

Novel Genome Editing With CRISPR-Cas9

Lara Janiszewski

**Honors Thesis
Molecular, Cellular, Developmental Biology
Defense date: April 6, 2015**

Thesis Committee

**Dr. Rui Yi, MCDB (Thesis Advisor)
Dr. Jennifer Martin, MCDB (Honors Council Representative)
Dr. Daniel Medeiros, EBIO**

Abstract

My first project focused on the deletion of the Let7 microRNA (miRNA) family in murine keratinocytes. There are 9 Let7 family members present in these cells, all acting as translational suppressors of the same target mRNAs. This redundancy implies that Let7 is crucial to cellular functioning, and it has indeed been found to act as a tumor suppressor, down regulating functions such as proliferation, cell motility, and de-differentiation. I hypothesized that increasing loss of Let7 loci would lead to progressively more tumor-like phenotypes in the cells. Phenotypic analysis of these cells has not yet been possible, but initial Let7 knockouts were successful.

The amyloid precursor protein (APP) is found on chromosome 21 and is strongly upregulated in individuals with Down Syndrome (DS). This increased protein level is thought to play a role in the early onset development of amyloid-beta plaques in the brains of those with DS, one of the hallmarks of Alzheimer's Disease. APP was therefor chosen as a target for CRISPR, not to knock it out, but to delete one copy and reduce its gene dosage to normal levels. This would allow us to assess protein and RNA levels between trisomic and disomic APP, against a background of trisomy 21. If successful, our work will serve as a proof of principle that this technique can be used to assess the dosage effect for any chromosome 21 gene. To attempt this, we need transient transfection, which is difficult to achieve with plasmids that can last in the cells for weeks. We hypothesized that brief periods of editing, and therefor incomplete knockouts, could be generated by transfecting cells with CRISPR RNAs: Cas9/EGFP mRNA, and two sgRNAs. This was attempted in both lymphoblastoid cells and HEK293T cells, but a knockout has yet to be realized.

Keywords: Down Syndrome, CRISPR-Cas9, Let7, APP, transfection

Introduction

CRISPR-Cas9

Prior to 2013, genome editing was done with custom nucleases designed for specific loci, such as zinc

finger nucleases or transcription activator-like effector nucleases. These are somewhat cumbersome molecules, as they must be custom designed for each site. Gene knockouts were often done more reliably *in vivo*, using homologous recombination within embryos and then selectively breeding until the desired genotype was reached in offspring. A novel approach to gene knockout has recently been adapted from a natural system in *Streptococcus pyogenes*. The bacteria uses Cas proteins, which are RNA guided endonucleases (RGENs), as an intracellular defense against exogenous DNA. It does so by using a DNA construct called CRISPR (clustered, regularly interspaced, short palindromic repeats). Should the cell detect DNA that constitutes a threat, it samples the DNA and inserts pieces of it between the short palindromic repeats in the CRISPR region of its own genome. These inserts are then transcribed, creating RNA fragments that match the invasive DNA with a tail transcribed from a CRISPR repeat. These transcripts are termed CRISPR RNAs (crRNAs), and each can then hybridize with an endogenous RNA termed the transactivating CRISPR RNA (tracrRNA). This piece is structural, forming hairpins, and allows the RNA complex to dock on a Cas endonuclease. The crRNA then acts as a search template, base pairing with its parental exogenous DNA and allowing Cas to cause a double stranded break (DSB). However, in order to be cleaved by Cas, the target DNA sequence must be immediately 5' to a PAM sequence (protospacer adjacent motif) which usually has the form of 5'-NGG. This restriction may serve as a mechanism to prevent cleavage of the bacteria's own genome at CRISPR sites. The DSB is fairly predictable, and is typically created three base pairs 5' of the PAM sequence (Ran et al, 2013).

