A clustered regularly interspaced short palindromic repeats knockout method to reveal methyl-CpG binding domain 4 function

Ayden Ng (nga8@bxscience.edu)
Bronx High School of Science  https://orcid.org/0000-0002-2541-2656

Method Article

Keywords: Epigenetics, MBD4, methylation, Astatotilapia burtoni, CRISPR, plasticity

Posted Date: November 14th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2266782/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

DNA methylation is an epigenetic mechanism tailored for DNA repression, engineered for regulating genetic expression without direct manipulation of the nucleotide sequence. One component of this process includes methyl-binding proteins (MBD), which have an affinity for methyl groups, and they competitively inhibit transcription factors from binding to genetic promoters. Interestingly, MBD4 is unique because, as opposed to transcriptional repression, it promotes gene repair & demethylation and is associated with various methylation-related diseases, such as Autism. By further studying MBD4, we can identify a potential therapeutic target for MRD and further understand the role of methylation on the epigenome in regards to seasonal plasticity. Therefore, this paper describes a CRISPR Knockout screen to isolate & repress MBD4 from its customary functionality with gRNA targets GGAAGGGGGUGCUUGUGAUG and GGAAGGGGGTGCTTGTGATGTGG in Astatotilapia burtoni Cichlid. I expect a morphological change in the Cichlid's skin color (such change can be identified with computer vision COCO-Style-Dataset-Generator-GUI), which substantiates our belief that MBD4 does, in fact, play a significant role in seasonally-regulated epigenetic switches and can be targeted in methylation treatments. However, the exogenous factors relating to MBD4’s role in methylation remain to be investigated.

Introduction

Methylation is an epigenetic mechanism tailored for DNA repression, capable of regulating genetic expression without directly editing the nucleotide sequence [1]. It functions by transferring a methyl group (CH$_3$) from S-adenyl methionine (SAM) to the 5’ end of a cytosine (mC), via DNA methyltransferases (DNMTs), especially within CpG dinucleotides [2]. This attracts a group of proteins that possess a particularly significant affinity for methyl groups, also known as methyl-binding protein (MBD). The majority of MBD domain proteins repress gene expression by binding to promoters, thus competitively preventing essential transcription factors from properly functioning [3]. They are also responsible for interacting with a host of other proteins, including histone deacetylases [4] and chromatin [5], which disturb histone and heterochromatin structure, respectively. Interestingly, one version of MBD demonstrate characteristics apart from repression, as MBD4 is evidenced to contribute towards gene repair in regards to CpG - TpG (methylated cases) and CpG - UpG (unmethylated cases) mismatches; the associated binding partners include MLH1, DNMT1-UHRF/USP7, and DNMT3B [6]. In addition to the N-terminal MBD domain that all MBD proteins possess, MBD4 also has a unique C-terminal glycosylate domain that catalyzes their DNA repair & demethylation characteristics [7]. In fact, there is evidence to suggest that MBD4 mutations/misregulation contribute to autism spectrum disorder [8], cancerous tumors (especially in the gastrointestinal tract) [9], hematopoiesis illness [10], and even memory bias in injured mice with olfactory methylation [11]. Therefore, it is essential that we fully grasp the mechanisms behind MBD4 and its contribution to disease progression & ecology, which are both factors not yet well understood. For example, what role does MBD4 play in DNA methylation & disease? In what ways do MBD4 facilitate molecular plasticity in response to the environment? Does methylation contribute to seasonally transcriptional phenotypes, or vice versa? [12]
As a result, targeting MBD4 in a CRISPR/Cas9 knockout screen would reveal protein function necessary for further understanding the role of methylation in the human body, especially regarding the pathogenesis of disease. Clustered regularly interspaced short palindromic repeats (CRISPR) is a genetic-editing tool that is used to alter sequences of DNA. Cas9 cleaves DNA, which allows for insertion or removal of a desired gene, and it targets that gene with a sgRNA binding sequence matching that of the desired target sequence [13]. In a CRISPR knockout screen, CRISPR/Cas9 would cleave the target site, which would generate random misplaced insertions/deletions (indels) as a result of non-homologous end joining. These “mistakes” from NHEJ DNA repair pave the way for (framshift) mutations, thus rendering the gene dysfunctional to perform reverse genetics [14]. In other words, knockouts include inhibiting gene expression and observing the phenotypic consequences to infer gene function. The CRISPR knockout screen protocol can also be slightly modified for fish subjects [15]:

