Title
Reporter Gene Silencing in Targeted Mouse Mutants Is Associated with Promoter CpG Island Methylation.

Permalink
https://escholarship.org/uc/item/35h3w8nnm

Journal
PloS one, 10(8)

ISSN
1932-6203

Authors
Kirov, Julia V
Adkisson, Michael
Nava, AJ
et al.

Publication Date
2015

DOI
10.1371/journal.pone.0134155

Peer reviewed
Introduction

Random integration of foreign DNA into mammalian genomes is known to provoke a response resulting in histone modification, and marked by DNA methylation at CpG dinucleotide sites, with the end result being the silencing of any potential transcriptional elements. This silencing is particularly effective against repeat elements [1] and retrotransposon sequences [2]. However, since the degree of silencing depends upon the site of insertion, local chromatin organization and features also must play a role. Silencing has been problematic in the construction of vectors for random transgene insertion and expression in the creation of animal models since vectors are often inserted as concatemers and provoke silencing [3]. Furthermore, the potential for silencing of viral sequence is an important consideration for developing strategies for gene therapy [4]. Silencing of engineered transgenes has been mitigated by avoiding viral
repeat elements known to provoke silencing [4], by engineering into the vector flanking insulators of DNA sequence to reduce local effects of the region on the transgene [5–6], and also by targeting the transgene to genomic regions thought to be less responsive to the presence of foreign DNA, e.g., the Rosa26 locus in mice [7–8].

The majority of gene targeting experiments in mammalian systems are designed to eliminate function of targeted genes, although many “knockins” have been designed to introduce specific mutations, or to express alternative sequences under the control of the native gene promoter. Often targeting vectors contain reporter sequence such as bacterial beta-galactosidase (LacZ) or green fluorescent protein, as well as selectable markers such as a neomycin resistance cassette in order to facilitate mutant selection in stem cell populations [9]. Since the majority of these gene targeting events are designed to knock out, or eliminate gene expression, then the consequences of silencing has been thought to be manageable, except that silencing of vector elements in the stem cells might interfere with selection of targeted cells using antibiotic resistance, or cause silencing of a reporter gene in the adult mutant. Strategies have been developed to eliminate most of the foreign DNA from targeting vectors after genomic integration by engineering recombinase sites flanking the selection cassette allowing the removal of vector components at any stage in the production of the model [10–11].

An earlier report in mouse studied the silencing of a randomly integrated transgene containing LacZ driven by a ubiquitous promoter. Silencing of transgene expression was correlated with the CpG content of the LacZ sequence in an allelic series of random integrants [12]. In that report, decreasing CpG content of the LacZ sequence was correlated with decreased methylation of CpGs in the heterologous promoter and reduced silencing assessed by enzyme activity. These data suggest that the CpG content of reporter genes, or other elements in transgenes, may have important local effects.

Recent large-scale mutagenesis programs in mice have created a valuable resource for studying the consequence of loss of function mutations in mammalian systems. Programs such as the European Mouse Disease Clinic (EUMODIC; www.europenome.org; [13]); the Sanger Center Mouse Genetics Project [14]; KOMP knockout projects [15], and other programs recently organized as the International Knockout Mouse Consortium (www.mousephenotype.org), are producing thousands of targeted loss-of-function mutations in mouse protein-coding genes. Many of the targeting vectors contain both a reporter gene (LacZ) along with a neomycin resistance selectable marker. These mutations will provide a valuable resource for studying gene function, but they also provide a remarkable resource for studying the unintended consequences of targeting such as silencing of the targeted gene, or effects on the expression of neighboring genes. Since the targeting vector sequence is constant, while the location and local environment changes with each targeting event, then it may be possible to use this resource as a tool to better understand and characterize the mechanisms provoking transgene silencing and by inference gene regulation.

In a pilot study characterizing LacZ reporter gene expression in KOMP mutants [16] we noted that a subset of mutants had no LacZ staining despite gene expression surveys indicating the gene was expressed at moderate to high levels in some tissues. This suggested to us that these mutations may not be staining for LacZ due to promoter silencing of the targeted gene. In order to assess this, and preliminarily evaluate if DNA methylation marks the silenced genes, we identified a set of KOMP mutants with expected patterns of LacZ staining, and another set of mutants with no LacZ staining although it was expected based upon other gene expression surveys. In these two sets of mutants, we evaluated promoter CpG island methylation, LacZ coding sequence methylation, and quantified mRNA of the targeted gene and the LacZ reporter in order to determine if gene silencing was a possible explanation for the lack of reporter gene staining.
Materials and Methods

Mouse Production

Homozygous mutant C57BL/6N mice, and wild type controls, created from Knockout Mouse Project (KOMP) targeted stem cells [9], were studied. The mutant allele was either the KOMP CHORI-SANGER-DAVIS (CSD) "Knockout First" conditional-ready allele or a CSD "deletion" mutant where the recombination removed the critical exon. These alleles are gene traps and carry a LacZ reporter gene driven by the targeted gene promoter, along with a Neo selection marker driven by a heterologous promoter. For a description of these alleles, visit the website of the International Mouse Phenotyping Consortium (IMPC: www.mousephenotype.org).

Targeting in the mice was confirmed by long-range PCR and zygosity by qPCR of LacZ coding sequence [17]. Homozygous Mutants, and Wild Type control (WT, Control) mice, were produced by heterozygous breeding of mutants and therefore shared a common C57BL/6N background strain and were reared in identical environmental conditions. Pups were weaned at ~21 days of age, and maintained on ad libitum Harlan Teklad Global Rodent Diet #2918 and water, in an environmentally controlled facility with a 12:12hr light:dark cycle. Mice were euthanized at ~7 weeks of age under isoflurane anesthesia with a thoracotomy and cervical dislocation. Tissues were rapidly harvested and quick frozen in liquid nitrogen, and stored at -80degC until used. All animal work followed the Guide for the Care and Use of Laboratory Animals of the Institute of Laboratory Animal Research of the National Institutes of Health and was approved by the Institutional Animal Care and Use Committee at the University of California, Davis.

