Duke University **Dec 06, 2021**
Dpt. of Biomedical Engineering **Yesh Doctor**

Mathematical Evalulation and Modeling of the Potency of CRISPR-Guided Epigenetic Modifiers (CRISPR-GEMs)

Contents

1 Introduction

The recent advent of CRISPR-Cas9 based technologies has done a tremendous amount to mature the field of DNA-based therapies. CRISPR-Cas9's key advantage is the ability target a precise region in the genome, with high accuracy, low cost, and modular capabilities. CRISPRi and CRISPRa use tools including dCas9-KRAB and dCas9-VP64 fusions allow specific transcriptional perturbations to occur with high fidelity. [1, 2, 3] However, there are currently limited resources available to specifically edit the epigenome. Existing strategies employ similar techniques used in transcriptional regulation: fusing epigenomic affecting domains (epieffectors) to the Cas9 system, allowing for targeted histone tail modification. These tools are what we define as CRISPR Guided Epigenetic Modifiers, or CRISPR-GEMs. At present, some of the tools available are dCas9-p300 fusions [4] to up-regulate histone lysine acetylation and dCas9-HDAC3 fusions [5] to down-regulate histone acetylation, among others. Unfortunately, these are not silver bullets; editing is context specific, and effectiveness varies between cell lines, genes, and specific epigenetic marks.[6] There is a need for more CRISPR-GEMs that bridge in some gaps in editing that existing biotechnologies cannot.

We aimed to bridge this gap by creating repressive CRISPR-GEM's that can effectively modulate the epigenome using epieffectors that have been shown to remove acetyl groups (HDAC8, SIRT6) and deposit methyl groups (SETDB2) on histone tails. This semester in particular, the aim was to investigate the kinetics of these CRISPR-GEMs and how their repressive ability on the genes DKK1, MLNR1, and IGFBPL1 change over time.

2 Brief Experiment Description

A preliminary screen (data not shown) of genes that were differentially expressed in a HDAC8-Knockout cell line (compared to WT) highlighted a few candidate genes that were amenable to repression by novel CRISPR-GEMs. We determined that the human genes Motilin Receptor (MLNR), Dickkopf WNT Signaling Pathway Inhibitor 1 (DKK1), and Insulin Like Growth Factor Binding Protein Like 1 (IGFBPL1) were good targets to test the repressive CRISPR-GEMs on. Furthermore, research from literature suggested that histone 3 lysine 27/9 and histone 4 lysine 16 may be highly susceptible to deacetylation, and histone 3 lysine 9 may be highly susceptible to methylation [9]. We decided to focus on these marks (H3K17ac, H3K9ac, H4k16ac, and H3K9me3) for epigenomic analysis. Approximately 350,000 HEK-293T cells (in triplicate) were transfected with each of the CRISPR-GEMs. 0.5, 1, 1.5, 2, 3, 4, and 5 days post transfection, cells were were harvested for RNA extraction and processing into cDNA. qPCR was performed with primers that span the exons of MLNR, DKK1, and IGFBPL1 to quantify mRNA expression changes over time. mRNA fold change was normalized to GAPDH (reference gene) and non treated (NT) samples.

