.9	54.5	56.2	57.8	59.5	61.1	62.7	64.4	66
P	*DB5*	*DB5*	*DB5*	*DB5*	*DB5*	*DB5*	*DB5*	*DB5*	*DB5*	*DB5*	*DB5*	*DB5*
P	*DB5*	*DB5*	*DB5*	*DB5*	*DB5*	*DB5*	*DB5*	*DB5*	*DB5*	*DB5*	*DB5*	*DB5*
P	*DB6*	*DB6*	*DB6*	*DB6*	*DB6*	*DB6*	*DB6*	*DB6*	*DB6*	*DB6*	*DB6*	*DB6*
P	*DB6*	*DB6*	*DB6*	*DB6*	*DB6*	*DB6*	*DB6*	*DB6*	*DB6*	*DB6*	*DB6*	*DB6*
P	*DB9*	*DB9*	*DB9*	*DB9*	*DB9*	*DB9*	*DB9*	*DB9*	*DB9*	*DB9*	*DB9*	*DB9*
P	*DB9*	*DB9*	*DB9*	*DB9*	*DB9*	*DB9*	*DB9*	*DB9*	*DB9*	*DB9*	*DB9*	*DB9*
P	*DB10*	*DB10*	*DB10*	*DB10*	*DB10*	*DB10*	*DB10*	*DB10*	*DB10*	*DB10*	*DB10*	*DB10*
P	*DB10*	*DB10*	*DB10*	*DB10*	*DB10*	*DB10*	*DB10*	*DB10*	*DB10*	*DB10*	*DB10*	*DB10*

Table 12: Gray and white area shown in italics represents PCR setup of a 96-well plate. Gray rows represent PCR reactions with input of 100 % methylated DNA and white rows represent PCR reactions with input of 0% methylated DNA. Blue row labeled "°C"

represents the temperature for each column. Blue row labeled "P" represents the primers used for each row.

5th round of PCR

Optimal annealing temperatures and primers were chosen from gel results and PCRs were prepped using these parameters on patient BS DNA. This step yielded 60 PCR reactions, all at a 2 ng input of patient BS DNA. ***Tables 13, 14, 15, and 16*** display the 60 PCR reactions. Gel electrophoresis was performed to obtain results.

#	1	2	3	4	5	6	7	8	9	10
P	*DA5*	*DA5*	*DA5*	*DA5*	*DA5*	*DA5*	*DA5*	*DA5*	*DA5*	*DA5*
P	*DB5*	*DB5*	*DB5*	*DB5*	*DB5*	*DB5*	*DB5*	*DB5*	*DB5*	*DB5*

Table 13: Gray and white area shown in italics represents PCR setup. Red row labeled "#" represents patient numbers. Patients are in case and control pairs – (1 and 2, 3 and 4, 5 and 6, 7 and 8, 9 and 10). Red column labeled "P" represent primer pair used for corresponding row. PCR reactions were performed at annealing temperature of 53.8° C

#	1	2	3	4	5	6	7	8	9	10
P	*DA6*	*DA6*	*DA6*	*DA6*	*DA6*	*DA6*	*DA6*	*DA6*	*DA6*	*DA6*

Table 14: Gray and white area shown in italics represents PCR setup. Red row labeled "#" represents patient numbers. Patients are in case and control pairs – (1 and 2, 3 and 4, 5 and 6, 7 and 8, 9 and 10). Red column labeled "P" represent primer pair used for corresponding row. PCR reactions were performed at annealing temperature of 54.5° C

#	1	2	3	4	5	6	7	8	9	10
P	*DB6*	*DB6*	*DB6*	*DB6*	*DB6*	*DB6*	*DB6*	*DB6*	*DB6*	*DB6*

Table 15: Gray and white area shown in italics represents PCR setup. Red row labeled "#" represents patient numbers. Patients are in case and control pairs – (1 and 2, 3 and 4, 5 and 6, 7 and 8, 9 and 10). Red column labeled "P" represent primer pair used for corresponding row. PCR reactions were performed at annealing temperature of 54.5° C

#	1	2	3	4	5	6	7	8	9	10
P	*DB9*	*DB9*	*DB9*	*DB9*	*DB9*	*DB9*	*DB9*	*DB9*	*DB9*	*DB9*
P	*DB10*	*DB10*	*DB10*	*DB10*	*DB10*	*DB10*	*DB10*	*DB10*	*DB10*	*DB10*

Table 16: Gray and white area shown in italics represents PCR setup. Red row labeled "#" represents patient numbers. Patients are in case and control pairs – (1 and 2, 3 and 4, 5 and 6, 7 and 8, 9 and 10). Red column labeled "P" represent primer pair used for corresponding row. PCR reactions were performed at annealing temperature of 57.8° C

Product Purification – Gel or PCR

Depending on gel electrophoresis results, products were either PCR or gel purified to equal a total of 57 purification reactions. PCR purification was done for all patient/primer combinations that displayed little to no byproduct bands and only correct bands on gel results. Gel purification occurred on patient/primer combinations that displayed the correct bands, but also showed byproduct bands. Correct gel bands are excised from agarose gel using a clean, sharp razor and long-waved UV handheld lamp. The gel purification (Qiagen QIAquick Gel Extraction kit, Valencia, CA) was performed on spliced gel fragments Eluted DNA is quantified via spectrophotometer (BioTek, Take3 – Winooski, VT). DNA yields were sufficient for the next step of the protocol, sequencing.

Sequencing

Ninety-seven patient samples in 10 primer regions were prepped, purified and sent to Emory University's Genomics core lab for Illumina V3 sequencing. We are waiting on sequencing data results.

Generally, this type of sequencing incorporates a library preparation, which involves shearing our DNA products and the ligation of adapters to the ends of the fragmented DNA. Sequencing occurs when a glass flow cell that has the same ligation adapters attached to its surface, is introduced to the library preparation DNA, allowing for hybridization of DNA fragments to the surface.[35] [36] The amplification occurs through bridge amplification where DNA fragments bend over and encounter a complimentary primer pair. The polymerase can then do multiple copies in one place resulting in cluster

amplifications.[35 36] This technology involves pair-end sequencing, which allows for sequencing in both the forward direction and the reverse direction. The reactions supply all 4 nucleotides, which are designed with a distinct fluorescence, reporting at a specific wavelength via imaging.[35 36]

At the time of sequence analysis, we hope to separate the sequences into their respective primer regions that we interrogated. Next, we propose to investigate the methylation percentages per patient at each CpG site, and then investigate the methylation at each primer region.

Results

Prioritization of target genes

Target genes from the 22q11 deletion region chosen were *CLTCL1* and *DGCR8*. Appropriate primer regions in these regions were then optimized in our study for efficient future methylation analysis.

Cell lines for optimization

In order to optimize primer regions in *CLTCL1* and *DGCR8*, we used cell lines (Coriell Cell Repository) from a DiGeorge Syndrome individual and an apparently healthy individual. ***Table 17*** displays the characteristics of the individuals and their cell lines.

Cell ID	GM17942	GM22296
Disease	DiGeorge Syndrome	Apparently healthy
Cell Type	B-Lymphocyte	B-Lymphocyte
Tissue Type	Human blood	Human blood
Transformant	Epstein-Barr Virus	Epstein-Barr Virus
Age/Sex	6 years/Male	48 years/Male
Race	Hispanic/Latino	Caucasian
Description	Clinically affected	Clinically normal

Table 17**:** Cell lines cultured for optimization of bisulfite conversion. Cell lines were obtained from Coriell Cell Repository (Camden, NJ) and used to optimize the bisulfite conversion protocol before use on more limited sources of patient DNA.

Primer Design

Using Sequenom Epidesigner, twenty-eight primer pairs were chosen from CpG regions in *CLTCL1* and *DGCR8*. In *CLTCL1*, one region was picked, which included the CpG island and a 2 kilobase-pair upstream and downstream CpG island shore.

Optimization of PCR input DNA

Due to our low patient B-lymphocyte counts, DNA input quantity for optimization of PCR was required at the initial steps of the study. Patient cell counts ranged from 10,000 to 30,000. Assuming a consistent DNA yield from each B-lymphocyte, more cells would allow for greater DNA extraction. Optimization of input DNA allowed us to find the lowest DNA amount that would still yield successful PCR results. ***Figure 5*** displays the PCR results for DNA inputs ranging from 2ng to 25ng in case and control DNA from cell lines. The gel displays a band between 400 and 500 base pairs. The gel in ***Figure 5*** and all gels thereafter display a 100 base pair (Thermo Scientific GeneRuler 100 bp DNA Ladder – Waltham, MA) ladder with a brighter band at the 500 base pair size, as indicated by the blue arrow. Since the target amplicon size is 436 base pairs we concluded that the primer pairs are amplifying the expected product. Furthermore, we saw a visible, but faint band at 2ng and 5ng. By using a 2 ng and 5 ng for the next round of optimization, we were able to determine the input to achieve optimal results. Using the minimum amount of DNA input possible that still achieves results ensures that we can interrogate as many primer regions as possible given our low patient DNA yields.