We selected six mutant lines for study based upon LacZ staining pattern in adult mutants, gene expression patterns described in two gene expression atlases, and the presence of CpG islands (CGI) in the promoter of the targeted gene. These mutant lines were selected from a pool of ~90 mutant lines that were part of a pilot study of 313 mutant lines with LacZ staining, for which we had also banked frozen tissue samples. It was noted that three of the mutant lines had no LacZ staining even though it was expected based upon gene expression atlases using Affymetrix Chip technology. We identified three mutant lines from the same set of ~ 90 lines, with similar gene expression data in these other atlases which did have LacZ staining. The LacZ staining pattern for these mutants is reported at: www.kompphenotype.org [16]. Presence of CGIs was determined from analyses of mouse genomic DNA sequence using the UC Santa Cruz Genome Browser (http://genome.ucsc.edu/index.html;[18]) with the criteria defining a CGI of: a minimum length of 200bp, a minimum GC content of 50% and observed-to-expected CpG ratio greater than 60%. Tissue expression patterns were assessed by consulting two mouse expression atlases in the BioGPS database (www.biogps.org;[19]) and the Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) available at the National Center for Biological Information. We identified three mutant lines (Lyplal1tm1a(KOMP)Wtsi, Rab32tm1a(KOMP)Wtsi, Rgcctm1(KOMP)Wtsi) with apparent reporter silencing (Silenced) based upon the lack of LacZ staining in mutant tissues despite expression data-bases reporting moderate to high expression of the native gene. Three mutant lines were identified (Ninj1tm1a (KOMP)Wtsi, Dstntm1a(KOMP)Wtsi, Arap1tm1a(EUCOMM)Wtsi) with LacZ staining as predicted by the gene expression atlases (Expressed). Four of these alleles were the "Knockout First" allele, the Rgcc mutant was a deletion mutant where the critical exon was deleted, and the Dstn mutant carried a promoterless neo selection marker. All of these targeted genes had CpG islands in the promoter. See S1 Table for a list of mutants and tissues, and a summary of the gene expression data from BioGPS and GEO. For the Silenced Group of mutants we analyzed tissues with the highest gene expression according to the databases. For the Expressed group of mutants we analyzed 7 tissues where available including: brain, spleen, heart, kidney, liver, lung, muscle in order to evaluate the same tissues as evaluated in the Silenced group. Plots of CpG islands in
Experimental Design

We used qRT-PCR to assess mRNA levels of each gene in wild type control and mutant mice, and to measure the mRNA of the LacZ reporter gene in mutants. By sequencing bisulfite-treated DNA, we assessed DNA methylation of CpG islands found in the promoter regions of each targeted gene, and methylation of the LacZ reporter gene coding sequence in mutants. Fig 1 presents a schematic of the gene structure and the DNA sequences assayed for CpG methylation and for quantitating expression by qRT-PCR.

RNA Isolation and Quantitative RT-PCR

Tissue RNA isolation was performed using TRI-reagent (Sigma, St. Louis, MO) according to the manufacturer’s protocol. To ensure homogeneous disruption, tissue samples weighing 30-60mg were homogenized in 1ml of TRI-reagent using a bead mill at 30Hz for 45sec x2. Heart and muscle were ground to a fine powder with a mortar and pestle on dry ice prior to bead mill homogenization. Phase separation was achieved using 0.2mL of chloroform followed by isopropanol precipitation of the RNA from the aqueous phase. The remaining phases were stored at 4degC for later DNA isolation. RNA precipitated from the aqueous phase samples was solubilized in DEPC-treated water. RNA concentrations were assessed using a NanoDrop 1000 (Thermo Scientific) and quality was assessed using a BioAnalyzer 2100 (Agilent). RNA Integrity numbers from the Bioanalyzer, an assessment of RNA quality, are reported in S2 Table.

All RNA samples were treated with DNase-I (Ambion, Turbo DNA-free DNase-I, Life Technologies). A total of 6–10ug of RNA was treated in a 50μL reaction. RNA (2ug) was transcribed to cDNA using the High Capacity RNA to cDNA kit (Applied Biosystems). Specific transcripts were quantified by qRT-PCR using pre-validated IDT PrimeTime qPCR primers and probes, or custom designed primer probe sets (Table 1). Specifics of the reaction mixes for Dnase-I treatment, reverse transcription, and qPCR are provided in S3, S4 and S5 Tables. Primers were designed to span exon junctions at the 3’ end of each targeted gene and to have high specificity (S6 Table). Primer probe pairs were analyzed by serial dilution against wild type cDNA to confirm high efficiency of the qRT-PCR reaction and these data are presented in S2 Fig and S7 Table.

Triplicate PCR reactions for each gene, tissue and biological replicate were set-up in a laminar flow hood at room temperature with a master mix containing hot start TaqDNA polymerase. Amplification and qPCR measurements were performed using the Applied Biosystems 7900HT Fast Real-Time PCR System, v 2.4.1 software. Thermal cycling conditions were: 10 min at 95°C for initial denaturation; and 40 cycles of 15 sec at 95°C and 30 sec at 60°C. Each reaction contained 2 μL of template cDNA and a reaction master mix containing 2X Thermo Scientific Maxima Probe/ROX qPCR Master Mix (Thermo Scientific, USA), 500 nM of each primer and 250 nM of probe. All qPCR assays were run with appropriate controls including the Non-Template Control and minus RT control. The qRT-PCR experiments were designed and conformed to the MIQE guidelines as described by Bustin [20]. Data were analyzed using ΔΔCt method with Actb as the internal reference [21].