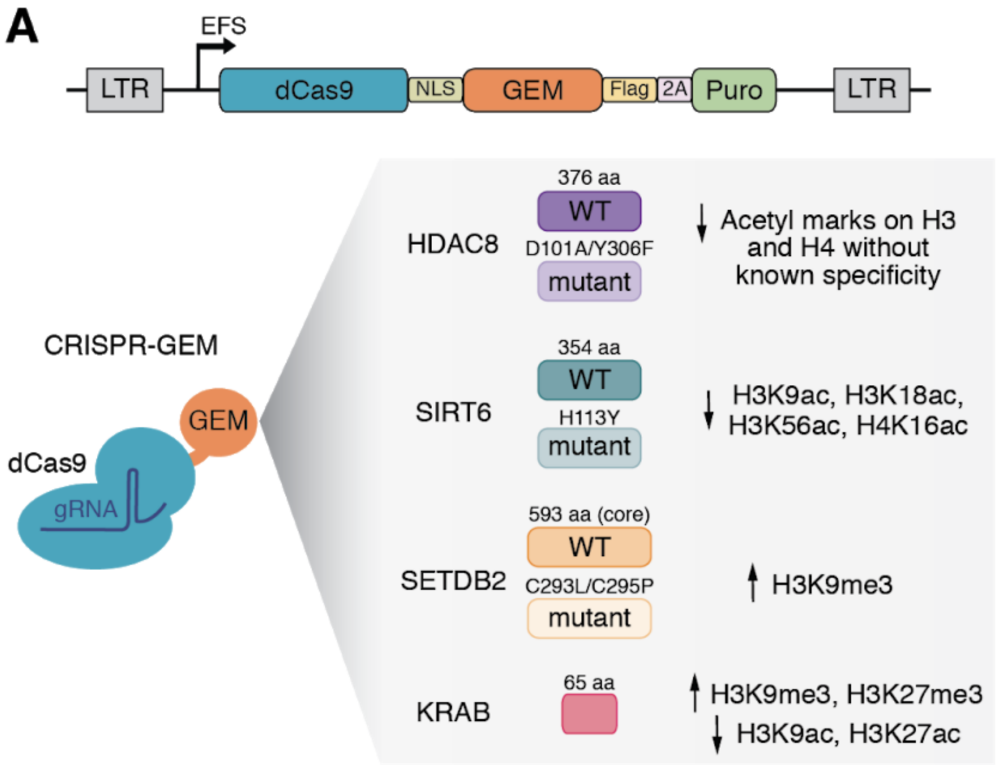

Figure 1: **List of Created CRISPR-GEMs**

3 Methods

Methods adapted from Holtzman et al. (prepublication).

3.1 Cell Culturing and Transfection.

HEK-293T cells (ATCC, Manassas VA, obtained through the Duke University Cell Culture Facility) were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin and maintained at 37C and 5% CO2. Tranfections were performed in 24-well plates using a total amount of 1000 ng plasmid DNA mixed with Lipofectamine 3000 (Life Technologies, L3000008) as per manufacturer's instruction. For gene expression assays, dCas9-effector fusions, gRNA-1 and gRNA-2 plasmids were transfected in a 2:1:1 mass ratio, respectively. Cells were propagated 3 days until harvesting.

3.2 RT-qPCR

RNA was isolated using the RNeasy Plus RNA isolation kit (Qiagen, 74136), and 500-1000 ng of purified RNA were used to synthesize cDNA with the SuperScript VILO cDNA Synthesis Kit (Invitrogen, 11754250). Real-time PCR was performed by either using PerfeCTa SYBR GreenFastMix (Quanta Bio-

sciences, 95072-012) with the CFX96 RealTime PCR Detection System (Bio-Rad). The results are expressed as fold change expression of the gene of interest normalized to GAPDH expression using the ΔΔCt method.

4 Experimental Results

4.1 mRNA Fold Change

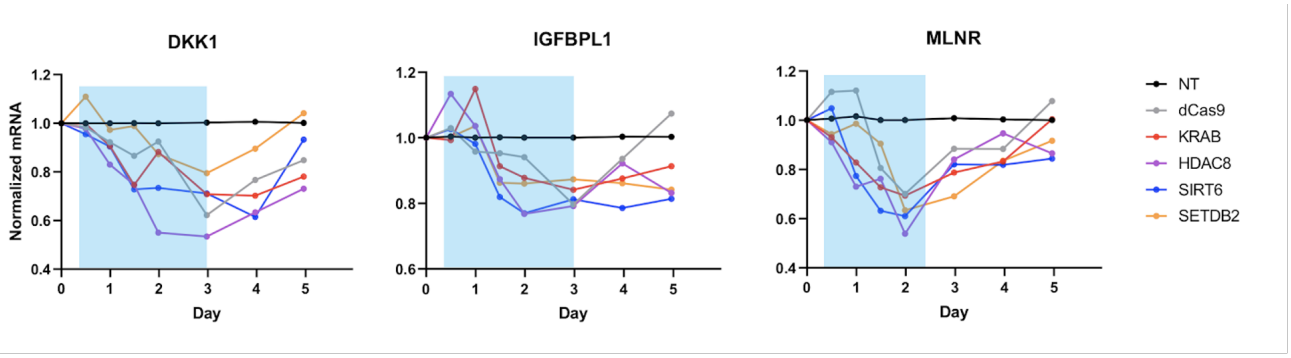

Figure 2: **Time Dependent Changes in Gene Expression**: Mean mRNA levels of DKK1, IGFBPL1 and MLNR in the presence of various CRISPR-GEMs at multiple timepoints (day 0.5, 1, 1.5, 2, 3, 4 and 5; N=4, on day 2 and 4, N=2 for all other timepoints). Blue rectangles denote day 0.5-day 3, which were used for logarithmic fitting.

In DKK1, all epieffectors tend to follow a logistic decay followed by logistic growth after day 3. HDAC8, SIRT6 both display lower expression than KRAB from days 0.5 - 3, and all three display lower expression than dCas9 up to day 3. In IGFBPL, all epieffectors tend to follow a logistic decay followed by logistic growth after day 3, with a spike in most epieffectors above non treated levels on day 0.5. Once again, HDAC8, SIRT6 both display lower expression than KRAB from days 0.5 - 3, and all three display lower expression than dCas9 up to day 3. SETDB2 elicits similar IFGBPL1 expression to KRAB. In MLNR, all epieffectors tend to follow a logistic decay followed by logistic growth after day 3, with a spike in most epieffectors above non treated levels on day 0.5. Once again, HDAC8, SIRT6 both display lower expression than KRAB from days 0.5 - 3, and all three display lower expression than dCas9 up to day 3. SETDB2 elicits higher MLNR expression than KRAB, HDAC8, SIRT6 up to day 3, where it it demonstrates lower expression than KRAB.