Researchers have adapted this mechanism to create genomic mutations or knockouts in eukaryotic cells by optimizing the codons in a particular Cas protein – Cas9. In order to edit the eukaryotic genome, cells can be transfected with a plasmid carrying the Cas9 gene and engineered sequences which will produce specifically targeted RNA guides. Post cleavage by Cas9, cellular DNA repair mechanisms are relied upon to complete the editing process. There are two general repair pathways: non homologous end joining (NHEJ) and homology directed repair (HDR). HDR allows an exogenous 'donor' sequence to be provided to the cell and swapped into the genome, causing specific changes in sequence. NHEJ is a repair mechanism that is useful when a disruption of any sort at the DSB site will give the desired effect. For example, NHEJ may leave behind various insertions or deletions (indels), which can serve the purpose of interrupting an exon or altering a promoter region. NHEJ is also what is relied upon when a specific large deletion is desired; if two guide RNAs are introduced to the cell Cas9 will

theoretically cleave both locations, snipping out the desired piece of DNA and allowing NHEJ to bind the two genomic ends back together. If this is successful, the cell now possesses a stable genomic knockout.

Various groups have researched the most efficient means of introducing CRISPR/Cas9 components into cells, either in vitro or in vivo. As discussed by Sander and Joung (2014), the Yi lab has adopted the method of creating single guide RNAs (sgRNAs) which are chimeric molecules of specifically designed crRNA template ligated to the scaffold-like tracrRNA. This creates one continuous transcript, the secondary structure of which allows it to complex with Cas9. The template for one or more sgRNAs can be added to a plasmid containing the Cas9 gene, and all pieces are placed under cell specific promoters. In order to induce the editing desired in my projects, two sgRNAs are introduced at once, which target regions flanking an area to be deleted. Inversions can also result from this process, instead of the desired deletion. This may or may not be problematic, depending on the intended effect of the deletion.

The use of these chimeric RNAs has been shown to be highly efficient (Cho et al, 2013), and editing rates are further enhanced when the entirety of the *S. pyogenes* tracrRNA sequence is used in sgRNAs. (Hsu et al, 2013). However, the efficiency with which crRNAs bind their target sites can vary dramatically; some must be abandoned due to unacceptably low cleavage rates. Some target sites may actually be inaccessible due to epigenetic markers, chromatin density, or other obstructions, and these are difficult to predict when designing constructs. Ran et al (2013) suggests a method of screening cleavage efficiency for multiple genomic targets at once. This entails ordering several complete sgRNA oligos with U6 promoters already on them. These would then be transfected en masse as dsDNA, with Cas9, and target sites would be genotyped to assess cleavage. Once high efficiency sgRNAs were determined they would be inserted into plasmids and cloned for future transfection. Unfortunately, ordering multiple 353 bp oligos (and their reverse complements) would quickly become cost prohibitive.

CRISPR/Cas9 offers efficient genome editing without residual sequences (loxP sites) or the need to breed transgenic animal strains. It also offers the ability to multiplex mutations to see combined effects, and to work directly with cell lines without designing specific enzymes for each cleavage site. Another advantage lies in the way cleavage sites are determined. The use of base pairing to locate sequence

targets is more predictable than the protein-DNA interactions relied upon for other cleavage mechanisms such as zinc finger nucleases. These can be context dependent, which makes possible off-target cleavage cryptic to determine or replicate. CRISPR activity is not without errors, and off-target effects are an issue if binding and cleavage can occur at sites which are mismatches to the crRNA. However, locations of Cas9 off-target effects can be guessed at via basic sequence homology.

Cho et al (2014) suggest that RNA guided endonucleases (RGENs) like Cas9 are inherently imperfect when it comes to off-target effects because they evolved in prokaryotes, with genomes much smaller than those of the eukaryotes we hope to modify. In other words, RGENs aren't better than they have to be for the uses of the bacteria. Concurring with this, Hso et al (2013) note that 1 base pair discrimination is not a level of granularity to be expected by these proteins. Within a 20 bp crRNA the 8 – 12 nucleotides at the 3' end serve as a sort of seed sequence, and typically need to be matched with high fidelity for cleavage to occur. This is not always the case though, and it has been seen that off-target cleavage is affected by a combination of the location and identity of each mismatch between crRNA and host DNA, as well as the total number of mismatches (Hso et al, 2013).