The qRT-PCR was performed with normalization to Actb gene expression and relative quantities of transcript were determined by the delta delta Ct method. Data are presented as the targeted gene, and LacZ reporter gene expression in each mutant tissue, relative to wild type control gene expression in the same tissue. The mRNA abundance of the native gene was
comparable for the Silenced and Expressed groups with a delta Ct relative to Actb averaged across all genes and tissues of 5.5 for the Silenced group and 4.86 for the Expressed group.

DNA Extraction, Bisulfite Treatment, PCR & Amplicon Sequencing

The general method for determining methylation status of CpGs in amplicons of bisulfite-treated DNA has been described and validated by Masser et al. [22]. DNA was extracted from the TRI-reagent interphase using Back Extraction Buffer (Life Technologies). Specifics of the Back Extraction mix are presented in S8 Table. After incubation and centrifugation, the aqueous phase containing DNA was removed and transferred to a clean 1.5mL tube. DNA was precipitated using isopropanol. The resulting DNA pellet was dissolved in TE Buffer (Sigma). Concentrations and quality of the DNA were assessed using a NanoDrop 1000. Genomic DNA (5µg) was treated with bisulfite using the MethylEasy Xceed kit (Human Genetic Signature, Australia) according to the manufacture’s protocol and eluted with 50µL of the elution buffer. Bisulfite conversion rates were confirmed to be >98% by analyzing the cytosine to thymine conversion for cytosines not in CpG motifs.

Amplification of bisulfite treated DNA was carried out using KAPA2G Robust HotStart DNA Polymerase (KAPA Biosystems). We used a semi-nested PCR to produce clean defined bands which were excised for library preparation and sequencing. Primers used for amplification of CpG Islands are listed in Table 2. For each CpG island, from 3 to 12 overlapping amplicons were sequenced. For LacZ methylation, a set of seven non-overlapping amplicons were used covering 67% of the coding sequence. The primers used for LacZ amplification are listed in Table 3. Each PCR reaction had unique conditions depending on the Tm’s of the primer sets. PCR products were separated by agarose gel electrophoresis, excised and DNA was purified using E&K Gel Purification kit (E&K Scientific).

Library preparation for Next Generation Sequencing. Indexed libraries were prepared using TruSeq Sample Prep Kit Sets A/B (Illumina RS-122-2001/2) with the protocol started at the end-repair step with amplicons sonicated to 150-200nt length. Sequencing was completed either with the MiSeq Platform at the UC Davis Genome Center (150nt reads/paired end), or with the HiSeq2500 platform at the QB3-Berkeley Sequencing Core (single end 50nt). The number of reads per amplicon passing QC averaged 37,000.
Bioinformatics and Statistical Analysis of Methylation Calling. Sequencing reads were aligned to the MM9 genomic sequence using the Bowtie2 algorithm [23], and Bismark [24], was used for Bisulfite Indexed Genome creation and methylation calling. CpGs with reads <5 and assumed to be artifacts of PCR, and those CpGs falling into primer annealing sequence, were not included in the analysis. For each CpG, the proportion of methylated vs. non-methylated was calculated. Alignment files were converted to BAM format using SAMtools for visual inspection using the Integrated Genomic Viewer from the Broad Institute.