1. Breed wild type fish and collect the founder eggs
2. Synthesize sgRNAs for high indel rates
3. Microinject the sgRNA with CRISPR/Cas9 into founder eggs
4. Phenotype the resulting founding adult Cichlid (F₀ generation)
5. Outcross the mosaic F₀ adults with wild types
6. Cross the confirmed F₁ heterozygotes and sequence for mutations
7. Phenotype the homozygous F₂ generation

In fact, Astatotilapia burtoni fish (blue-yellow color morph) serves as the ideal model for a knockout screen, given the reliability of their observable plastic phenotypes. For instance, such males exhibit dominant sexual behavior and are known to be aggressive when protecting their territory during mating season. Though, this behavior can be easily influenced by its dynamic social environment [16], making A. burtoni a reliable model to study the mechanisms in which methylation could contribute to their plastic behavioral traits. In fact, the upregulation of various neuroendocrine genes in territorial males suggests that the "socially-regulated phenotype switch" is particularly apparent in this species of Cichlid [17], and observable color changes in Cichlid skin have similarly been exploited for reverse genetics [18]. Therefore, the unique plastic phenotype of A. burtoni allows us to study the effect of the environment on epigenetic genomics and will be the model of choice as a result.

I hypothesize that MBD4 plays an important role in the environmentally-regulated switch that Cichlid phenotypes (behaviorally & physiologically) function upon, especially given the host of evidence demonstrating a correlation between methylation levels and transcriptional phenotypes [19]. For instance, hibernation-specific protein-27 (HP-27) in chipmunks is upregulated during hibernation [20], suggesting a seasonal or situational effect on epigenetics. Since MBD4 is directly associated with the manner in which methylation functions [21, 22], it's reasonable to infer that MBD4 would play a similar role. Of course, there is limited evidence to suggest that MBD4 directly affects the adaptive plasticity of vertebrate species (as opposed to exogenous factors), which is the main factor to be investigated in this study. Therefore, if knocking out the MBD4 gene in A. burtoni fish presents any significant alterations in phenotype, it can be assumed that MBD4 is the main contributor to plasticity or the earliest contributor to
phenotype in the methylation pathway. This could reveal a therapeutic target to methylation-related illnesses [23–26] and hint at the role of epigenetics on the genome in a dynamic environment.

Methods

Subjects

I will use Astatotilapia burtoni fish, including 20 females/1 male, collected from Lake Malawi [27]. Adult fish will be held in 30 L tanks with regulation [28] pH (8.0–8.2), salinity (320-480ppm), and temperature (25°C) with 12-hour light-dark schedules. All experimental procedures will be approved by the Animal Care Committee.

In mating, all fish will be injected with Ovaprim [29] (Syndel) at 0.5 µL per gram of body mass and allowed to mate for 30 minutes. Embryos (flushed from the oral cavity) will be collected with 4mm transfer pipettes and stored in well plates treated with 6mL of 1mg/L methylene blue antifungal reagent (Sigma) water after collection. 100 embryos will be used (80 experimental, 20 negative control), since evidence shows that it takes 6 founder fish to chance a mutation [30].

Experimental Procedure Overview

I will perform a CRISPR/Cas9 knockout screen on the MBD4 gene. This includes a CRISPR/Cas9 injection into A. burtoni embryos to breed the founder F₀ generation. I then will outcross the founder and wild type fish to limit off-targets, and will later screen the F₁ generation for mutation germline transmission with fluorescent PCR & sequencing. Lastly, after crossing the mutant heterozygotes, the homozygous F₂ generation will be finally examined for phenotype. Untreated wild-type genotypes & phenotypes are used as controls. The procedures below follow the workflow described here in this overview.

sgRNA Design and Synthesis & Cas9 Preparation

I summarize below the set of rules that have been previously established for effective gRNA design [31–33] along with my experimental specifics in parentheses immediately following each rule:

1. Target gene must be defined (MBD4)
2. Cas9 protein & PAM sequence must be defined (SpCas9 & NGG)
3. gRNA Promoter must be defined for expression (T7 Polymerase)
4. High GC Content (45–90%), limited predicted off-targets (< 500), exon targeting (close to 5’ end; region should be present in all isoforms and in a functional protein domain) are also important factors that were considered.