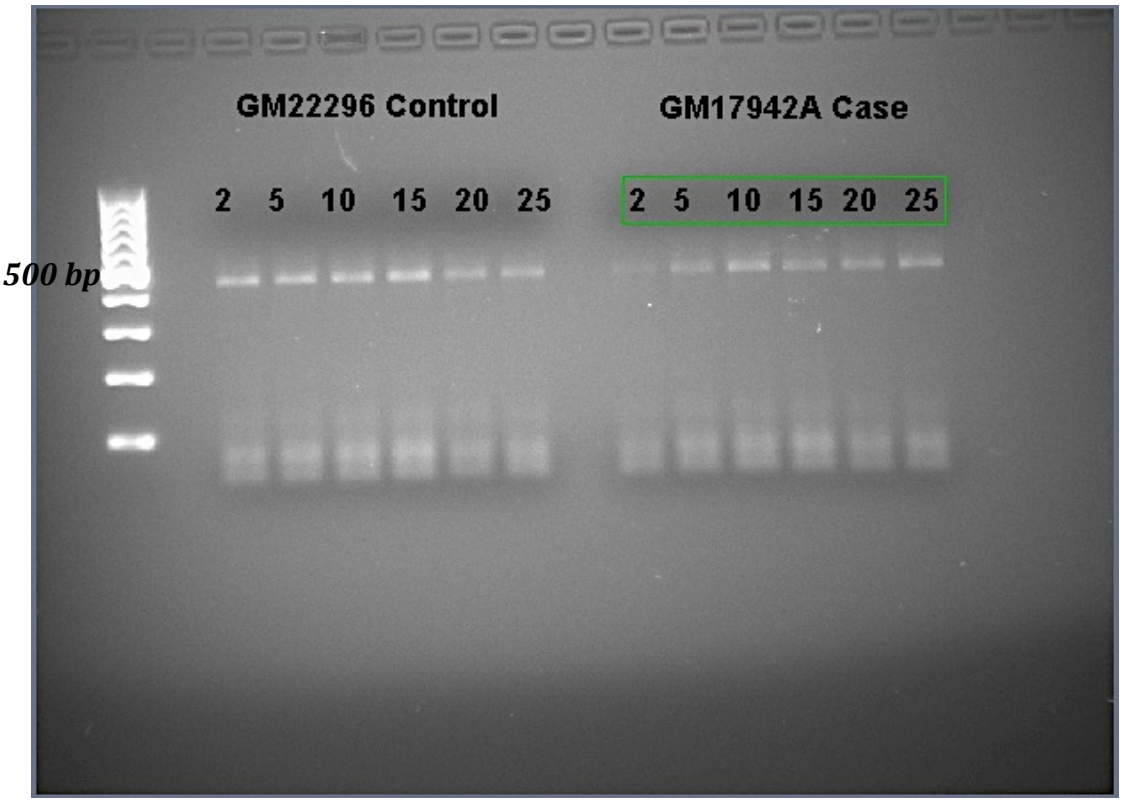
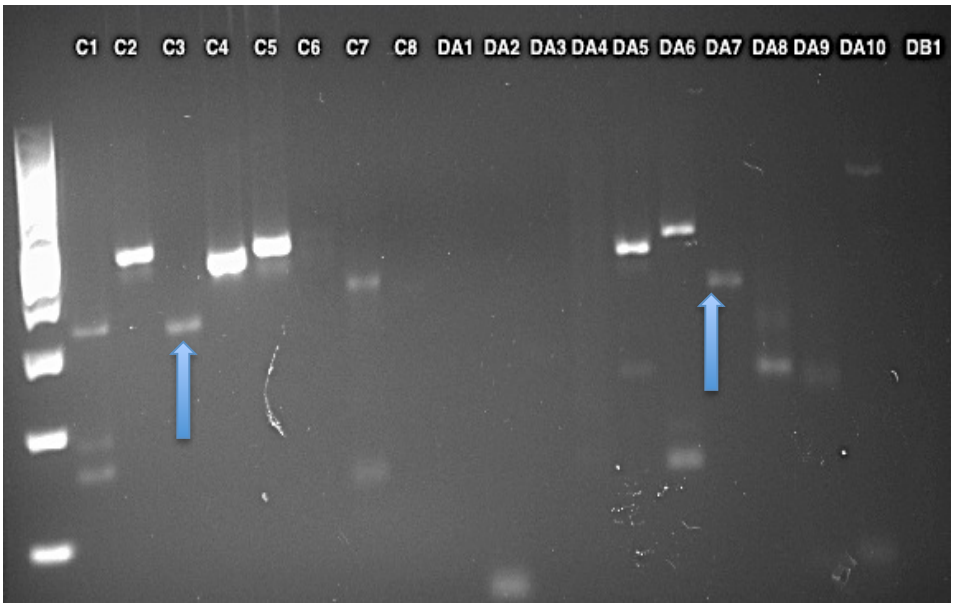


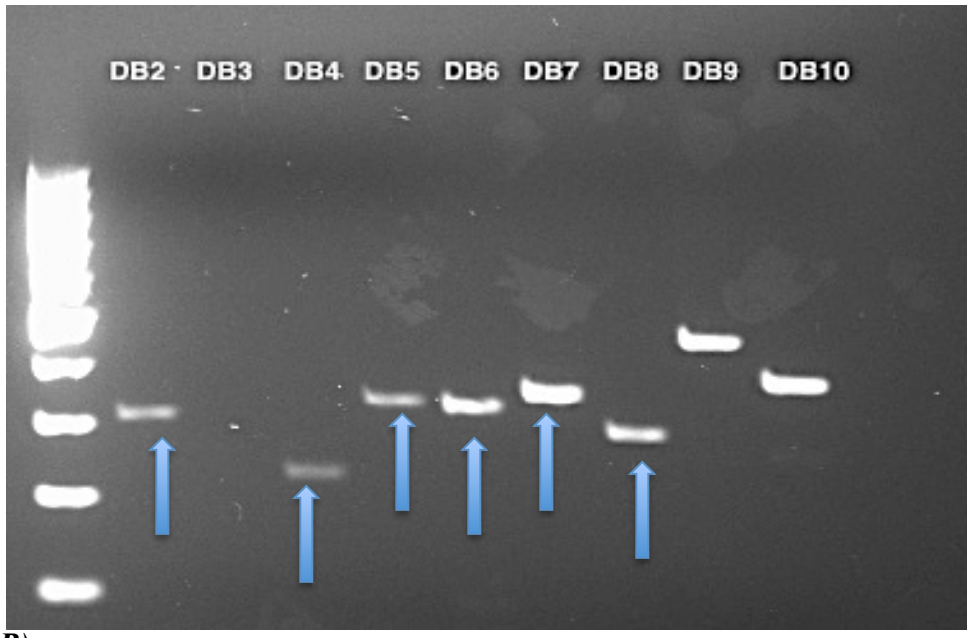
Figure 5*:* Gel electrophoresis imaging shows bands at all DNA inputs at about 436 bp, ranging from 2 ng to 25 ng. We proceeded forward with the next round of optimization with a 2 ng and 5 ng input.

1st round of PCR optimization

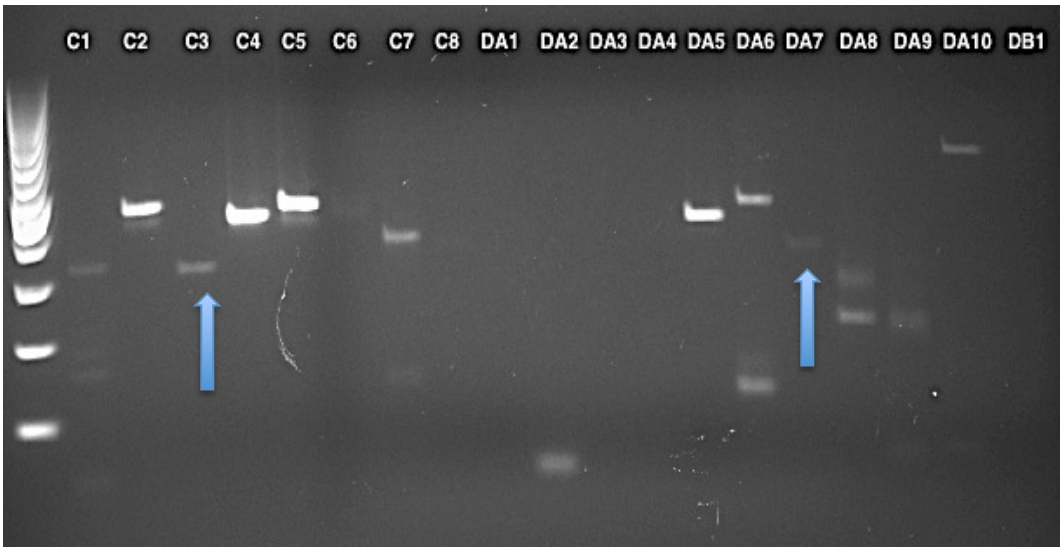
For the 1st round of PCR optimization, a PCR annealing temperature of 51°C was chosen. Since the majority of primer pairs had similar melting temperatures this temperature was chosen for being at or just below the primers lowest Tms. All 28-primer pairs that were chosen from Sequenom EpiDesigner were interrogated in this round of PCR. We input 2ng and 5ng for each primer pair to again ensure that our previous DNA input optimization results would replicate. As ***Figure 6*** shows, the gel electrophoresis results display single bands at correct corresponding amplicon sizes for 8 primer pairs. These 8 primer pairs moved on to the 2nd round of PCR optimization while all other primer pairs were held aside for later optimization.