Statistics
At each specific CpG site within a CGI, the significance of differences between the mutant and wild type allele were calculated with the non-parametric Fisher Exact test. For overall CGI differences in methylation, the mutant and controls were compared using the Cochran-Mantel-Haenszel statistic for categorical data. Comparing the Silenced and Expressed groups for percent CGI methylation, and for relative gene expression by ΔΔCt, we used a one-tailed t-test. Regression analysis was completed by the least mean squares method and an F statistic calculated to determine significance.
| Primer   | Sequence                           | Sequenced Amplicon length |
|----------|------------------------------------|---------------------------|
| Arap1-BS-F1 | GTA GTTTAGGATAAGGAGAGATATAAAAT     | 224                       |
| Arap1-BS-R2 | CAA CA CCTA ACCT AT TT TAAAC      |                           |
| Arap1-BS-R3 | CCCCCCAA AAAACRCRCACCAAAT AAC     | 344                       |
| Dstn-BS-F1 | GGT GTTTTTTGGGAGTGTTTTTGAAGTTG    | 225                       |
| Dstn-BS-R1 | CTACACAA ATTTTTCTCTTCTCTAAAC      |                           |
| Dstn-BS-F2 | GAA ATAGYGGTTGATTTATATAGTATTGG    | 170                       |
| Dstn-BS-R2 | CCRATAAAC AATCTCCTACTACAC         |                           |
| Dstn-BS-F3 | GTTTTTGAAGTGTTTTGAGAGTT          | 150                       |
| Dstn-BS-R3 | CTCACAAA TCTACCTACCTCAAAAC        |                           |
| Dstn-BS-F4 | GGT GGYGATTTTTATGTAAGGTAAGGTT    | 162                       |
| Dstn-BS-R4 | CCCCCCAAAA ATCCCAAAT AAC           |                           |
| Dstn-BS-F5 | GGGTTTTTTGTTTTGGAGATTATG         | 230                       |
| Dstn-BS-R5 | CCCCCGAA ACCTACACCTACATTC        |                           |
| Lyplal1-BS-F3 | GGAAGAGTGGAGAGTGGTTTTGAGGTT     | 173                       |
| Lyplal1-BS-R3 | CTCACAA ACTACCTACCTACCAAC        |                           |
| Lyplal1-BS-F4 | GGGTYGGAGATTAYGAGATTGTTATTGAGG    | 147                       |
| Lyplal1-BS-R4 | CCRACACCA ATACCTACCTACATACAC     |                           |
| Ninj1-BS-F1 | GGGGATATTAGGGTYGTTGGAGGTT         | 198                       |
| Ninj1-BS-R1 | CTCACAACTCTCAATACCCRAACTCCATAAAC |                           |
| Ninj1-BS-F2 | GGGGYYGGTTAGGGYGGTTTTGAGGTT      | 177                       |
| Ninj1-BS-R2 | CCAACCAAACRCRCACCAAATAC          |                           |
| Ninj1-BS-F3 | GGGTGGTTTAGTGTGTGTGATTTTGT       | 162                       |
| Ninj1-BS-R3 | CCCCRAACCTCTACCTACCTACCTACAC     |                           |
| Ninj1-BS-F4 | GGGTTTTTTGTTTTGGAGATTATG         | 184                       |
| Ninj1-BS-R4 | CCCCCAATCTACCTACCTACATACAC       |                           |
| Ninj1-BS-F5 | GGGTTTTGATTATTTTGGGGGAGGTT       | 156                       |
| Ninj1-BS-R5 | CCCCCAAA CCTACCTACACRAACGTGTAATAC |                           |
| Ninj1-BS-F6 | GGGTTGGAAGTGGAGATTAGTAGGATAG     | 198                       |
| Ninj1-BS-R6 | CCAAACATCCCAATACRGTATACCTACCTA  |                           |
| Rab32-BS-F1 | GGGTTGGATAGGAGGATTTTTATAGTAAGAAGGTTG | 147                   |
| Rab32-BS-R1 | CACCAACCTCTCTCCCAAACCT           |                           |
| Rab32-BS-F2 | GAGGTTTTTGGTGTGGTTTTTGAGGTT     | 181                       |
| Rab32-BS-R2 | CCAACAACTACCTACCTACCAAC          |                           |
| Rab32-BS-F3 | GAGGTTGGATAGGAGGATTTTTATAGTAAGAAGGTTG | 161                   |
| Rab32-BS-R3 | CACCACAAA ACTACCTACCTACCTAC     |                           |
| Rgcc-BS-F1 | GAGGTTTTAGGAGGATTTTTATAGTAAGAAGGTTG | 231                   |
| Rgcc-BS-R1 | CCCCATAACCTACCTACCTACCAATAC     |                           |
| Rgcc-BS-F2 | GGGTTGTAGGGTTTAGGGTTTTTGAAGT     | 178                       |
| Rgcc-BS-R2 | CRCRTAAAATAAAATAACTACCTACCTACCTACCTAC | 207                   |
| Rgcc-BS-F3 | GGGTTGGATAGGAGGATTTTTATAGTAAGAAGGTTG | 199                   |
| Rgcc-BS-R3 | CTACCAACCTCTACCTACCTACATC       |                           |
| Rgcc-BS-F4 | GGGTYGGAGATTAYGAGATTGTTATTGAGG    | 200                       |
| Rgcc-BS-R4 | CCCCCTCAACGGCAAACAC             |                           |
| Rgcc-BS-F5 | GTAGGGTTAGGAGGATTTTTATAGTAAGAAGGTTG | 207                   |
| Rgcc-BS-R5 | CCCCCTCAACGGCAAACAC             |                           |
| Rgcc-BS-F6 | GAGGTTGGATAGGAGGATTTTTATAGTAAGAAGGTTG | 234                   |
Results & Discussion

Overall, relative mRNA quantity for the targeted gene, and mRNA for LacZ, was less in the Silenced compared with the Expressed group (Fig 2). In the Silenced mutants (*Lyplal1*, *Rab32*, *Rgcc*), the targeted gene mRNA and LacZ mRNA was 0.6 ± 1.21% and 10.4 ± 11.6% of Wild Type Control native gene, whereas for the Expressed group of mutants (*Ninj1*, *Dstn*, *Arap1*) mRNA for the targeted gene and mRNA for LacZ were 32.9 ± 20.5% and 53.4 ± 43.9% of controls (all data are presented as mean ± standard deviation).

In wild type control animals, methylation of the CpG islands found in the promoters of genes characterized in both Silenced and Expressed groups was very low, 0.71 ± 0.25% and 0.66 ± 0.44% respectively. However, in the mutant Silenced group the average promoter methylation was 3.98 ± 3.21%, but promoter methylation remained very low in the Expressed group of mutants at 0.85 ± 0.67% (Fig 3).

The methylation data at each CpG site within each CGI was analyzed using a non-parametric test (Fisher Exact test). Since the average number of reads per amplicon was 37,000, this statistical method had the power to score as significant ($p < 10^{-6}$) even very small differences in methylation. When the Cochran-Mantel-Haenszel test was used to compare Mutant versus Wild Type control there was a significant difference at $p < 0.0001$ for each island in every tissue. However, it was clear from inspection of the data (Fig 3) that there were differences in the overall percent methylation in the Silenced compared with the expressed group.

The methylation and gene expression data for each group were averaged across mutants and tissues and compared by t-test and the data are presented in Fig 4. There were significant differences between the Silenced and Expressed group for the targeted gene mRNA abundance ($p < 0.0001$), for LacZ mRNA ($p < 0.01$), and significant differences for the percent methylation of the promoters ($p < 0.0001$).

These data support the hypothesis that in a subset of targeted mutants carrying this allele there is a reduced expression of the reporter gene and the increased methylation of the promoters of the Silenced group supports the conclusion that this was due to promoter silencing. Not only was LacZ mRNA lower, mRNA for the targeted gene was significantly lower in Silenced compared with Expressed mutant tissue (Fig 4). The presence of mRNA for the gene-trap alleles is not surprising and likely represents splicing between the neo or LacZ reporter.
sequence and 3’ exons of the targeted gene, or leakiness of the gene-trap allele [25–28]. Of interest is that the quantity of targeted gene mRNA assessed by qRT-PCR was significantly lower in the Silenced compared with Expressed mutants.