4.2 CRISPR-GEM Potency Evaluation

Based on the psuedo-logarithmic decay observed in most of the CRISPR-GEM's between day 0.5 and 3, we fit a logarithmic decay function of the form:

$$[mRNA] = [mRNA]_0 e^{-kt}$$

. This form assumes that the rate of decay in mRNA concentration is proportional to the constant k and the current mRNA concentration;

$$d[mRNA]/dt = -k[mRNA]$$

Therefore, a higher k-value in fitting can be associated with a faster rate of mRNA decay which is associated with a greater level of gene repression and more potent CRISPR-GEM. Based on the k value, we can rank the potency of each CRISPR-GEM.

In figure 3, across all three models, SIRT6 displayed the highest decay constant, indicating that it produced the fastest attenuation in mRNA expression from day 0.5 to 3. In both IGFBPL1 and MLNR, all three designed epieffectors had higher decay constants than KRAB.

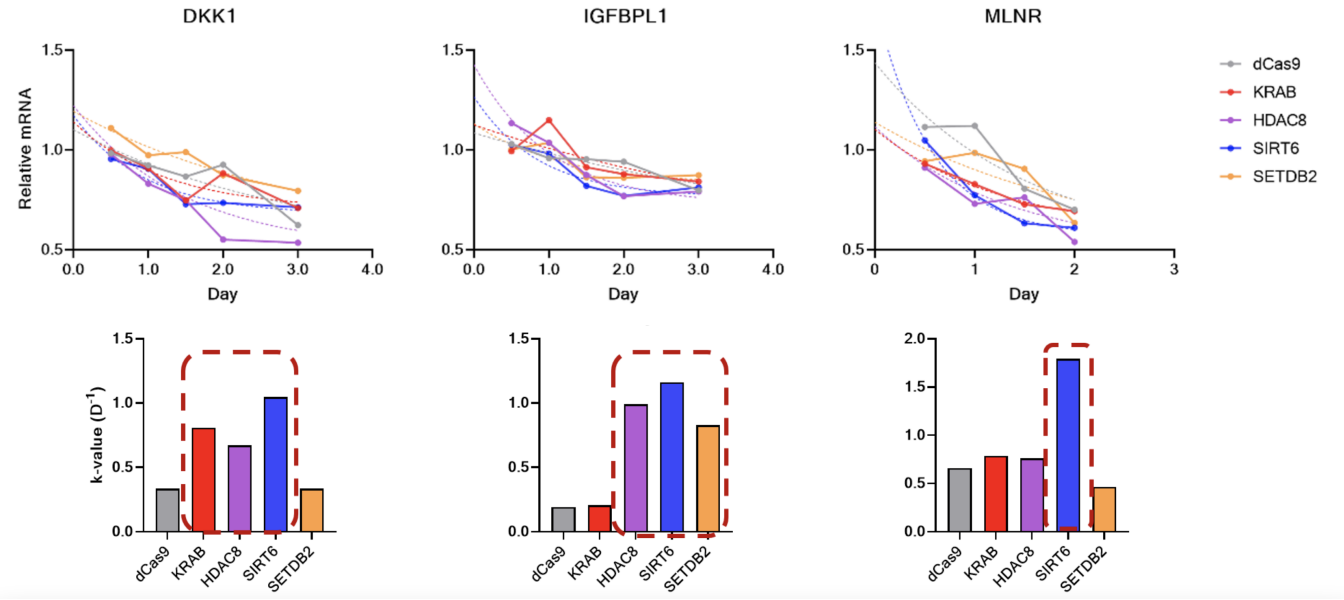

Figure 3: **Logarithmic Fitting to Mean mRNA expression Day 0.5 - Day 3**: The top row represents logarithmic (exponential) fits to the data from day 0.5 to day 3 using the proposed kinetics above for DKK1, IGFBPL1, MLNR. The bottom row shows a graph of the associated k-value for each fit, with a larger k indicating a higher decay constant.

Figure 4 represents correlation between the quantity of actyl/methyl marks deposited (from ChIP-qPCR) and mRNA expression. H3K9me3 deposition correlates well with DKK1 repression at Day 4 and MLNR at Day 10. Additionally, at day 2, H3K9 deacetylation correlates strongly with IGFBPL1 repression.