C1 C2 C3 C4 C5 C6 C7 C8 DA1 DA2 DA3 DA4 DA5 DA6 DA7 DA8 DA9 DA10 DB1

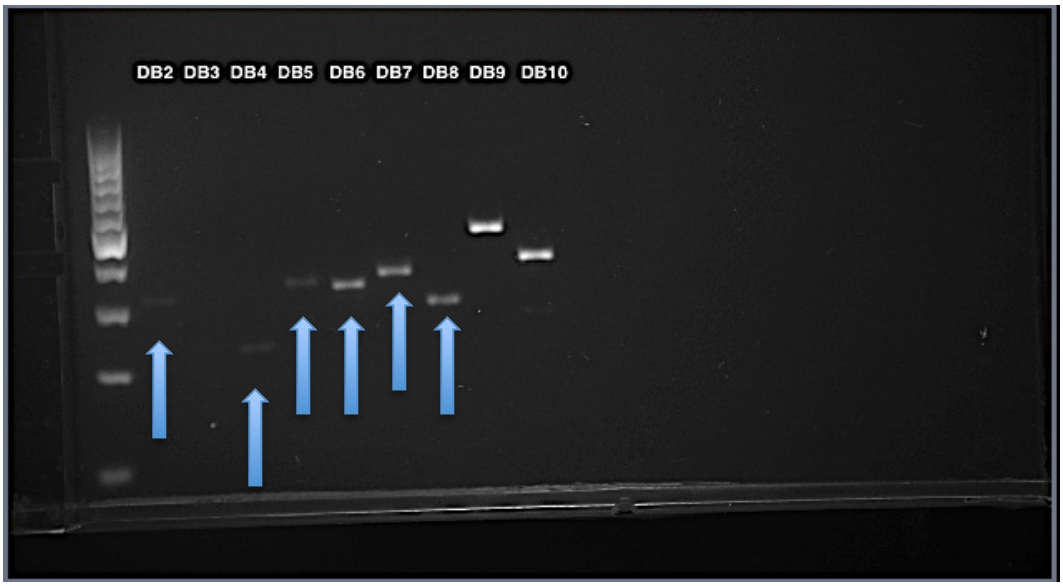
(6A)


DB2 DB3 DB4 DB5 DB6 DB7 DB8 DB9 DB10

(6B)



(6C)



(6D)

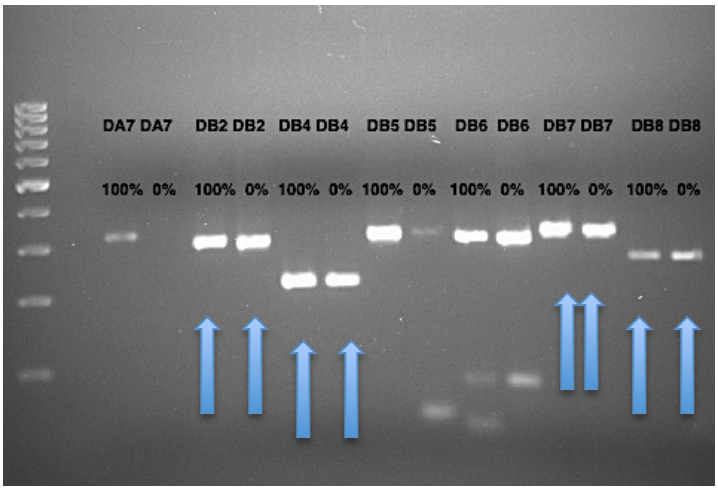
Figure 6: Single bands at the correct corresponding amplicon sizes are seen at the following regions: C3 (expected amplicon 307 bp), DA7 (expected amplicon 329 bp), DB2 (expected amplicon 308 bp), DB4 (expected amplicon 222 bp), DB5 (expected amplicon 327 bp), DB6 (expected amplicon 319 bp), DB7 (expected amplicon 344 bp), and DB8 (expected amplicon 281 bp). All PCRs were performed at an annealing temperature of 51°C. ***(6A and 6B)*** Displays gels for PCR DNA input of 2 ng. ***(6C and 6D)*** Displays gels for PCR DNA input of 5 ng.

2nd Round of PCR

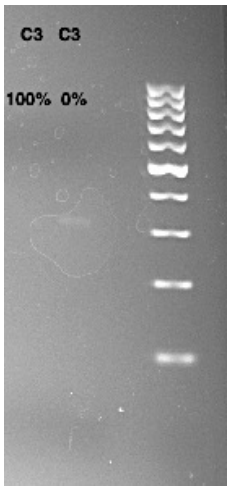
Control DNA inputs of methylated (100%) and non-methylated (0%) human DNA were used in at this step of the primer optimization workflow. The non-methylated human DNA is purified from HCT116 double-knockout cells of both types DNA methyltransferases (DNMTs); This control DNA, therefore, has a methylation level of less than 5 percent and can be used as a negative control. The methylated human DNA, a positive control, was also purified from HCT116 double-knockout cells that have been enzymatically methylated at all cytosine residues in the CpG context, by CpG methylase (Zymo Research - Irvine, CA). Bands at both 100% and 0% methylated DNA ensures that strand bias is not occurring. Strand bias is the term to describe the preferential amplification of unmethylated DNA over methylated DNA, which can occur because unmethylated DNA is less complex and therefore easier for DNA polymerase to replicate. PCR optimization with positive and negative control DNA allowed us to ensure our methylation interrogation was valid for the lowest and highest theoretical percentages of methylation variation. The human methylation variation can then be consistently investigated with our protocol.

The two control DNAs were used for each successful primer pair from the 1st round of PCR optimization. We performed PCR with input DNA of 2ng and 20ng. We used 20ng to verify no byproduct appeared from nonspecific amplification. The PCR annealing temperature remained at 51°C. As seen in ***Figure 7, f***our primer pairs, *DGCR8*-B2, *DGCR8*-B4, *DGCR8*-B7, *DGCR8*-B8 showed equal bands in both intensity and amplicon size and proceeded to the next step of optimization. Primers that were successful showed

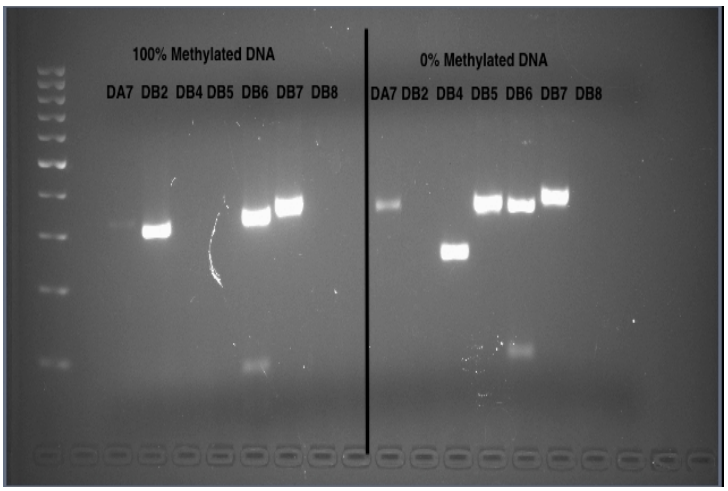
ability to optimally amplify the correct amplicon at both the lowest and highest methylation levels, ensuring that human methylation variation would be covered and the likelihood of strand bias would be low.



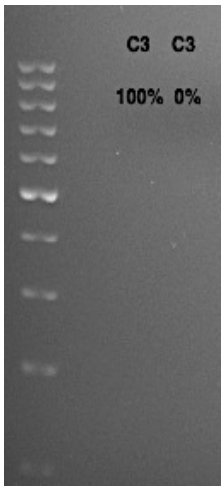
(7A)



(7B)



(7C)

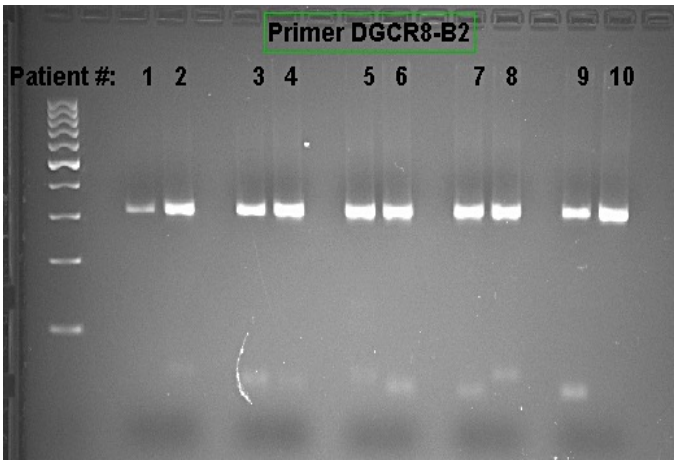


(7D)

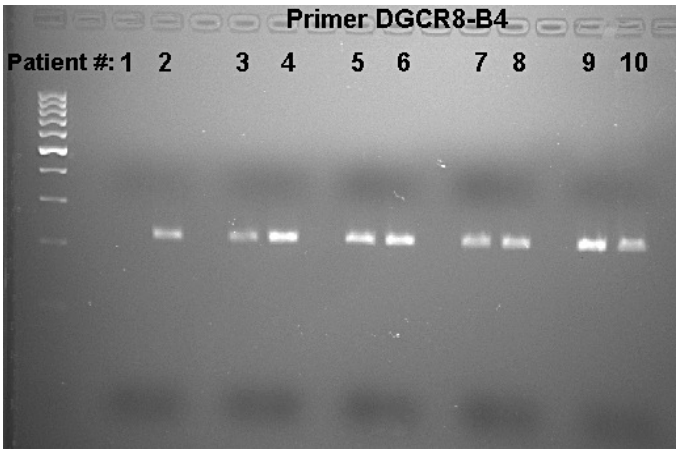
Figure 7 (A-D): In *7A,* equal bands of intensity and amplicon size are seen for DB2 (expected amplicon 308 bp), DB4 (expected amplicon 222 bp), DB7 (expected amplicon 344 bp), DB8 (expected amplicon 281 bp) in as seen by the blue arrows in. ***(7B and 7D)*** No products were seen in gels for the CLTCL1 primer region of C3 All PCRs were performed at annealing temperature of 51°C. ***(7A and 7B)*** Displays gels for PCR using DNA inputs of 100% and 0% control DNA at 2ng. ***(7C and 7D)*** Displays gels for PCR using DNA inputs of 100% and 0% control DNAs at 20 ng, which do not show same consistency as the 2ng gels, probably due to technical errors.