As reviewed by Deaton & Bird [29] and by Illingworth and Bird [30], CpG islands (CGI) are regions of DNA rich in CpG sequences compared with the rest of the genome. CGIs are found in the promoter regions of ~70% of all genes. Generally, the cytosines are not methylated in these CGIs in either expressed or inactive genes, while there is a high percentage of CpGs methylated outside of CGIs. It is not known what protects promoter CGIs from methylation of the CpG cytosines but likely histone modifications and transcription factor binding are protective. However, when methylation of CpGs does occur within CGIs this is strongly correlated with the silencing of transcription. Methylation of CpGs may not be the initiating event of silencing but instead may mark chromatin modifications with cytosine methylation locking in and stabilizing the silenced state. CpG methylation is of significant importance during development [31], for X inactivation [32], imprinting [33], and is observed in silenced tumor suppressor genes in cancer tissue [34]. As noted in the introduction, CpG methylation has also been associated with the silencing of transgene promoters.
We report here a small but statistically significant increase in CGI methylation in the promoters of targeted genes that did not have LacZ staining. This promoter methylation was correlated with LacZ methylation in the Silenced group ($R^2 = 0.74$, $p < 0.013$; Fig 5). However, in the Silenced group there was only a weak and non-significant correlation of promoter methylation.
methylation and LacZ mRNA ($R^2 = 0.22$; Fig 5). This suggests that there are certainly other factors, in addition to promoter methylation, that are contributing to the silencing of gene expression in the Silenced group. As would be expected since the overall promoter methylation was uniformly very low in the Expressed group of mutants and tissues, there was no correlation between CpG% methylation and LacZ gene expression was negative as expected, this did not reach statistical significance.

Although the amount of overall methylation of the promoters in the Silenced group was elevated it was still quite low. However, the pattern of methylation may be the important factor in driving silencing and not the overall percent methylation at a CGI. There is precedent that
methylation of specific CpG sites in promoter regions correlates with silencing. For example, Furst et al. [35] have reported that methylation status of a single CpG site in the promoter for the estrogen receptor alpha gene is correlated with transcriptional silencing. In a genome wide survey, Medvedeva [36] found that methylation of ~16% of CpGs in CGIs near the transcriptional start site were correlated with gene repression. However, these CpGs were generally not found at transcription factor binding sites. Of the three CGIs in the promoters of the Silenced group, only one had a uniform increase in CGI methylation across all the CpG sites in the mutant (Lyplal1). The other two CGIs had increased methylation in only a subset of CpGs. The patterns of CpG methylation for each of the silenced promoters in representative tissues are presented in S3 Fig.

LacZ was highly methylated for all of the mutant alleles and tissues (Fig 6). The percent methylation of LacZ CpGs in the Silenced group was 21.9 ± 18.5%, whereas the percent methylation in the Expressed group was 45.4 ± 18.3%. This was significantly different by t-test (p < 0.008). Gene bodies (exons and introns) are generally CpG poor relative to the rest of the genome, but the CpGs in gene bodies are highly methylated although the functional consequences of gene body methylation are not understood [37]. Gene body methylation does not block transcription or elongation [37]. In fact, there are reports that methylation of gene bodies is correlated with high gene expression on the active X chromosome [38]. Aran et al. [39] demonstrated that actively expressed gene bodies are hypermethylated compared with flanking sequences and compared with gene bodies that are not expressed. Whereas, Jjingo et al. [40] showed that gene bodies with the highest level of methylation are the genes with mid-level of expression, and genes with low and highest levels of expression have low methylation levels. Therefore, the relationship between exon/intron methylation and gene expression is not a simple one. One possibility is that the differential LacZ coding sequence methylation in the Expressed and Silenced group also contributed to the differential reporter gene expression in these two groups, with the higher methylation of LacZ coding sequence in the Expressed group leading to higher transcription. This may be one additional factor, along with promoter silencing, responsible for regulating reporter gene expression in this system but further work will be needed to determine this.

This study does not definitively prove that the presence of the LacZ reporter with a high CpG content is causal for the silencing of the targeted promoter in the Silenced mutants as would be suggested by the data from Chevalier-Mariette [12]. LacZ sequence has a higher GC content and more CpGs than most of the mammalian genome, containing 3061 nucleotides, a GC content of 56.3% with an observed CpG over expected CpG ratio of 1.19. The presence of unprotected CpGs in the LacZ coding sequence may recruit methylation factors which may extend their methylation activity to nearby promoter CGIs. However, these mutants also carried a Neo selection cassette which is high in GC% and CpG content. Neo coding sequence length is 794nts, the GC content is 59.9%, and the observed-to-expected ratio of CpGs is 1.03. A number of factors could be at play in determining if a promoter is silenced with this targeting vector, including the presence of the exogenous high CpG content DNA as well as unique characteristics of the targeted gene sequence and local chromatin environment. It is likely not simply the presence of the LacZ coding sequence in a permissive environment producing silencing.