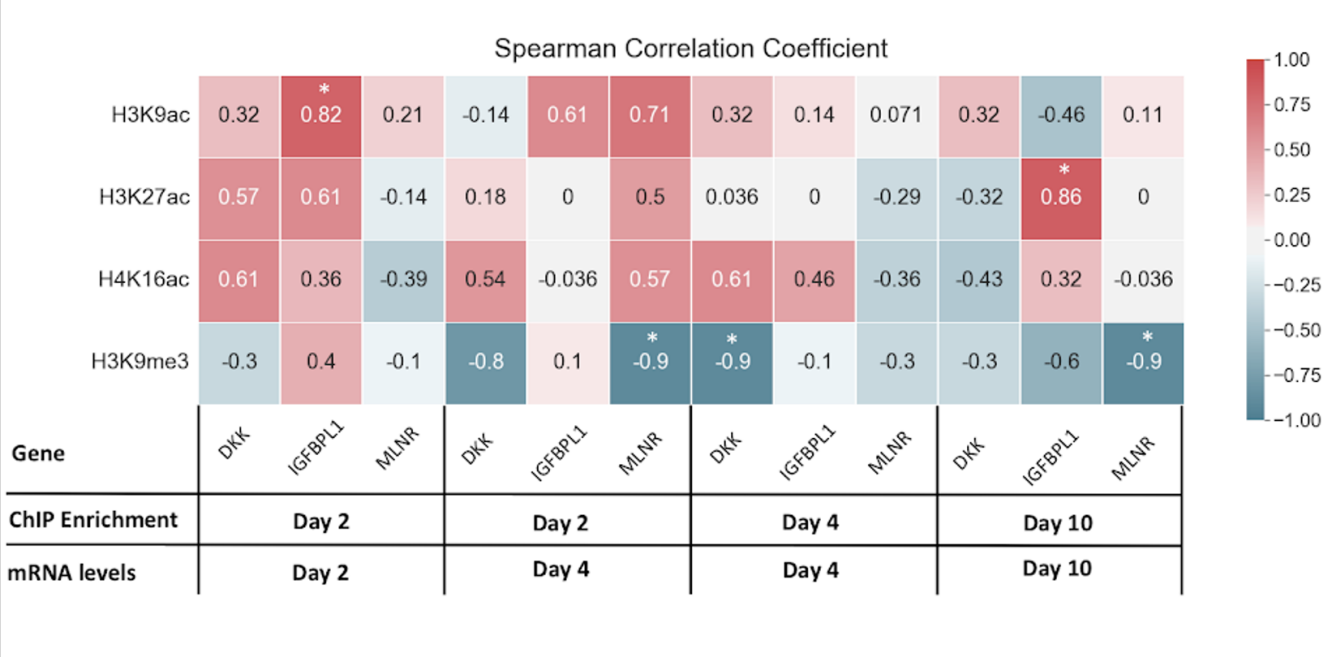

Figure 4: **Correlation Heatmap Between Epigenetic and Expression Changes**: Spearman's rank correlation coefficients (ρ) between fold enrichment of histone mark deposition and relative mRNA levels X days post lentiviral infection with CRISPR-GEM were calculated and plotted on a heatmap (for H3K9ac, H3K27ac and H4K16ac n=7, and for H3K9me3 n=5). Significance of Spearman's rank correlation analysis was calculated by a two-tailed test. $^*P < 0.05$.

5 Model Construction, Assumptions

We propose a mechanism of action for the CRISPR-GEMs by which dCas9 localizes the EpiEffector (HDAC, HMT) to the gRNA programmed site (DKK1, MLNR, IGFBL1) due to physical linkage of the EpiEffector to dCas9. For simplicity, we assume that dCas9 and the bound epieffector are coexpressed effectively (i.e. functional dCas9 implies functional EpiEffector). There is some proportion of the localized epieffector that deposits/removes epigenetic marks from histone lysines. There is a further proportion of these histone mark changes that impact chromatin structure, restricting transcription factor accessibility. Additionally, dCas9 being bound near the programmed site provides steric blockage to transcription factors. These actions can be (for simplicity) lumped into a single parameter that alters the basal transcription rate of the programmed site to a level below the degradation rate, leading to net transcriptional inhibition. This process can be modeled through Hill Kinetics wherein the basal transcription rate (k_{basal}) is adulterated through the Hill parameter (θ), treating the CRISPR-GEM as a Hill Repressor.

5.1 Transient Expression of CRISPR-GEM

Experimentally, each CRISPR-GEM is delivered to the cell via transient transfection. This manifests in a rapid increase in CRISPR-GEM concentration between 0-36 hours post transfection, followed by clearance of the protein via ubiquitination and degradation of the underlying plasmid vector. Through a time course study (measuring both gene target mRNA and dCas9 mRNA levels via qPCR 0.5, 1, 1.5, 2, 2.5, 3, 4, and 5 days post transfection), we observe that this transient process under our experimental conditions can be best estimated by initial logistic growth followed by exponential clearance of the CRISPR-GEM. We model the CRISPR-GEM concentration as a function of time ($E(t)$) as follows:

$$E(t) = \begin{cases} \frac{100.0}{1+e^{-0.95(t-1.5)}} - 20 & t \leq 10 \\ 80e^{\frac{t-10}{1.5}} & t > 10 \end{cases}$$

The initial logistic growth is modeled by a standard logistic function of the form $\frac{C}{1+e^{-k(t-t_{1/2})}}$. We assume a carrying capacity, C of $100 - 20 = 80$ to force the logistic function to cross through the origin ($(0, 0)$). Setting the growth rate, k to 0.95 closely models rapid transcription and translation of transfected epieffector, and including a phase delay $t_{1/2} = 1.5$ models observed EpiEffector saturation between 24 and 36 hours post transfection.