I therefore curated a target sequence with the nucleotide pattern GG - N₃₋₂₀ - NGG. CHOPCHOP [34] (https://chopchop.cbu.uib.no/) is a reliable tool that is commonly used for curating gRNAs with such specificity, as it automatically searches within the A. burtoni genome to rank potential targets. I set the target to be MBD4 in AstBur1.0 with CRISPR/Cas9 for knock out and adjusted the target specificity to
match the rules above in the “Options” bar. I decided to choose the top 2 sequences targeting different exons, as there's evidence to suggest that this allows for the greatest possibility (99%) for germline transmission [35].

**Remote gRNA Sequences Can be Found in Results***

The target sequences were then cross-checked with Synthego's Guide Design Verifier (https://design.synthego.com/#/validate) to confirm the effectiveness of each sequence as suggested by CHOPCHOP. The final oligonucleotides will be created by annealing together a 5’ T7 promoter and a 3’ overlapping gRNA sequence to the target sequence, not including the PAM site (ordered on IDT). This is considered the top strand:

5’ - TAATACGACTCACTAT - gRNA - GTTTTAGAGCTAGAAATAGC - 3’

**Remote Full Oligos Can be Found in Results**

In order to synthesize the crRNA-tracRNA, I will use a recently developed cloning-free method [36] that require the expanding of the top strand and a universal bottom strand (AAAAGCACCGACCTCGGTGCCACTTTTTCAAGTTGATAACGGACTA GCCTATTTTTAACTTGCTATTTCGCTCTCTAAAC). I will mix the top & bottom strand solutions and use PCR amplification to extend the new oligo to a confirmed 120bp according to protocol. Lastly, I will combine a T7 RNA synthesis mix (NEB Catalog #E2040S) to the new oligo, PCR the solution, add DNase, and purify the final reaction, all according to the manufacturer’s instructions. The resulting 2 sgRNAs (2 targets) will be stored at -80° C and will be co-injected.

Cas9 mRNA will be derived by digesting pT3TS-nls-zCas9-nls plasmid (Addgene, cat no. 46757) with XbaI and preparing RNA transcription reaction with mMESSAGE mMACHINE T3 kit (Life Technologies) [37].

**Microinjection**

I will prepare micropipette needles according to methods used in Sticklebacks with MPPI-3 Pressure Injector (Applied Scientific Instrumentation), adopting air supply, pressure, dissecting microscope magnification, and other accepted settings [38]. The needles will be held on ice vertically in a capillary storage jar.

To prepare the microinjection contents, I will combine 200ng of the synthesized sgRNA (25pg), 400ng of Cas9 mRNA (300pg), 0.5µl of 0.5% DPBS (Elabscience, EP-CM-L0409), and 5µl of RNAse free water. I will then pipet these contents into the blunt-end of the needle at 0.5µl increments until the needle is filled and the very end of the needle is broken with watchmaker’s forceps at a 60° angle. 25 minutes post fertilization, I will separate the embryos into each indentation of a saw blade and flush each embryo with water for 2 seconds to swell the chorion. To start the injection, I will rotate the embryos until the blastomere is directly perpendicular to the injection needle, and then push the needle into the cytoplasm outside of the yolk. I inject the solution with a foot pedal until a red spot with diffuse edges fills about of
the cytoplasm's diameter. A transfer pipette will be used to transfer the injected embryos into the well plates. I will check the health of these embryos daily.