Many groups are working to find ways of managing off-target effects when using CRISPR/Cas9. Careful selection of cleavage sites is key; Cho et al (2014) reported undetectable levels of off-target cleavage when cleavage sites were unique, with no homologous sequences elsewhere in the genome. Also, it has been found that modifications to the sgRNA can reduce off-target effects, but at the cost of decreased on-target efficiency as well. Interestingly, Fu et al (2015) found that reducing the target sequence length from 20 nt to 17 or 18 nt decreased off-target effects without drastic reduction of on-target efficiency. When inserting a Cas9/sgRNA plasmid into eukaryotic cells, transcription is often under the control of a U6 promoter, which requires a leading G on the transcript. Ran et al (2013) found that this leading G may increase target fidelity. Interestingly, Cho et al (2014) noted that, when compared to plasmid encoded sgRNAs, synthetic GG(N)₂₀ sgRNAs produced by T7 enzymes had high discrimination abilities, though some were then less active at their on-target sites.

There are mixed opinions regarding the use of modified Cas9 nickases instead of endonucleases. Cho et al (2014) propose that paired Cas9 nickases can yield highly specific DSBs in human cells. They suggest that, while each Cas9 may produce off-target nicks, DSBs should only occur where both

nicks can hybridize. However, Sander and Joung (2014) noted that when they attempted this, they found indels in regions where only one Cas9 should have nicked the DNA. They hypothesize that DSBs might occur at unrepaired nicks when helicase separates the strands for DNA replication.

Though off-target effects will be difficult to eliminate from CRISPR genome editing, suggestions have been made for differentiating between phenotypes caused by on and off-target mutations. One option would be attempting to have a reporter gene construct that is affected by the on-target mutation. Alternately one could attempt a rescue of the wild type phenotype. Sander and Joung (2014) suggest using various sgRNAs to edit the same target area in separate cell populations. Each sgRNA should have different off-target effects, so if an altered phenotype is present in all mutated strains it is more likely to be from the intended mutation. However, these techniques rely on the mutation producing an easily detectable phenotype, or the researchers knowing what phenotype to expect. When particular off-target loci are of concern one might try the method suggested by Cho(2014). They incubated purified Cas9 and sgRNAs with PCR amplicons containing particular off-target sequences, and used genotyping to assess cleavage.

MicroRNA

My initial genes of interest were the Let-7 family of microRNA (miRNA), which is a type of small noncoding RNA. These molecules are transcribed and processed but are not known to be translated. Instead, each miRNA targets and hybridizes to the 3' UTRs of various mRNAs, affecting their stability and translation; this control is down-regulatory in nature. An mRNA may be regulated by multiple miRNAs, and one miRNA may have multiple binding sites on one mRNA. Additionally, a miRNA family can have dozens to hundreds of mRNA targets.

Functional miRNAs are 21-25 nt in length, but are transcribed as parts of a larger molecule. Often the genes for miRNAs will reside in an intron of a protein-coding gene they regulate. MiRNAs are transcribed from these introns as a series of connected hairpins termed primary miRNA (pri-miRNA), which has a 5' cap and poly-A tail. A microprocessor complex within the nucleus snips apart the pri-miRNA into precursor miRNAs (pre-miRNAs), which have stem and loop configurations. This microprocessor consists of Drosha, an RNase and DGCR8, a dsRNA binding protein. Pre-miRNAs are subsequently exported from the nucleus by the exportin 5 and a Ran-GTPase complex. Once in the cytoplasm pre-miRNAs are pruned by an RNase Dicer, which snips their loop structures and creates

double stranded miRNAs with sticky ends. Dicer then initiates the formation of an RNA induced silencing complex (RISC), which includes Argonaute endonucleases. The double-stranded miRNA becomes unannealed, and one strand is taken up by the RISC complex as a guide RNA while the other 'passenger' strand is degraded. The strand that will become guide RNA is chosen for its lower 5' base pairing stability. Once complexed with RISC the guide RNA can bind an mRNA and cause its translational suppression, or induce the endonuclease activity in RISC to cleave the mRNA. These two outcomes are determined by the complementarity between the guide RNA and the mRNA UTR target. When the two match up imperfectly they anneal to make stretches of dsRNA on the mRNA, suppressing translation. When the guide and target complement perfectly, mRNA cleavage is induced. Each miRNA has what is termed its seed sequence – the 2nd through 8th nucleotides of the guide RNA. This sequence perfectly complements the target mRNA, and is conserved among 'families' of miRNAs that have the same mRNA targets. There are multiple other ways in which miRNA interacts with the ribosome and transcriptional functioning, though the Yi lab does not focus on these.