The subsequent exponential decay is modeled by a standard exponential of the form $Ae^{(t-t_0)/t_{1/2}}$. We introduce a time-shift of $t_0 = 10$ to account for the piece wise nature of transient transfection ($t > 10$). Furthermore, we include a rapid decay half-life ($t_{1/2} = 1.5$) translating to 3.6 hours. We use $A = 80$ to preserve the same initial EpiEffector concentration at $t = 10$ between the end of logistic growth and start of exponential decay portions.

5.2 CRISPR-GEM Regulation of mRNA Expression

As described previously, we use Hill kinetics to model transcriptional repression of the target sequence. The mRNA transcription rate is modeled as follows (note that $[E]$ represents CRISPR-GEM concentration and $[mRNA]$ represents mRNA concentration):

$$\frac{d[mRNA]}{dt} = k_{basal} * \theta - k_{deg} * [mRNA]$$

$$\theta = \frac{K_d^n}{K_d^n + [E]^n}$$

In the case that there is no CRISPR-GEM present (i.e. $\theta = 1$) we assume zero order generation and first-order decay of mRNA. The zero order generation assumption arises from knowledge of the functionality of DKK1, IGFBPL1, MLNR which are not known to be transcription regulating proteins (no positive feedback). The first order decay assumption rises from observed homeostasis of DKK1, IGFBL1, MLNR mRNA levels compared to a control gene (GAPDH) in the control (non-treated group) that is known to

fluctuate in concentration (increase and return to basal level) over the cell cycle, suggesting typical mammalian gene behavior [10]. Chen et al. [11] describe mRNA degradation in mammalian cells as containing an initial lag phase followed by first-order decay. We make the simplifying assumption that the lag phase is sufficiently short in the context of the conducted 5-day experiment, and therefore model mRNA turnover as first-order.

The Hill parameter θ represents a value between 0 and 1 representing how the CRISPR-GEMs attenuate the basal production rate (k_{basal}) of mRNA. The hypothesized mechanism of action of HDAC/HMT is through removing acetylation marks on H4K16 and depositing trimethylation marks on H3K27, HDAC/HMT cause chromatin constriction, inhibiting transcription by transcription factors. Traditional transcriptional repressors work in a similar way; providing steric hinderance to RNA polymerase II; we model CRISPR-GEMs through the same Hill equation used to model transcriptional repressors.

K_d represents the half-repression threshold; for $[E] = K_d$, $\Theta = 0.5$, and therefore $k_{effective} = 0.5 * k_{basal}$. K_d is one measure of CRISPR-GEM potency, as the lower K_d is, the lower the required concentration for half-repression. n is a measure of CRISPR-GEM cooperativity; for $n > 1$ this can be interpreted as each CRISPR-GEM facilitates the action $n - 1$ other CRISPR-GEMs. There does not exist evidence of di/polymerization of HDAC8, SIRT6, SETDB2, so a n value closer to 1 is expected.

6 Model Results

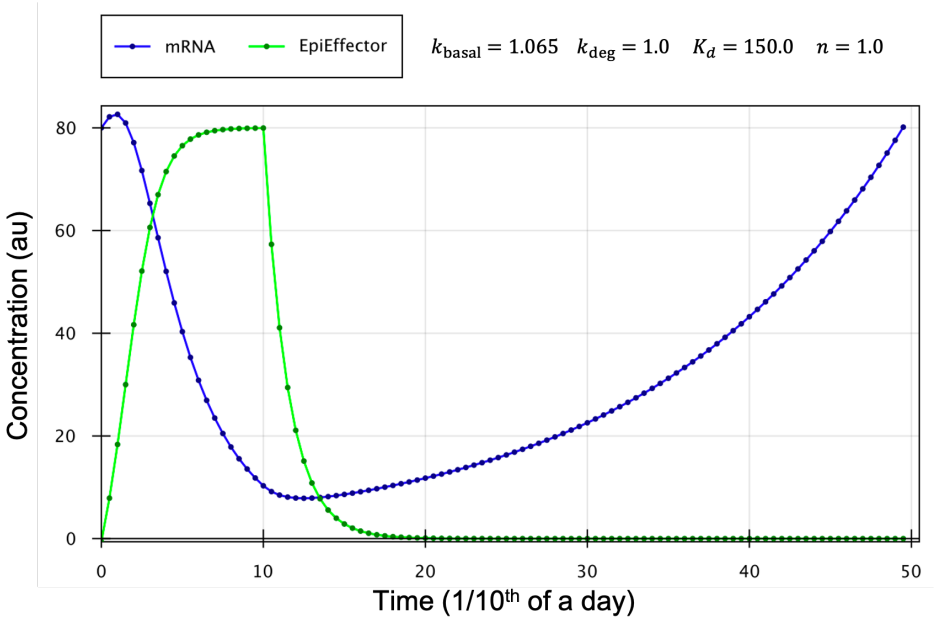

Figure 5: **Model prediction closely follows experimentally derived data.**

We implement the model described above in Dynetica and simulate CRISPR-GEM and mRNA expres-