**Verify Mutations with Fluorescent PCR & Sequencing**

At 48 hours post injection, I will collect diluted DNA samples from the Cichlid embryos with Extract-N-Amp (Sigma, E7526), according to the manufacturer's instructions. Additionally, I designed primers to begin fluorescent PCR by adding a M13F sequence to the forward primer and a PIG-tail sequence [39] to the reverse primer, all annealed at the 5' end of the gene-specific primers suggested by CHOPCHOP, as follows:

Remote Designed Primers Can be Found in Results***

I will create the fluorescent primer mix by mixing 5 µl of these 100µM forward/reverse primers, 485 µl of TE pH 8.0 (Quality Biological, cat no. 351-011-131) and 5 µl of 100µM M13-FAM primer (/56-FAM/TGTAACGACGGCCG). I will then add 6 µl of the fluorescent primer mix to 100 µl of PCR master mix [40] to synthesize the final mix, and finally add 1.5 µl of diluted DNA to 5 µl of the final mix. The resulting solution is deemed the analyzing product.

To start fluorescent PCR, I will amplify the analyzing product in a thermocycler following these conditions: 94°C, 12 min, denature cycle 1; 94°C, 30 sec, denature cycle 2–36; 55°C, 30 sec, anneal cycle 2–36; 72°C, 30 s, extend cycle 2–36; 72°C, 10 min, extend cycle 37; 4°C, hold cycle 38. I will then analyze the resulting product with a Genetic Analyzer 3130xl (Life Technologies, cat no. 3130XLR) on CRISPR-STAT parameters [41]. GeneMapper (Life Technologies, cat no. 4475073) with the accepted settings [42] in combination with Peak Scanner 2 (Thermofisher) can be used to display peak sizes from 3130xl, and the embryos that illustrate frameshift mutations, shown by more than 1 peak present, will be incubated to start the mutant lines. A low ratio between areas under the wiltype peak vs all peaks indicates high mutation frequency [43]. In adult Cichlid, I will examine DNA from their servered caudal fin. Sanger Sequencing with 3730xl [44] and PolyPeakParser [45] provides the exact nucleotide sequence to confirm genotype and allele mutation. I will lastly examine the fish for phenotypic expressions as a result of the induced mutation. Observations can be identified automatically with computer vision COCO-Style-Dataset-Generator-GUI (https://github.com/hanskrupakar/COCO-Style-Dataset-Generator-GUI).

**Results**

Presented are the computational results for designer strands:

**gRNA**

5’ - GGGAGGTGTCCACTTCTTGAAGG – 3’

5’ - GGAAGGGGGTGCTTGTGATGTGG – 3’

These sequences were verified for efficacy with Synthego software. (Fig. 1).
**Oligos**

Created by annealing a T7 promoter and an overlapping universal sequence to the gRNA targets, as explained in the methods section.

5’–TAATACGACTCACTATGGGAGGTGTCCACTTCTTGAAGGCTAGAAATAGC–3’

5’–TAATACGACTCACTATGGGAAAGGGGTGCTTGTGATGGTTTAGAGCTAGAAATAGC–3’

**Primers**

**Target:** GGGAGGCTTGCCACTTTCTTGAGG

Forward: 5’ - TGTAAAACGACGGCCAGTCCTTATTTCAGTGGGAAATCCA − 3’

Reverse: 5’ - GTGTCTTCAGGGATCATGGAAAAGAGTTT − 3’

**Target:** GGAAGGGGGTGCTTGTGATGTGG

Forward: 5’ - TGTAAAACGACGGCCAGTTTCACCACAGAAGCCCTTAGAT − 3’

Reverse: 5’ - GTGTCTTGAGGTGCCTCTAGAATTGCTGT − 3’

**Discussion**

Given the demonstrated efficiency of the designer strands, the probability for success is significantly high in this experiment. gRNA strands have been optimized for efficiency. Between my selected target strands GGGAGGCTTGCCACTTTCTTGAGG and GGAAGGGGGTGCTTGTGATGTGG, both exhibited satisfactory qualities for a traditional CRISPR knockout, including a high GC content, limited off-targets, and wide exon targeting [46]. Primers were also optimized for a CRISPR knockout, especially regarding base pair length from 200bp-300bp [47]. After injecting the CRISPR knockout solution, we can observe phenotype either manually or with COCO-Style-Dataset-Generator-GUI (https://github.com/hanskrupakar/COCO-Style-Dataset-Generator-GUI), as described earlier.