3rd Round of PCR

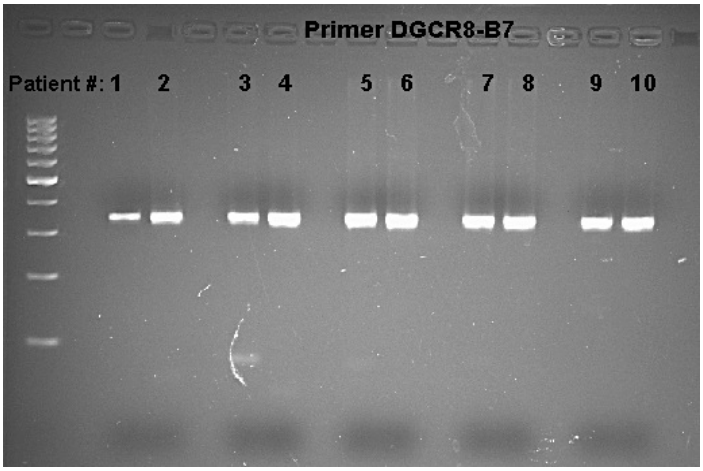
The 3rd round of PCR involved using our optimized PCR parameters on extracted and bisulfite-converted patient DNA (***Figure 8***). Patient 1 did not amplify primer region *DGCR8*-B4 and *DGCR8*-B8 and Patient 6 did not amplify at primer region *DGCR8*-B8. This suggested one of two rationales: First, that the patients did not contain complimentary primer regions, either by mutation or deletion, so that the primers could not bind to produce the correct amplicon size. Second, pipetting errors during the PCR preparation occurred, such as errors in DNA input, primer input or *Taq* ReadyMix input.



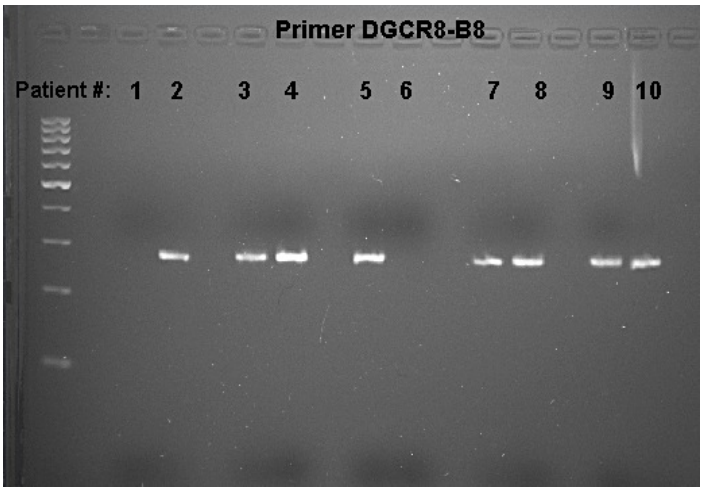
(8A)



(8B)



(8C)



(8D)

Figure 8 (A-D): Gel results for patient PCRs of successful primers from 2nd round of optimization. PCRs involved patient DNA input of 2ng and an annealing temperature of 51°C. ***(8A)*** Displays gels for patient PCRs of primer *DGCR8*-B2 with an expected amplicon of 308 bp ***(8B)*** Displays gels for patient PCRs of primer *DGCR8*-B4 with an expected amplicon of 222 bp. Patient 1 does not display band. ***(8C)*** Displays gels for patient PCRs of primer *DGCR8*-B7 with an expected amplicon of 344 bp. ***(8D)*** Displays gels for patient PCRs of primer *DGCR8*-B8 with an expected amplicon of 281 bp. Patient 1 and 6 do not display bands.

Repeated PCRs

DNA input of 2ng for the patient PCRs from the 3rd round was at such a low quantity, it was possible that DNA was not mixed into the PCR preparation. To control for the possibility of either patient DNA variation (mutation or deletion) or pipetting error, PCRs were repeated on the case/control group of the patients where the product was absent. ***Figure 10*** shows these PCRs had the correct amplicon size at all re-investigated primer regions for both case/control patients.

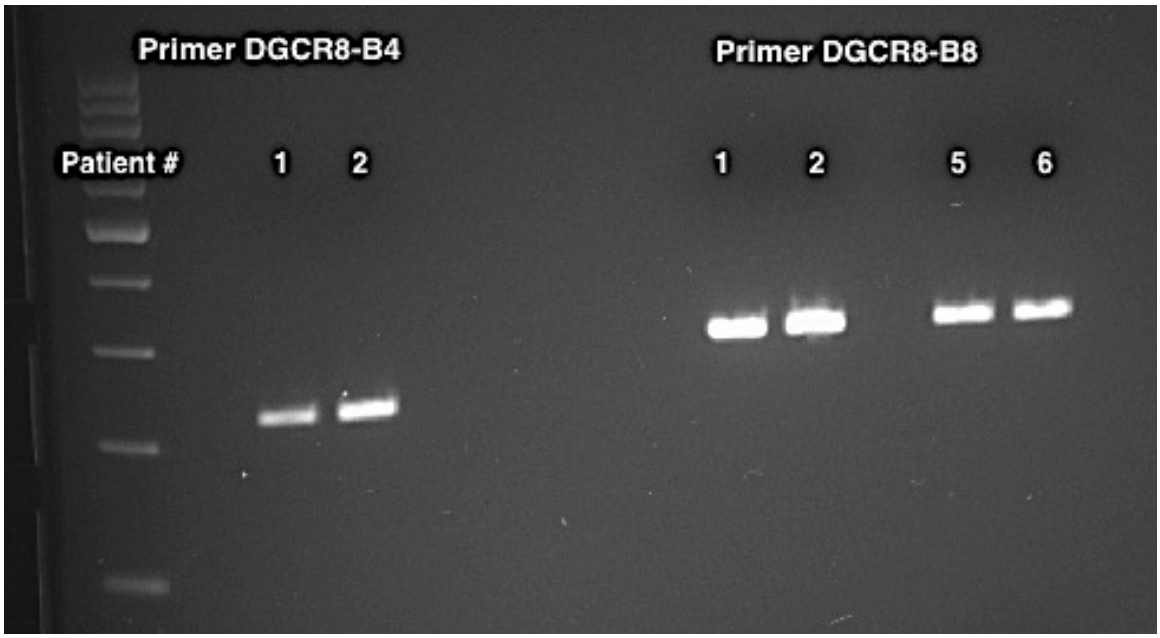
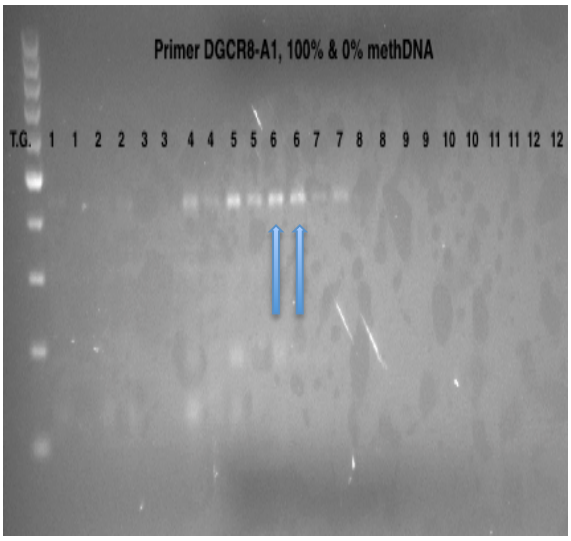


Figure 9: Displays gels for patient case and control pairs from primer regions that did not show bands from the 3rd round of PCR. Expected amplicon for *DGCR8*-B4 is 222 bp and for *DGCR8*-B8 is 281 bp. PCRs were redone on *DGCR8*- B4 for patient case and control pair 1 and 2. PCRs were also redone on *DGCR8*-B8 on patient case and control pair 1 and 2, 5 and 6. PCRs involved patient DNA input of 2 nanograms and an annealing temperature of 51°C.