sion using the Runga-Kutta-Fehlberg algorithm for 5000 points (corresponding to a simulation time of 5 days). The transient piece wise equations create a pulse in epieffector concentration that plateaus 24 hours post transfection ($t = 0$) and exponentially decays back to 0 within 48 hours post transfection. This is consistent with experimentally measured dCas9 concentrations between day 0 and 2 (data not shown). In turn, the mRNA concentration responds to the increase in CRISPR-GEM as the concentration drops from 80 (arbitrary units) at basal to its minimal value (10 au) at Day 1.3 in a psuedo-logarithmic manner similar to experimentally observed results (Fig 2). Full recovery of basal mRNA concentration is achieved 5 days post-transfection which also follows experimental results from Fig 2. The parameters used to create these results are: $k_{basal} = 1.065, k_{deg} = 1.0, K_d = 150.0, n = 1.0$. $n = 1.0$ can be interpreted as CRISPR-GEMs do not act in a cooperative manner, which is to be expected as epigenetic mark deposition from testing in previous semesters (data not shown) with CRISPR-GEMs is independent of the number of guide RNAs used (same effect using one or multiple guides). Furthermore, $K_d = 150.0$ implies that CRISPR-GEMs do not reach the half-repression threshold (max concentration of 80.0), yet still provide substantial repression. This is likely due to the first-order turnover of mRNA further accelerating repression.

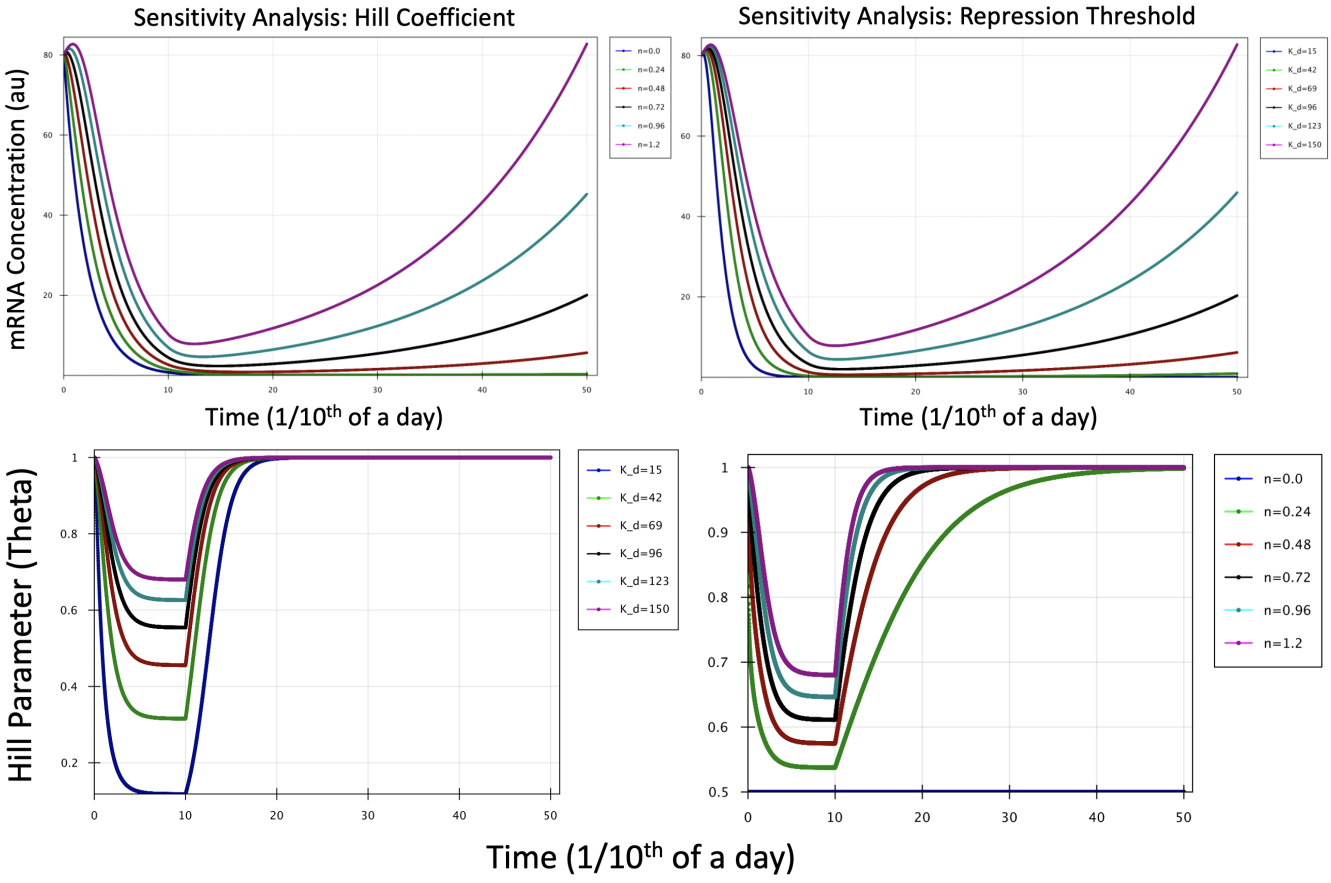

Figure 6: K_d **and** n **dependent modeling of effector potency**

The model allows for flexible parameter turning of both K_d and n, which alter CRISPR-GEM potency.