**Conclusions**

Therefore, I conclude that this CRISPR KO experiment will function effectively and will reveal that the MBD4 gene plays a significant role in an environment-regulated switch that facilitates phenotypic plasticity. CRISPR knockout will likely disable MBD4 gene expression [48] theorized to play a key role in DNA methylation. Additionally, it’s well known that hibernation-specific protein-27 (HP-27) in chipmunks is upregulated during hibernation [49], suggesting seasonal regulation on epigenetics. By relation, it’s inferable that DNA methylation, an epigenetic process [50], is similarly manipulated via MBD4 with seasons. In Cichlid, this relationship can usually be indicated by a morphological physical change [51] MBD4 would result in a difference in skin color and/or sexual behavior in comparison to the control.
MBD4-intact Cichlid. This is particularly significant because the unique plastic phenotype of *A. burtoni* allows us to study the effect of the environment on epigenetic genomics and provides us the opportunity to reveal a direct correlation between MBD4 & plasticity. A change in skin color and/or behavior is evidence of such proposition and could reveal a therapeutic target to methylation-related illnesses [52–55].

However, many questions are still left unanswered: What exogenous factors could still affect the phenotypic expressions of methylation? What characteristics of each season prompt molecular plasticity? How can we apply the role of MBD4 on methylation in therapeutic treatments?

**Abbreviations**

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

CRISPR KO: Clustered Regularly Interspaced Short Palindromic Repeats knock out

*MBD*: Methyl-CpG binding domain

MBD4: Methyl-CpG binding domain 4

DNA: Deoxyribonucleic acid

*A.Burtoni*: Astatotilapia burtoni

gRNA: Guide ribonucleic acid

sgRNA: Single guide ribonucleic acid

TracrRNA: Trans-activating clustered regularly interspaced short palindromic repeats ribonucleic acid

SpCas9: *Streptococcus pyogenes* clustered regularly interspaced short palindromic repeats associated protein 9

DNMT: Deoxyribonucleic acid methyltransferases

SAM: S-adenyl methionine

**Declarations**

**Conflicts of Interest**

The author declares no conflict of interests.

**Author Contributions**
AHN: made substantial contributions to the design of the research protocol, the collection of data as well as interpretation and analysis of the data, revised the manuscript critically, and gave final approval of the version to be published.

Acknowledgements

I thank Dr. Emily Schmidt for providing writing assistance on this manuscript.

Funding

This study was not funded.

References


55. Tajerian M, Alvarado SG, Clark JD. Differential olfactory bulb methylation and hydroxymethylation are linked to odor location memory bias in injured mice. Molecular pain. 2019; 15. https://doi.org/10.1177/1744806919873475

Figures
Synthego verification. Target sequence GGGAGGTGTCCACTTCTTGAGG (PAM GGA) correlates with gRNA sequence GGAGGUGUCCACUUCUUGA (Above). There’s evidence that this gRNA will target a common exon, exhibit high activity, and have minimal off-targets. Target sequence GGAAGGGGTGCTTTGTGATGTGG (PAM GGT) correlates with gRNA sequence
GGAAGGGGGUGCUUGUGAUG (Below). There’s evidence that this gRNA will target a common exon in an early coding region, exhibit high activity, and have minimal off-targets.

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>NU_005179658.1:286654-286676</td>
<td>CTTATTCATGGGAAATCCA</td>
<td>69.2</td>
<td>0</td>
<td>NU_005179658.1:286868-286890</td>
<td>CAAGGATCATGGAAAAGTTT</td>
<td>59.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>NU_005179658.1:285180-285411</td>
<td>TTCACCCAGAAGCCTTAGAT</td>
<td>60.1</td>
<td>0</td>
<td>NU_005179658.1:285626-285648</td>
<td>GAGGTCCTCTAGAATGTCTGT</td>
<td>50.9</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Figure 2**

Primer synthesis. *Target GGGAGGTGTCCACTTCTTGAAGG is calculated to have left primer CTTATTCAGTGGGAAATCCA and right primer CAGGGATCATGGAAAAG7jAGTTT (3). Target GGAAGGGGGTGCTTGTGATGTGG is calculated to have left primer TTCACCCAGAAGCCTTAGAT and right primer GAGGTCCTCTAGAATGTCTGT (4). Efficiency is centered around 60.0, and base pair length is centered around 245 bp for each primer.*