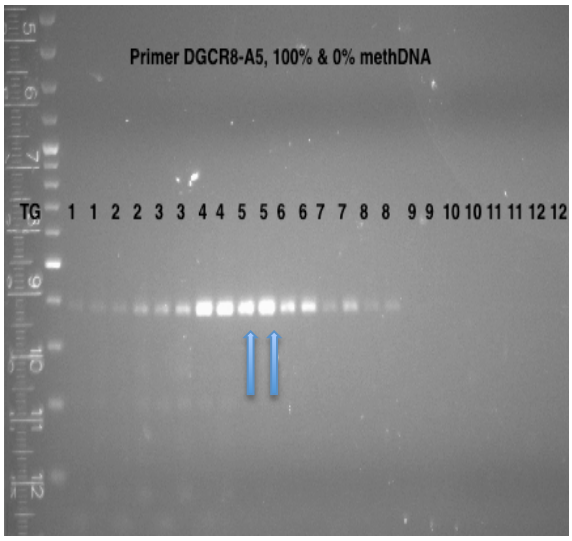
4th Round of PCR Optimization – Temperature Gradient

The above experiments yielded 4 primers, which were successfully optimized, all in island B of *DGCR8* gene. Due to limitations, further optimization steps were restricted to

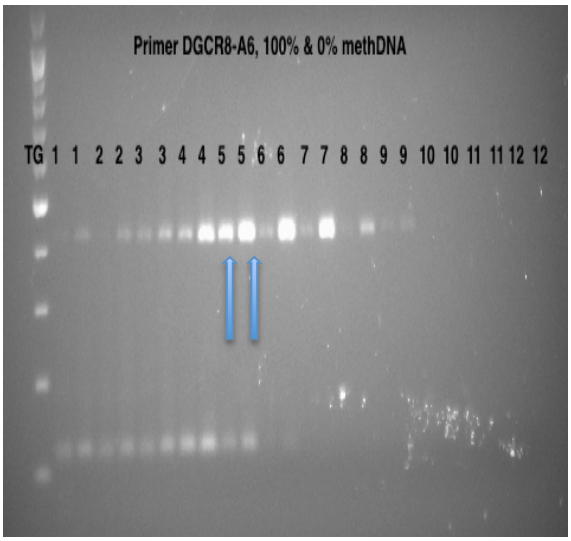
DGCR8. The remaining 16 primers, 10 from *DGCR8* island A and 6 from *DGCR8* island B, underwent PCR temperature gradient optimization. The 4th round of PCR optimization efficiently combined the control DNA (100% and 0% methylated DNA) step with the temperature gradient step. Hence our 96-well PCR plates had a row for 100% methylated DNA and a row for 0% methylated DNA for each primer pair. Each plate investigated 4 primer pairs, with 4 plates total. The temperature gradient was automatically set by the thermal cycler (Fisher Scientific - Eppendorf Mastercycler, Hamburg, Germany) with 12 different annealing temperatures set, one for each of the twelve columns of the 96-well plate. The annelaing temperature range of the gradient was 48°C to 66°C, giving a difference between each of the 12 rows of 1.63°C. The temperature gradient allowed for an annealing temperature interrogation that covered an 18-degree span, in one step. In the gel electrophoresis, successful primers showed equal bands in intensity and amplicon size at both 100% and 0% methylated DNA at one temperature in the 12 temperature gradient. In other words, a temperature that did not have two equal bands was considered not optimal and was discarded. In the temperatures that did show two equal bands, the temperatures with the brightest bands were picked. Optimal temperatures were found for *DGCR8*-A1 (expected amplicon 440 bp), *DGCR8*-A5 (expected amplicon 392 bp), *DGCR8*-A6 (expected amplicon size 426 bp), *DGCR8*-B5 (expected amplicon 327 bp, *DGCR8*-B6 (expected amplicon 319 bp), *DGCR8*-B9 (expected amplicon 439 bp), and *DGCR8*-B10 (expected amplicon 364 bp), as indicated by the blue arrows ***(Figure 11)***. The interval numbers displayed on the gels correspond to the TGs in ***Table 18***.



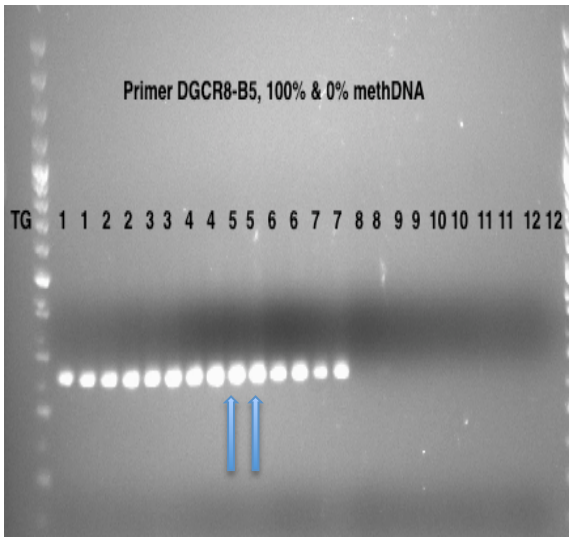
(10A)



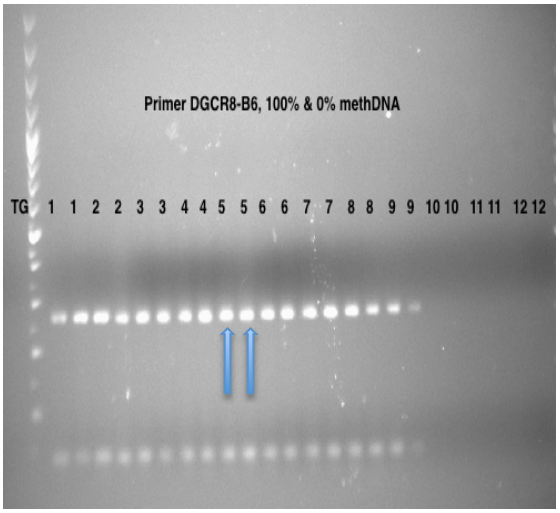
(10B)



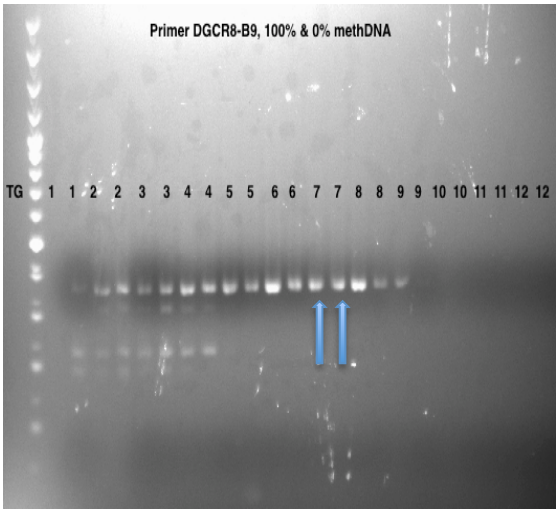
(10C)



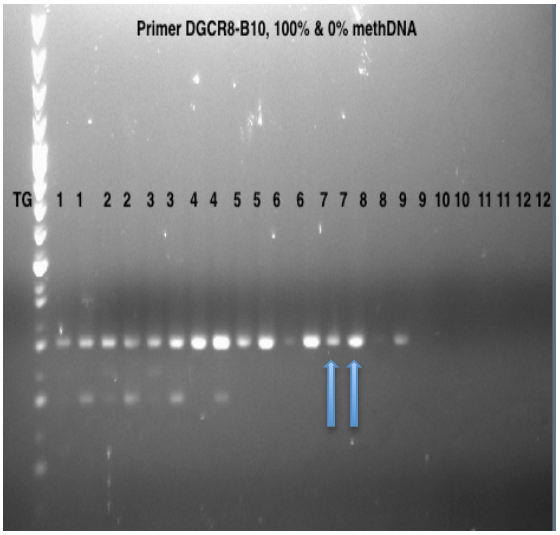
(10D)



(10E)



(10F)



(10G)

Figure 10 (A-G): Displays gels from temperature gradient PCRs from the remaining 16 *DGCR8* primer regions. Temperatures on the gradient (TG) correspond to the respective TGs on the table below (***Table 18)***. Optimal temperatures are chosen on the basis of consistent bands in both intensity and respective amplicon size for both 100% and 0% methylated DNA ***(10A)*** Optimal temperature is at the 6th interval at 56.2°C. ***(10B-10E)*** Optimal temperature is at the 5th interval at 54.5°C. ***(10F, 10G)*** The optimal temperature is at the 7th interval at 57.8°C.