We looked for unique features of the Silenced vs. the Expressed gene structure/organization, vector insertion, and histone modifications in mouse genome screens that might explain why the CSD targeting vector insertion results in silencing in some genes but not others. All of the genes targeted in this study contain multiple exons, and the size and location of the promoter CGI is similar among the mutants. The CGIs were of similar length, GC content, and ratio of observed to expected CpGs. The vector insertion site tended to be closer to the CGIs in the
Silenced compared with the Expressed group, and two of the vector targeting arms overlapped with the CGIs in the Silenced group while none of the Expressed group vectors had CGI homology arm overlap. Overlapping vector arms could have resulted in disruption of the CGIs, but we found no differences in CGI sequence compared with the reference genome. We also found that two of the promoters in the Silenced group, and none of the promoters in the Expressed group, were associated with H3K27 histone trimethylation, a marker of Polycomb silencing [41]. These studies were done in wild type mouse tissues and the H3K27 methylation may indicate that these specific genes are predisposed to silencing. Other than these findings, we found no other differences in gene structure, vector characteristics, or histone marks that might explain differences in silencing between the Silenced and Expressed groups of mutants. However, the sample used is too small to make generalizations. Future studies with larger sets of genes, may reveal structural elements correlated with promoter silencing.

The focus of this report is the promoter silencing of the targeted gene in these mutants. However, the possibility exists that the targeting vector may influence the expression of genes in close proximity. This had been reported in other targeted mutant mouse models, either due to disruption of intragenic regulatory elements [42–43], or the presence of the heterologous promoter and neo sequence [44–46]. In a recent report in a targeted Slc25a21 mutation using the same vector as described in the present report [47], they observed reduced expression of a nearby gene. Using an allelic series including the intact vector, and a vector with the neo cassette removed but retaining the LacZ reporter, they demonstrated that the presence of the Neo cassette was responsible for producing the phenotypes of dental and craniofacial abnormalities along with otitis media and hearing impairment. The presence of the Neo decreased the expression of the 3’ Pax9 gene, and a previous publication of a mutation in mouse Pax9 recapitulated some of the same phenotypes. Examining the GEO mouse expression databases revealed expression in kidney, liver, spleen and other tissues for mouse Slc25a21. However, the Slc25a1 adult mutant has no staining for the LacZ reporter except in the testis (See: https://www.sanger.ac.uk/mouseportal/), which could be nonspecific or ectopic. Therefore, the down
regulation of the neighboring Pax9 gene in the Slc25a21 mutant may also be associated with down regulation of the targeted promoter.

Conclusions

In a subset of targeted gene trap mouse mutants, we have observed the apparent silencing of the targeted gene promoter reflected by reduced LacZ mRNA from the reporter gene. In this subset of mutants, the degree of LacZ methylation is significantly correlated with CGI methylation, but CGI methylation is only weakly negatively correlated with LacZ mRNA levels. The data support the hypothesis that presence of the exogenous DNA in the targeting vector, interacting with local chromatin environment, may lead to promoter silencing of the target and that this silencing is marked by CpG methylation. These findings emphasize the need to consider the local effects of targeting vectors on reporter gene expression, and possibly local effects on neighboring genes. Although we believe that local promoter silencing is a relatively rare event in mutants carrying this allele, additional work will be required to define the frequency of these events, and also to understand which features of the targeted gene environment, interacting with the vector, promote silencing.

Supporting Information

S1 Fig. Visual representation of CGI’s of Expressing and Silenced groups. CpG sites in each cartoon are represented by vertical lines. The number of nucleotides covered by the images is indicated in the upper right corner of each CpG plot. Flanking the actual CpG islands in these plots are 200nts not included in the island definition.

S2 Fig. qRT-PCR Probe Efficiency for Actb, LacZ, Arap1, Dstn, Ninj, Rgcc, Lyplal1 and Rab32. Each plot point is the Ct value obtained using serious dilution of target cDNA. Efficiency values were measured using the Ct slope method constructing a plot of Ct vs. log cDNA dilution factor.

S3 Fig. CpG Methylation Wild Type vs. Mutant. Methylation percent representation of individual CpG sites in the promoter regions of Silenced group.

S1 Table. Mutants, Tissues, LacZ Staining, and Gene Expression. Columns under BioGPS and GEO indicate fluorescent intensity of the signal from probes for the specific gene transcripts. Numbers separated by slashes indicate values for different probe-sets, hyphenated numbers indicate range of values from different probe-sets.

S2 Table. Total RNA quality. Quality evaluation of pooled biological replicates of RNA samples with BioAnalyzer prior to reverse transcription to cDNA.

S3 Table. Conditions for DNAse Treatment

S4 Table. Reverse Transcription. Reagents, their amounts and temperature cycling conditions used for reverse transcription.
S5 Table. qPCR Reaction Conditions. Reagents and their amounts for qRT-PCR reaction.

(DOCX)

S6 Table. qRT-PCR probe design. Information for number of exons for all the genes investigated, NCBI accession numbers, in-silico specificity, primer/probe annealing location, length of an amplicon, and splice variants targeted.

(DOCX)

S7 Table. Table of Primer/Probe Efficiencies. Efficiency values that were used to correct relative expression for genes of interest. Efficiencies were determined using the Ct slope method for which a plot of Ct vs. log cDNA dilution factor was constructed.

(DOCX)

S8 Table. Back Extraction Buffer. Reagents and their amounts for DNA isolation from interphase and organic phase left after RNA isolation.

(DOCX)

Acknowledgments

We thank Dario Boffelli for assistance analyzing the bisulfite treated DNA sequencing results, the UC Davis Mouse Biology Program staff for mouse production, genotyping and animal husbandry, and Ginny Gildengorin for assisting with the statistical analysis.

Author Contributions

Conceived and designed the experiments: DBW. Performed the experiments: JVK AJN MA. Analyzed the data: JVK EKE. Wrote the paper: JVK DBW. Consulted on primer design: BW. Harvested tissue from mouse: AC. Consulted on bioinformatics analysis: EKE. Reviewed manuscript: KCKL PdJ.