Decreasing K_d and n models a stronger repressor. These changes manifest in an altered θ parameter. For $K_d = 150.0$ like in the model in Fig 5, the lowest value of θ is 0.71. Decreasing the repression threshold also decreases θ leading to modeling a stronger repressor. A similar trend is seen when decreasing n. Altering n values also models speed of mRNA recovery following repressor clearance.

7 Discussion

7.1 Experimental Results

Figure 2 Demonstrates that there is a tendency for epigenetic changes to manifest relatively quickly, but not be permanent in nature. From Day 0.5 to day 3, the level of mRNA expression decreases in MLNR, IGFBPL1, and DKK1, however for all three genes, after day 3 appears to pick back up. It would be interesting to monitor the expression profile with greater granularity between day 2 and day 3 to understand the criticial transition point between repression and expression recovery. Additionally, figure 4 along with data collected in previous semesters suggests that repression is seen at day 10 as well, which indicates that there may be another cycle where the effects of epigenetic deposition return. Experiments with greater resolution need to be taken between day 5 and day 10 to tease out this effect. It is possible that there is long term oscillatory behaviour, or the initial signal at day 0.5-3 is primarily predominated by steric effects from dCas9 binding and/or cell response to transfection.

Figure 3 attempts to use the k-value of exponential effect to explain the potency of a particular epieffector (pertaining to its repression ability). The results from the curve fitting to indicate that SIRT6 has the greatest decay constant, and therefore leads to mRNA expression downregulation in DKK1, MLNR, and IGFBPL1 at the greatest rate. However, this analysis assumes that the goodness of fit of all of the eponential models are the same, which is not true. Specifically, looking at the fit from DKK1 on HDAC8, its clear to see that the fit line is not very strong. The associated mRNA levels are lower than SIRT6 on days 2 and 3, however the poor fit likely results in a lower k value being reported. Hence, these results are not entirely representative of the rate of mRNA down regulation by the epieffectors. One possible solution could be the log-transformation of the first order rate equation, where

$$ln[mRNA] = -kt + ln[mRNA]_0$$

From this, the natural log of mRNA concentration versus time can be plotted, and a linear fit can be produced, which may more accurately fit the HDAC8 data, producing a more robust analysis. Nevertheless, all active epieffectos (KRAB, SIRT6, HDAC8, SETDB2) appear to have larger k-values and elicit lower mRNA expression from days 0.5 to day 3 than dCas9 in all three genes, suggesting that the epieffector effects are not exclusively steric.

Figure 4 provides evidence that there is a relationship between epigenetic changes and expression changes. Specifically, these effects can be seen strongly in DKK1 and MLNR on day 2, and MLNR on day 10. This suggests that there is potentially some form of causal relationship between epigenetic modifications and

mRNA expression change.

We posit in figure 7 that epigenetic changes precede transcription level changes. At $t = 0$, CRISPR-

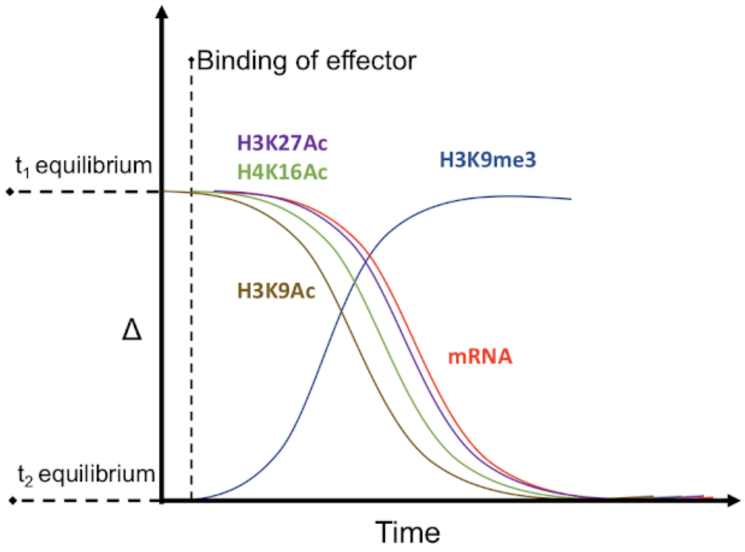

Figure 7: **Hypothesis on Time-Relationship Between Epigenetic Changes and Expression Level Changes**

GEM's are added to cells, and begin to deposit / remove marks from histone tails. This cascades into expression level changes, primarily chromatin constriction, which disables transcriptional machinery from transcribing mRNA. This system will reach a new equilibrium, t_2 These marks however, are degraded / removed by endogenous cell machinery, and hence the mRNA expression will restore to t_1 after some time. In particular, data from figure 3 suggests this recovery will ensue after 3 days.