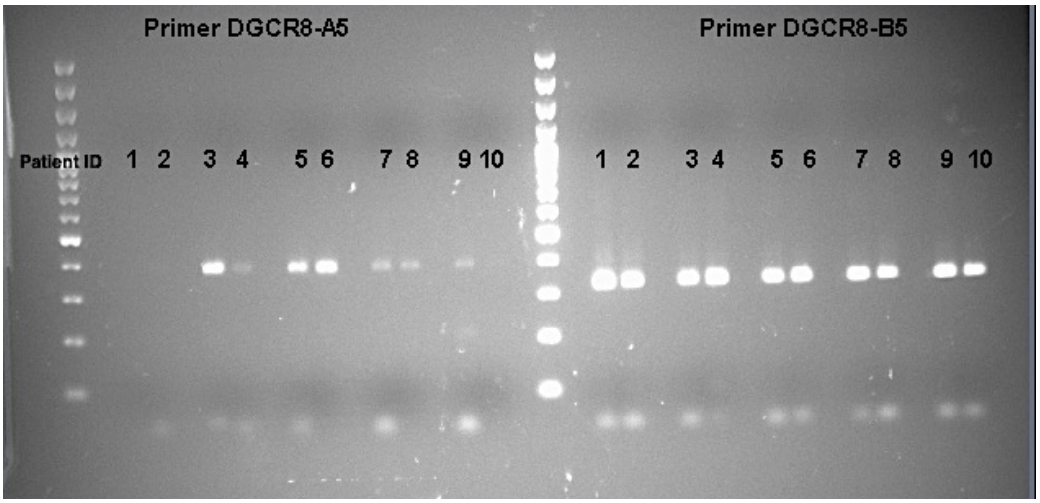
TG	1	2	3	4	5	6	7	8	9	10	11	12
°C	48	49.6	51.3	52.9	54.5	56.2	57.8	59.5	61.1	62.7	64.4	66

Table 18: TG corresponds to the temperatures for each column of the PCR temperature gradient optimization. Each column of the 96-well PCR plate corresponds to a specific temperature as specified. On the last column of this table TG is 12 and corresponding °C is 66

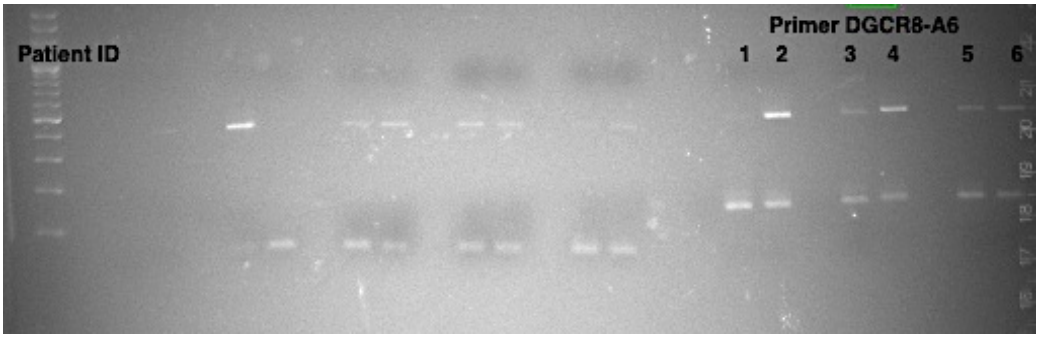
5th round of PCR

Patient PCRs were run on all of successful primer pairs from the 4th round, except for *DGCR8*-A1. The faint, yet consistent bands at optimal temperature of *DGCR8*-A1 indicated that the PCR step may require more DNA input. However, due to the limited patient DNA available, we did not pursue this primer region and decided to optimize the other 6 regions that did work.

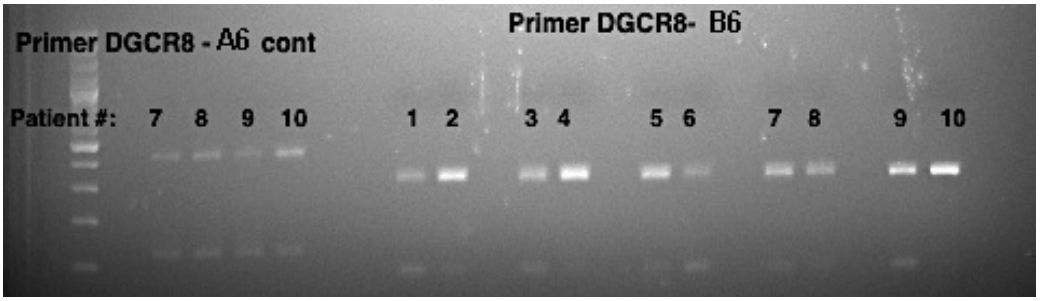
All patient PCRs were performed at the corresponding optimal annealing temperatures (***Figure 11***) and purified, either directly from PCR product or from extracted gel fragments. Gel purification was chosen for primer regions that displayed at least moderate primer dimers or any non-specific products. This included primers *DGCR8*-A5, *DGCR8*-A6, *DGCR8*-B5, *DGCR8*-B6, *DGCR8*-B9. The gel for region *DGCR8*-A5 showed several patients with faint or non-existent bands. Patients 3, 4, 5, 6, 7, 8, and 9 were gel purified, ensuring that only the correct amplicon would be purified in the sequence sample (***Figure 11A***). Primer *DGCR8*-B10 showed clean gels with very faint bands (***Figure 11E***), and proceed with direct PCR purification. Both PCR and gel purification protocols purified out products under 100 base pairs, ensuring that purified product did not contain primer dimers. In total, 10 primer regions had successful optimizations. DNA purifications were sent to sequencing.


Primer DGCR8-A5
Primer DGCR8-B5
Patient ID
1 2 3 4 5 6 7 8 9 10
1 2 3 4 5 6 7 8 9 10

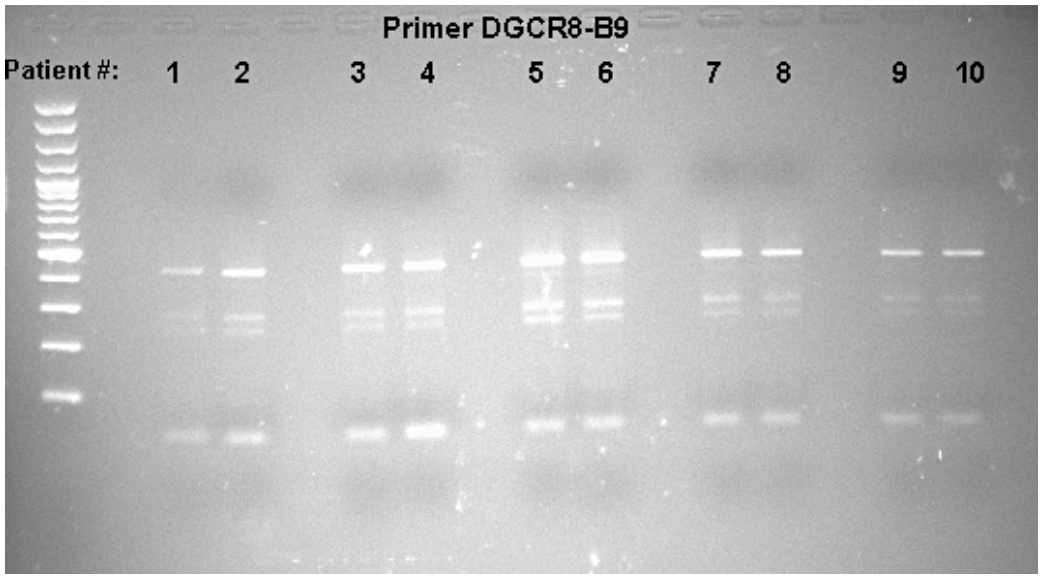
(11A)


Primer DGCR8-A6
Patient ID
1 2 3 4 5 6

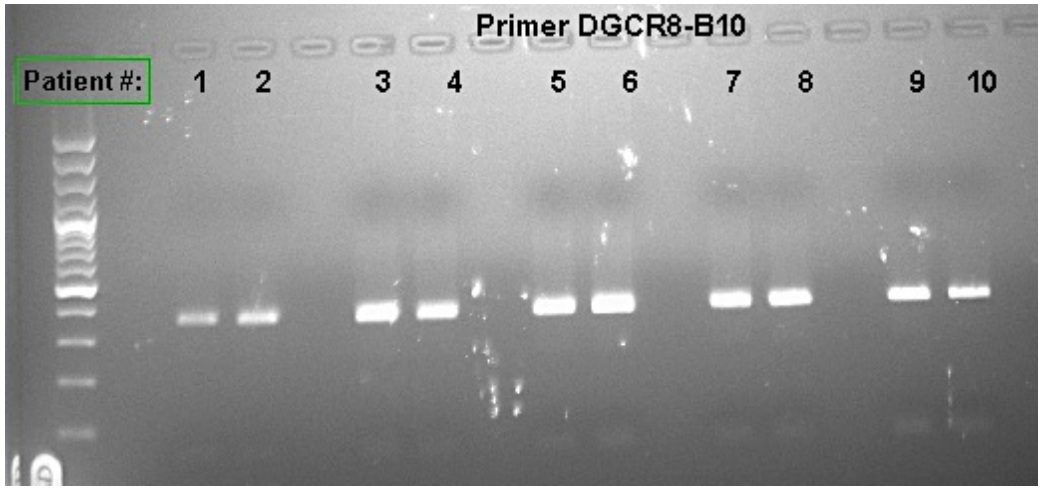
(11B)


Primer DGCR8 - A6 cont
Primer DGCR8- B6
Patient #:
7 8 9 10 1 2 3 4 5 6 7 8 9 10

(11C)



(11D)



(11E)

Figure 11: Gel results for patient PCRs of successful primers from 4th round of optimization. PCRs involved patient DNA input of 2ng and an annealing temperature corresponding to the optimal temperature found in the TG. ***(11A)*** Displays gels for patient PCRs of primer *DGCR8*-A5 (expected amplicon size 392 bp) and *DGCR8*-B5 (expected amplicon 327 bp) at an annealing temperature of 54.5°C. Patient 1, 2 and 10 do not display bands for primer region *DGCR8*-A5. ***(11B)*** Displays gels for part of the patient PCRs of primer *DGCR8*-A6, (expected amplicon 426 bp). Disregard non-labeled lanes on gel. *DGCR8*-A6 continues on ***(11C)*** which displays the rest of the patient PCRs of *DGCR8*-A6 and all patient PCRS of *DGCR8*-B6 (expected amplicon 319 bp). The annealing temperature for *DGCR8*-A6 was set to 54.5°C, and 56.2°C for *DGCR8*-B6. ***(11D)*** Displays gels for patient PCRs of primer *DGCR8*-B9 (expected amplicon 439 bp)

set at an annealing temperature of 57.8 C. ***(11E)*** Displays gels for patient PCRs of primer *DGCR8*-B10 (expected amplicon 364 bp) set at an annealing temperature of 57.8°C.