References

1. Rosser JM, An W. Repeat-induced gene silencing of L1 transgenes is correlated with differential promoter methylation. Gene. 2010 May 15; 456(1–2):15–23. doi: 10.1016/j.gene.2010.02.005 PMID: 20167267

2. Whitelaw E, Martin DI. Retrotransposons as epigenetic mediators of phenotypic variation in mammals. Nat Genet. 2001 Apr; 27(4):361–5. PMID: 11279513

3. Garrick D, Fiering S, Martin DI, Whitelaw E. Repeat-induced gene silencing in mammals. Nat Genet. 1998 Jan; 18(1):56–9. PMID: 9429901

4. Herbst F, Ball C, Tuorto F, Nowrouz A, Wang W, Zavidij O, et al. Extensive Methylation of Promoter Sequences Silences Lentiviral Transgene Expression During Stem Cell Differentiation In Vivo. Mol Ther. 2012 May; 20(S):1014–21. doi: 10.1038/mt.2012.46 PMID: 22434137

5. Tajima S, Shinohara K, Fukumoto M, Zaitsu R, Miyagawa J, Hino S, et al. Ars insulator identified in sea urchin possesses an activity to ensure the transgene expression in mouse cells. J Biochem. 2006 Apr; 139(4):705–14. PMID: 16672271

6. Ciavatta D, Kalantry S, Magnuson T, Smithies O. A DNA insulator prevents repression of a targeted X-linked transgene but not it’s random or imprinted X inactivation. Proc Natl Acad Sci U S A. 2006 Jun 27; 103(26): 9958–9963. PMID: 16777957

7. Soriano P. Generalized lacZ expression with the ROSA26 Cre reporter strain. Nat Genet. 1999 Jan; 21(1):70–1. PMID: 9916792

8. Srivivas S, Watanabe T, Lin CS, William CM, Tanabe Y, Jessell TM, et al. Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. BMC Dev Biol. 2001; 1: 4. PMID: 11299042

9. Skarnes W, Rosen B, West A, Koutsourakis M, Bushell W, Iyer V, et al. A conditional knockout resource for the genome-wide study of mouse gene function. Nature. 2011 Jun 15; 474(7351):337–42. doi: 10.1038/nature10163 PMID: 21677750
10. Birling MC, Gofflot F, Warot X. Site-specific recombinases for manipulation of the mouse genome. Methods Mol Biol. 2009; 561:245–63. doi: 10.1007/978-1-60327-019-9_16 PMID: 19504076

11. Hadjantonakis AK, Pirity M, Nagy A. Cre recombinase mediated alterations of the mouse genome using embryonic stem cells. Methods Mol Biol. 2008; 461: 111–132. doi:10.1007/978-1-60327-483-8_8 PMID: 19030793

12. Chevalier-Mariette C, Henry I, Montfort L, Capgras S, Forlani S, Muschler J et al. CpG content affects gene silencing in mice: evidence from novel transgenes. Genome Biol. 2003; 4(9):R53. PMID: 12952532

13. Gates H, Mallon AM, Brown SD; EUMODIC Consortium. High-throughput mouse phenotyping. Methods 2011 Apr; 53(4):394–404. doi:10.1016/j.ymeth.2010.12.017 PMID: 21185382

14. Ayadi A, Birling MC, Bottomley J, Bussell J, Fuchs H, Fray M, et al. Mouse large-scale phenotyping initiatives: overview of the European Mouse Disease Clinic (EUMODIC) and of the Wellcome Trust Sanger Institute Mouse Genetics Project. Mamm Genome. 2012 Oct; 23(9–10): 600–610. doi: 10.1007/s00335-012-9418-y PMID: 22961258

15. Lloyd KC. A knockout mouse resource for the biomedical research community. Ann N Y Acad Sci. 2011 Dec; 1245:24–6. doi: 10.1111/j.1749-6632.2011.06311.x PMID: 22211970

16. West DB, Pasumarthi R, Baridon B, Djan E, Trainor A, Griffey S, et al. A LacZ reporter gene expression atlas for 313 adult KOMP mutant mouse line. In press: Genome Res. 2015 Jan 15.

17. Ryder E, Gleeson D, Sethi D, Vyas S, Miklejewska E, Dalpi V, et al. Molecular characterization of mutant mouse strains generated from the EUCOMM/KOMP- CSD ES cell resource. Mamm Genome. 2013 Aug; 24(7–8):286–94. doi:10.1007/s00335-013-9467-x PMID: 23912999

18. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 2001 Dec; 25(4):402–8. PMID:11846609

19. Krueger F, Andrews SR. Focused, high accuracy 5-methylcytosine quantitation with base resolution by benchtop next-generation sequencing. Epigenetics Chromatin. 2013 Oct 11; 6(1):33. doi: 10.1186/1756-8935-6-33PMID: 24279302

20. Langmead B, Salzberg SL: Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Mar 4; 9(4):357–9. doi: 10.1038/nmeth.1923 PMID: 22386286

21. Voss AK, Thomas T, Gruss P. Efficiency Assessment of the Gene Trap Approach. Dev Dyn. 1998 Jun; 212(2):171–80. PMID: 9629493

22. Voss AK, Thomas T, Gruss P. Compensation for a gene trap mutation in the murine microtubule-associated protein 4 locus by alternative polyadenylation and alternative splicing. Dev Dyn. 1998 Jun; 212(2):258–66. PMID: 9629500

23. Illingworth RS, Bird AP. CpG islands and the regulation of transcription. Genes Dev. 2001 May 15; 25(10):1010–22. doi: 10.1101/gad.2037511 PMID: 21576282

24. Deaton AM, Bird A. CpG islands and the regulation of transcription. Genes Dev. 2011 May 15; 25(10):1010–22. doi: 10.1101/gad.2037511 PMID: 21576282

25. Illingworth RS, Bird AP. CpG islands—a rough guide’. FEBS Lett. 2009 Jun 5; 583(11):1713–20. doi: 10.1016/j.febslet.2009.04.012 PMID: 19376112