7.2 Model Limitations, Experimental Interpretation

The developed mathematical model provides a strong framework to evaluate relative CRISPR-GEM potency. Fitting the model to the average dCas9 mRNA (qPCR obtained) for each CRISPR-GEM allows for robust modeling of transient CRISPR-GEM expression. Furthermore, fitting an appropriate K_d and n value to the mRNA degradation and recovery curves like those in Fig 2 gives insight into the potency and cooperativity of each CRISPR-GEM relative to each other and control (dCas9 alone). This is an improvement over fitting a logistic decay model to data from day 0 to day 3 and extracting a k value as the model better accounts for transient transfection and mRNA recovery. However, to produce a proper fitting to the model, more time course data is required with higher granularity to produce meaningful results for specific K_d and n values. We collected data every 12 hours, however the mathematical model we developed suggests that changes can occur on the 3-4 hour scale, suggesting greater granularity in qPCR time course data is required before parameter fittings can be made. Furthermore, stochastic modeling will also likely

need to be included in the model to account for random variations at the cellular level and measurement (qPCR) level. The model also makes a first-order mRNA turnover assumption with constant degradation rate, which is likely overly simplified as there are many cellular regulatory factors that contribute to homeostasis maintenance; degradation rates likely change over the course of the cell cycle. Finally, the model also struggles to segregate repression due to epigenetic mark deposition (EpiEffector related) and sterically driven repression (dCas9 related) as these are both lumped into parameter θ. Nevertheless, the model is able to capture the general experimentally observed mRNA expression trends. This justifies assuming that CRISPR-GEMs act kinetically in a similar way to traditional transcriptional repressors. Furthermore, there is a noticeable lag phase between the start of CRISPR-GEM clearance ($t = 10$) and start of mRNA recovery ($t = 13$) which further bolsters our theory that epigenetic changes preceded transcriptional changes with CRISPR-GEMs.

8 Bibliography

1. Guilinger, J. P.; Thompson, D. B.; Liu, D. R. Fusion of Catalytically Inactive Cas9 to FokI Nuclease Improves the Specificity of Genome Modification. Nat. Biotechnol. 2014, 32, 577–582.

2. Konermann, S.; Brigham, M. D.; Trevino, A. E.; et al. Genome-Scale Transcriptional Activation by an Engi- neered CRISPR-Cas9 Complex. Nature 2015, 517, 583–588.

3. Gilbert, L. A.; Horlbeck, M. A.; Adamson, B.; et al. GenomeScale CRISPR-Mediated Control of Gene Re- pression and Activation. Cell 2014, 159, 647–661

4. Hilton, Isaac B., Anthony M. D'Ippolito, Christopher M. Vockley, Pratiksha I. Thakore, Gregory E. Crawford, Timothy E. Reddy, and Charles A. Gersbach. 2015. "Epigenome Editing by a CRISPR-Cas9-Based Acetyl- transferase Activates Genes from Promoters and Enhancers." Nature Biotechnology 33 (5): 510–17.

5. Kwon, Deborah Y., Ying-Tao Zhao, Janine M. Lamonica, and Zhaolan Zhou. 2017. "Locus-Specific Histone Deacetylation Using a Synthetic CRISPR-Cas9-Based HDAC." Nature Communications 8 (May): 15315.

6. Tadic , Vanja, Goran Josipovic , VlatkaZlotosš, and Aleksandar Vojta. 2019. "CRISPR/Cas9-Based Epigenome Editing: An Overview of dCas9-Based Tools with Special Emphasis on off-Target Activity." Meth- ods 164-165 (July): 109–19.

7. Lavarone, Elisa, Caterina M. Barbieri, and Diego Pasini. 2019. "Dissecting the Role of H3K27 Acetylation and Methylation in PRC2 Mediated Control of Cellular Identity." Nature Communications 10 (1): 1679.

8. Wilson, Stephen, and Fabian Volker Filipp. 2018. "A Network of Epigenomic and Transcriptional Cooperation Encompassing an Epigenomic Master Regulator in Cancer." NPJ Systems Biology and Applications 4 (July): 24.

9. Ernst, Jason, Pouya Kheradpour, Tarjei S. Mikkelsen, Noam Shoresh, Lucas D. Ward, Charles B. Epstein, Xiaolan Zhang, et al. 2011. "Mapping and Analysis of Chromatin State Dynamics in Nine Human Cell Types." Nature 473 (7345): 43–49.

10. Mansur, N. R., K. Meyer-Siegler, J. C. Wurzer, and M. A. Sirover. 1993. "Cell Cycle Regulation of the Glyceraldehyde-3-Phosphate Dehydrogenase/uracil DNA Glycosylase Gene in Normal Human Cells." Nucleic Acids Research 21 (4): 993–98.

11. Chen, Chyi-Ying A., Nader Ezzeddine, and Ann-Bin Shyu. 2008. "Messenger RNA Half-Life Measurements in Mammalian Cells." Methods in Enzymology 448: 335–57.