Discussion

The main accomplishment of this thesis is the optimization of 10 primer pairs that can be used to query differential methylation in the DGCR8 gene, which is located in the DiGeorge region of chromosome 22q11.2. These primer pairs collectively encompass approximately 3.4 kilobase pairs of CpG islands and shores of this gene. To our knowledge, the current literature is lacking any comparable coverage of this gene region that would allow subsequent sequencing of bisulfite converted DNA. Moreover, this is the largest region for such coverage of any gene in the typically deleted region of 22q11.2. Additionally, these studies demonstrated that these optimized PCR primer pairs could amplify bisulfite converted DNA from patients with D22q11 DS, in addition to control subjects without a known chromosomal deletion. Amplicons to query differential methylation in this 3.4kb region are now undergoing sequencing.

Insight from primer design

In comparison to hand-designing primers, using the primer-design software, Sequenom EpiDesigner proved to be more efficient and produced more successful primers. The software is a tool for primer design of bisulfite converted genomic DNA. Sequenom EpiDesigner allowed us to pick primer pairs within a set of parameters and a certain meting temperature (Tm) range of each other. The major limitations of our hand-picked primers were being able to find a variety of primers that fit our bisulfite design guidelines while also having an appropriate Tm range between the forward and reverse primers to successfully perform PCR. Hand-picked primer pairs had Tms at below 50°C and Tm differences of greater than 5°C. Sequenom EpiDesigner, however, not only recommended a number of primer pairs for each sequence range, but also allowed us to pick a majority

of our primer pairs with a Tm between 51°C - 56°C and within 5°C of each other, allowing for more successful amplification. However, the Sequenom EpiDesigner system is tailored for designing primers that are optimized for the company's downstream applications, and thus the choice of primers could have constraints that are not evident.

Optimization of *DGCR8* proved more difficult in island A region, with only 3 out of the 7 primer regions in the area showing any promise in the temperature gradient step. The remaining 8 regions either showed no bands or showed preferential amplification in one of the control DNAs (100% or 0% methylated DNA). The temperature gradient results for *DGCR8*-A2 showed no product amplification, but only amplification in what would seem to be primer dimers. The temperature gradient results for *DGCR8*-A3 and *DGCR8*-A4 showed no amplification. The temperature gradient results for *DGCR8*-A7, *DGCR8*-A9, and *DGCR8*-A10 showed preferential methylation bias with bands only at the 0% methylated DNA inputs. The temperature gradient results for *DGCR8*-A8 showed faint and inconsistent amplification, with no single temperature displayed equal bands at both control DNA inputs. Zero percent methylated DNA involves unmethylated cytosines at all sites (CpG and non-CpG sites) meaning that successful bisulfite treatment would theoretically convert all cytosine nucleotides in the genome to uracil nucleotides. The genome sequence, hence, has become less complex with an increased level of uracils (or thymines during PCR). Therefore during PCR, a possible explanation for preferential amplification of 0% methylated DNA may be due to a tendency to amplify less complex DNA sequences.

Implications of the findings

There are no comparable studies in the literature, though a paper was published just recently that examined a different gene in the 22q11-deleted region.[37] This paper examined CpG methylation in DGCR6/DGCR6L in whole blood of patients with 22q11.DS (N=16 subjects) and compared percent methylation to that observed in whole blood from N=3 control subjects.[37] The study used a similar technique to the one used in this thesis, but was only able to query methylation with four primer pairs.[37] The CpG sites in these promoters had a level of methylation that was low (2-8%), and did not differ in these whole blood samples between cases and controls, though only three controls were assessed.[37]

With 10 amplicons able to be sequenced, our DNA methylation analysis could uncover a number of possibilities. Studies have shown promise in using DNA-methylation signatures as epigenetic biomarkers for disease such as schizophrenia.[23 38 39 40 41] Melas et al. used one PCR primer pair to examine differences in methylation of 5 CpGs sites in COMT by comparing blood cells from patients with schizophrenia versus controls.[23] They found that this region of S-COMT was hypermethylated in schizophrenia patients, though the study did not consider 22q11DS, and thus patients had diverse risks factors for the disease.[23] Many were on antipsychotic medications, which influenced methylation patterns. Our study design overcomes many of the limitations of the Melas *et al.* study. All of our patients have the same main risk factor for schizophrenia (i.e., 22q11DS). A possible influence of antipsychotic medication is minimized in two ways:

1) Some of the patients are too young to fully manifest psychosis and are not likely to be on antipsychotic medication (this will be confirmed by medical records); 2) We cultured the patients' cells for 10 days to eliminate exposure to such medications, and thus normalize the extracellular environment for patient and control groups.

Future DNA analysis of our primer regions has potential to uncover DNA methylation signatures that exist in 22q11 DS patients with severe psychiatric symptoms. Therapeutically, several studies are underway investigating the potential of DNA-methylation modulation, as an epigenetic solution, specifically DNMT interference.[42] In Alzheimer's studies, trials of DNA-methylation inhibitors show a return in neprilysin (NEP) mRNA expression in circulating endothelial cells.[43]

Once the sequencing data returns, it will be analyzed for methylation percentages between 22q11 patients and control patients. There are a number of proposed outcomes we may find. The first possible outcome would verify our long-term hypothesis; specifically hypomethylation in DGCR8 of 22q11DS patients compared to control subjects. Hypomethylation in the region could result in increased gene transcription of *DGCR8,* potentially compensating partially or fully for the missing copy in the 22q11 deletion region. By correlating this hypomethylation with our 22q11 patient psychological data, we could possibly find association between levels of psychosis or other psychiatric morbidity and levels of hypomethylation. The second possible outcome is hypermethylation in the 22q11 patients. Hypermethylation in the region could result in decreased gene transcription, decreasing the already low gene expression of *DGCR8* in

patients with a hemizygous deletion. Since other members of Dr. Pearce's groups are examining mRNA expression of DGCR8, this correlation should be forthcoming. The third potential outcome is a combination of hypomethylation and hypermethylation depending on region of interrogation of *DGCR8*. For example, regions in the CpG island could be methylated at a different level (either hypomethylation or hypermethylation) from regions in the CpG shores, in comparison to control patients, potentially contributing to an aberrant gene expression in the *DGCR8* that remains in 22q11 patients. The final possible outcome is that there is no apparent differential methylation percentages found between 22q11 patients and control patients. This could be due to our low sample size, making it difficult to find differences. However it could also mean that methylation on *DGCR8* does not contribute to the variation of schizophrenia seen in patients with 22q11, but instead it is possible that other genes in the 22q11 deletion may play a contributing factor.

Limitations of our study include potential methylation differences between blood and brain tissue, using cultured day-10 B-lymphocytes, restricted region of investigation, and few 22q11 case and control patients. We decided to pursue our methylation protocol on B-lymphocytes because of their accessibility from live patients of various age groups and because they show some level of expression of 22q11.2 gene of interest. Studies have been inconclusive in showing a methylation difference between blood and brain tissue with some studies showing methylation differences between the two tissues while others show methylation pattern is mainly conserved.[44] Although methylation patterns in peripheral tissue may not account for psychiatric phenotype, it could still serve the

important function of disease biomarkers.[23] Additionally, our study is setup to investigate methylation differences in patients' B-lymphocytes of day-10 cultured cells, which could have obtained aberrant methylation in the culturing process. Another important limitation to reiterate is the small quantity of DNA used in PCR and the effects it undoubtedly had on our results. Successful results may have been obtained for if a greater input of DNA was used. Additionally, our study produced successful investigation in only one area of one gene in the 22q11.2 deleted region. Even in this case, our study covers more CpG sites, including CpG islands and shores, than much of the established gene methylation literature.[19 23 26] With only five 22q11 patient cases and five patient controls, our future methylation analyses has limited potential for broad categorical conclusions. More patient controls and cases must be ascertained to establish a more robust methylation analysis, however our study is definite start with other methylation literature studying as few as 3 individuals.[44]

Although containing a number of limitations, our study expresses a number of strengths. Firstly, we investigate not only a CpG island but also 2 kilobase pairs upstream and downstream of that island, allowing us to cover more CpG sites in our future methylation analysis. The idea of exploring both the CpG island and CpG shores is novel, let alone that a number of methylation analysis studies only look at very limited numbers of CpG sites.[23 29] Additionally, we performed our methylation protocol according to methylation literature's established gold-standard technique of sodium bisulfite treatment. [25, 26, 27, 28] With regards to cell type, although we do not investigate brain tissue, our patient blood cells are from individuals who were alive at time of blood draw. Brain tissue studies are

post-mortem, introducing a series of limitations that our study will not contain, such as methylation difference ascertained because of age, drugs and extraneous disease from the one studied, along with epigenetic changes that happen during and after tissue death.[44] Finally, primer regions and DNA have been optimized to a consistent level so that future methylation analysis for this region can occur more efficiently. The process of finding and optimizing successful primers that thoroughly span a continuous region is not only challenging, it is quite novel.