26. Borgel J, Guibert S, Li Y, Chiba H, Schübeler D, Sasaki H, et al. Targets and dynamics of promoter DNA methylation during early mouse development. Nat Genet. 2010 Dec; 42(12):1093–100. doi: 10.1038/ng.708 PMID: 21057502
32. Cotton AM, Price EM, Jones MJ, Balaton BP, Kobot MS, Brown CJ. Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation. Hum Mol Genet. 2015 Mar; 24(6): 1528–39. doi: 10.1093/hmg/ddu564 PMID: 25381334
33. MacDonald WA, Mann MR. Epigenetic regulation of genomic imprinting from germ line to preimplantation. Mol Reprod Dev. 2014 Feb; 81(2):126–40. doi: 10.1002/mrd.22220 PMID: 23893518
34. Jones PA, Baylin SB. The Epigenomics of Cancer. Cell. 2007 Feb 23; 128(4):683–92. PMID: 17320506
35. Fürst RW, Klem H, Meyer HH, Ulbrich SE. A differentially methylated single CpG-site is correlated with estrogen receptor alpha transcription. J Steroid Biochem Mol Biol. 2012 May; 130(1–2):96–104. doi: 10.1016/j.jsbmb.2012.01.009 PMID: 22342840
36. Medvedeva YA, Khamis AM, Kulakovskiy IV, Ba-Alawi W, Bhuyan MS, Kawaji H, et al. Effects of cytosine methylation on transcription factor binding sites. BMC Genomics. 2014 Mar 26; 15:119. doi: 10.1186/1471-2164-15-119 PMID: 24669864
37. Cotton AM, Price EM, Jones MJ, Balaton BP, Kobot MS, Brown CJ. Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation. Hum Mol Genet. 2015 Mar; 24(6): 1528–39. doi: 10.1093/hmg/ddu564 PMID: 25381334
33. MacDonald WA, Mann MR. Epigenetic regulation of genomic imprinting from germ line to preimplantation. Mol Reprod Dev. 2014 Feb; 81(2):126–40. doi: 10.1002/mrd.22220 PMID: 23893518
34. Jones PA, Baylin SB. The Epigenomics of Cancer. Cell. 2007 Feb 23; 128(4):683–92. PMID: 17320506
35. Fürst RW, Klem H, Meyer HH, Ulbrich SE. A differentially methylated single CpG-site is correlated with estrogen receptor alpha transcription. J Steroid Biochem Mol Biol. 2012 May; 130(1–2):96–104. doi: 10.1016/j.jsbmb.2012.01.009 PMID: 22342840
36. Medvedeva YA, Khamis AM, Kulakovskiy IV, Ba-Alawi W, Bhuyan MS, Kawaji H, et al. Effects of cytosine methylation on transcription factor binding sites. BMC Genomics. 2014 Mar 26; 15:119. doi: 10.1186/1471-2164-15-119 PMID: 24669864
37. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012 May 29; 13(7):484–92. doi: 10.1038/nrg3230 PMID: 22641018
38. Hellman A, Chess A. Gene body-specific methylation on the active X chromosome. Science. 2007 Feb 23; 315(5815):1141–3. PMID: 17322062
39. Aran D, Toporoff G, Rosenberg M, Hellman A. Replication timing-related and gene body-specific methylation of active human genes. Hum Mol Genet. 2011 Feb 15; 20(4):670–80. doi: 10.1093/hmg/ddq513 PMID: 21112978
40. Jingo D, Conley AB, Yi SV, Lunyak VV, Jordan IK. On the presence and role of human gene-body DNA methylation. Oncotarget. 2012 Apr; 3(4):462–74. PMID: 22577155
41. Banski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. Cell. 2007 May 18; 129(4):823–37. PMID: 17321214
42. Lettice LA, Horikoshi T, Heaney SJ, van Baren MJ, van der Linde HC, Breedveld GJ, et al. Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. Proc Natl Acad Sci U S A. 2002 May 28; 99(11):7548–53. PMID: 12032320
43. Zuniga A, Michos O, Spitz F, Haramis AP, Panman L, Galli A, et al. Mouse limb deformity mutations disrupt a global control region within the large regulatory landscape required for Gremlin expression. Genes Dev. 2004 Jul 1; 18(13):1553–64. PMID: 15198975
44. Meier ID, Bernreuther C, Tilling T, Neidhardt J, Wong YW, Schulze C, et al. Short DNA sequences inserted for gene targeting can accidentally interfere with off-target gene expression. FASEB Journal 2010 Jun; 24(6):1714–24. doi: 10.1096/fj.09-140749 PMID: 20110269
45. Scacheri PC, Crabtree JS, Novotny EA, Garrett-Beal L, Chen A, Edgemon KA, et al. Bidirectional transcriptional activity of PGK-neomycin and unexpected embryonic lethality in heterozygote chimeric knockout mice. Genesis. 2001 Aug; 30(4):259–63. PMID: 11536432
46. Olson EN, Arnold HH, Rigby PW, Wold BJ. Know your neighbors: three phenotypes in null mutants of the myogenic bHLH gene MRF4. Cell. 1996 Apr 5; 85(1):1–4. PMID: 8620528
47. Maguire S, Estabel J, Ingham N, Pearson S, Ryder E, Carragher DM, Walker N; Sanger MGP Slc25a21 Project Team, Bussell J, Chan WI, Keane TM, Adams DJ, Scudamore CL, Lelliott CJ, Ramirez-Solis R, Karp NA, Steel KP, White JK, Gerdin AK. Targeting of Slc25a21 is associated with orofacial defects and otitis media due to interrupted expression of a neighbouring gene. PLoS One. 2014 Mar 18; 9(3):e91907. doi: 10.1371/journal.pone.0091807 PMID: 24642684