Conclusion

By using a distinct cell type that was not transformed (B-cell), I was able to optimize PCR for 10 regions of DGCR8 (CpG islands and shores) and prepare amplicons from these regions from ten 22q11DS patients and controls to determine CpG methylation patterns. Given that DGCR8 plays an essential role in processing of microRNAs, an understanding of how CpG methylation patterns influence expression of this gene could have implications for the pathophysiology of schizophrenia and other mental illnesses in 22q11 DS, as well as immune abnormalities commonly found in this patient group.

References

[1] Weksberg, Rosanna, et al. "Molecular characterization of deletion breakpoints in adults with 22q11 deletion syndrome." *Human genetics* 120.6 (2007):837-845.

[2] Yamagishi, Hiroyuki, and DeepakSrivastava. "Unraveling the genetic and developmental mysteries of 22q11 deletion syndrome." *Trends in molecular medicine* 9.9 (2003):383-389.

[3] Sullivan, Kathleen E. "The clinical, immunological, and molecular spectrum of chromosome 22q11.2 deletion syndrome and DiGeorge syndrome." *Current opinion in allergy and clinical immunology* 4.6 (2004):505-512.

[4] Schreiner, Matthew J, et al. "Converging levels of analysis on a genomic hotspot for psychosis: Insights from 22q11.2 Deletion Syndrome." Neuropharmacology (2012)

[5] Bassett, A S, and E W WChow. "22q11 deletion syndrome: a genetic subtype of schizophrenia." *Biological psychiatry* 46.7 (1999):882-891.

[6] Levinson, Douglas F, et al. "Genome-wide association study of multiplex schizophrenia pedigrees." *The American journal of psychiatry* 169.9 (2012):963-973.

[7] Murgatroyd, Chris, et al. "Dynamic DNA methylation programs persistent adverse effects of early-life stress." *Nature neuroscience* 12.12 (2009):1559-1566.

[8] Nguyen, AnhThu, et al. "Global methylation profiling of lymphoblastoid cell lines reveals epigenetic contributions to autism spectrum disorders and a novel autism candidate gene, RORA, whose protein product is reduced in autistic brain." *The FASEB journal* 24.8 (2010):3036-3051.

[9] Petronis, Arturas, et al. "Monozygotic twins exhibit numerous epigenetic differences: clues to twin discordance?." *Schizophrenia bulletin* 29.1 (2003):169-178.

[10] Dempster EL, Pidsley R, Schalkwyk LC, et al. Disease-associated epigenetic changes in monozygotic twins discordant for schizophrenia and bipolar disorder. *Hum Mol Genet* 2011.

[11] Kaminsky, Zachary A, et al. "DNA methylation profiles in monozygotic and dizygotic twins." *Nature Genetics* 41.2 (2009):240-245.

[12] Mastroeni, Diego, et al. "Epigenetic mechanisms in Alzheimer's disease." *Neurobiology of Aging* 32.7 (2011):1161-1180.

[13] Bird, A.P., Wolffe, A.P., 1999. Methylation-induced repression--belts, braces, and chromatin. *Cell* 99, 451–454.

[14] Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Lands- berger, N., Strouboulis, J., Wolffe, A.P., 1998. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.* 19, 187–191.

[15] Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev*., 16, 6–21.

[16] Sweatt, J D. "Experience-dependent epigenetic modifications in the central nervous system." *Biological psychiatry* 65.3 (2009):191-197.

[17] Portela, Anna, and Manel Esteller. "Epigenetic modifications and human disease." *Nature biotechnology* 28.10 (2010):1057-1068.

[18] Cooper DN, Krawczak M (1989): Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes. *Hum Genet* 83:181–188.

[19] Irizarry, Rafael A, et al. "The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores." *Nature Genetics* 41.2 (2009):178-186.

[20] Rao, X., J. Evans, H. Chae, J. Pilrose, S. Kim, P. Yan, R. L. Huang, H. C. Lai, H. Lin, D. Miller, J. K. Rhee, Y. W. Huang, F. Gu, J. W. Gray, T. M. Huang, and K. P. Nephew. "CpG Island Shore Methylation Regulates Caveolin-1 Expression in Breast Cancer." *Oncogene* (2012): 1-10.

[21] Feinberg, A P, and BVogelstein. "Hypomethylation distinguishes genes of some human cancers from their normal counterparts." *Nature* 301.5895 (1983):89-92.

[22] Villa, R, et al. "Epigenetic gene silencing in acute promyelocytic leukemia." *Biochemical pharmacology* 68.6 (2004):1247-1254.

[23] Melas, Philippe A, et al. "Epigenetic aberrations in leukocytes of patients with schizophrenia: association of global DNA methylation with antipsychotic drug treatment and disease onset." *The FASEB journal* 26.6 (2012):2712-2718.

[24] Hamid Mostafavi Abdolmaleky, Kuang-hung Cheng, Stephen V. Faraone, Marsha Wilcox, Stephen J. Glatt, Fangming Gao, Cassandra L. Smith, Rahim Shafa, Batol Aeali, Julie Carnevale, Hongjie Pan, Panagiotis Papageorgis, Jose F. Ponte, Vadivelu Sivaraman,Ming T. Tsuang, Sam Thiagalingam Hypomethylation of MB-COMT promoter is a major risk factor for schizophrenia and bipolar disorder *Hum. Mol. Genet.* (2006) 15(21): 3132-3145

[25] Clark SJ, Statham A, Stirzaker C, Molloy PL, Frommer M (2006) DNA methylation: bisulphite modification and analysis. *Nat Protoc* 1: 2353–2364.

[26] Laird, Peter W. "Principles and challenges of genomewide DNA methylation analysis." *Nature reviews. Genetics* 11.3 (2010):191-203.

[27] Robinson, Mark D, et al. "Evaluation of affinity-based genome-wide DNA methylation data: effects of CpG density, amplification bias, and copy number variation." *Genome research* 20.12 (2010):1719-1729

[28] Schumacher, Alex. "Schumacher's Guide For: Bisulfite Conversion of DNA Methylation Fine-mapping." www.methylogix.com. N.p., 10 Feb. 2009. Web. 09 Nov. 2012.

[29] Murphy, Brenda C, Richard LO'Reilly, and Shiva M Singh. "Site-specific cytosine methylation in S-COMT promoter in 31 brain regions with implications for studies involving schizophrenia." *American journal of medical genetics.* Part B, *Neuropsychiatric genetics* 133B.1 (2005):37-42

[30] Karayiorgou, Maria, Tony J Simon, and Joseph A Gogos. "22q11.2 microdeletions: linking DNA structural variation to brain dysfunction and schizophrenia." *Nature reviews. Neuroscience* 11.6 (2010):402-416.

[31] Moreau, Michael P, et al. "Altered microRNA expression profiles in postmortem brain samples from individuals with schizophrenia and bipolar disorder." *Biological psychiatry* 69.2 (2011):188-193.

[32] Schofield, Claude M, et al. "Monoallelic deletion of the microRNA biogenesis gene Dgcr8 produces deficits in the development of excitatory synaptic transmission in the prefrontal cortex." *Neural development* 6(2011):11-11.

[33] Gregory RI, Yan KP, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, Shiekhattar R: The Microprocessor complex mediates the genesis of microRNAs. *Nature* 2004, 432:235-240.

[34] Beveridge, N J, et al. "Schizophrenia is associated with an increase in cortical microRNA biogenesis." *Molecular Psychiatry* 15.12 (2010):1176-1189.

[35] Metzker, Michael L. "Sequencing technologies - the next generation." *Nature reviews.* Genetics 11.1 (2010):31-46.

[36] Mardis, Elaine. "Current Topics in Genome Analysis 2012." Lipsett Amphitheatre of National Institutes of Health Clinical Center, Bethesda. 14 Nov. 2012. Lecture.

[37] Das Chakraborty, R, et al. "Dysregulation of DGCR6 and DGCR6L: psychopathological outcomes in chromosome 22q11.2 deletion syndrome." *Translational psychiatry* 2(2012):e105-e105.

[38] Heichman, Karen A, and Jorja DWarren. "DNA methylation biomarkers and their utility for solid cancer diagnostics." *Clinical chemistry and laboratory medicine* 50.10 (2012):1707-1721.

[39] Markopoulou, Soultana, Georgios Nikolaidis, and Triantafillos Liloglou. "DNA methylation biomarkers in biological fluids for early detection of respiratory tract cancer." *Clinical chemistry and laboratory medicine* 50.10 (2012):1723-1731.

[40] Wiencke, John K, et al. "Epigenetic biomarkers of T-cells in human glioma." *Epigenetics* 7.12 (2012)

[41] Wiencke, John K, et al. "Epigenetic biomarkers of T-cells in human glioma." *Epigenetics* 7.12 (2012)

[42] Nebbioso, Angela, et al. "Trials with 'epigenetic' drugs: An update." *Molecular Oncology* (2012)

[43] Urdinguio, Rocio G, Jose V Sanchez Mut, and Manel Esteller. "Epigenetic mechanisms in neurological diseases: genes, syndromes, and therapies." *Lancet neurology* 8.11 (2009):1056-1072.

[44] Davies MN, et al.(2012) Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. *Genome Biol* 13:R43.