Down Syndrome

Down Syndrome (DS), or trisomy 21, occurs in 0.14% of live births, and fetuses with DS are 7x more likely to spontaneously abort than non-trisomic fetuses. This aneuploidy occurs via meiotic non-disjunction or translocation. In 1% of cases the non-disjunction occurs in the mitotic division of a somatic cell after fertilization; this results in genetic mosaicism and often yields a less severe phenotype than in non-mosaic individuals. DS results in a variety of physical and cognitive delays and its severity differs among individuals. Typically people with DS have moderate cognitive impairment, dystonia, and characteristic facial features. The disorder can also result in deformities of the heart or gut, and immune insufficiencies. On a cellular level DS results in many markers of cellular stress, including high levels of reactive oxygen species (ROS), increased DNA damage and decreased DNA repair mechanisms, increased chromosomal instability, and dysfunctional mitochondria. It is not currently known what aspect of trisomy 21 causes the DS phenotypes. It has been suggested that the presence of aneuploidy is a global cellular stressor, capable of causing developmental impairments and cellular dysregulation. While this makes sense given the constitutive cellular stress responses seen in DS, it does not explain why triploidy of most other chromosomes is embryonically lethal. (Trisomy of chr13 and chr18 are technically survivable, but at a much lower frequency). This suggests that specific chr21 genes have an effect on the DS phenotype.

In support of this idea is the fact that trisomy 21 also predisposes individuals to a specific disease profile. Individuals with DS have a very low incidence of cancerous solid tissue tumors (with the exception of testicular cancers), but have much higher rates of childhood leukemias. Compared to the general population, DS is associated with a 33x increase in B cell lymphoblastic leukemia, and a 150x increase in acute myeloid leukemia. Work done by Nižetić et al (2012) indicates that chromosome 21 (chr21) carries genes that can impede cell proliferation and migration, as well as angiogenesis, which are traits needed for solid tumor formation. Conversely, chr21 also appears to encode oncogenic factors that specifically increase the likelihood of leukemias and germ cell cancers.

Unfortunately, another aspect of the DS disease profile is the seemingly universal development of early onset Alzheimer's Disease (AD) (Wilcock and Griffin, 2013). Not all individuals with DS will develop AD related dementia, though dementia symptoms appear by the age of 40 for 50% of those with DS. Additionally, cytological indications of AD, including plaques of amyloid-beta and tangles derived from mis-cleaved tau protein, are ubiquitous in adults with DS and can be seen even in some children with DS. It has been suggested that the link between DS and AD is the presence of the amyloid precursor protein (APP) on chromosome 21. Over-expression of APP appears to mean that more APP is available to be cleaved by the proteinase gamma secretase, producing excessive amyloid-beta fragments.

Cell types and genes studied

The name of the miRNA Let7 is derived from the word lethal, as its loss in *C. elegans* is not survivable. Worms only possess one Let7 gene, while the mammalian genome appears to have accumulated redundant copies – the Let7 family in mammals has 11 loci. The lethality of a full worm knockout combined with the insurance against loss seen in mammals indicate that Let7 is likely crucial to cellular functioning. This is corroborated by the fact that most research into the targets and effects of Let7 focuses on skin diseases and cancers. Decreased Let7 appears associated with skin fibrosis, and supplementation with Let7 can alleviate symptoms (Jinnin, 2014). Interestingly, an increase in Let7 levels has been found in skin samples from psoriasis patients. Let7 appears to function as a tumor suppressor by decreasing the translation of Ras GTPase and the activating transcription factors Myc and E2f2, all of which can be oncogenic (Johnson, C, 2007; Johnson, S, 2005). Let7 is also known to function as a suppressor in cellular processes that occur in development and cancer, including cell cycle regulation, DNA replication, cell division, and apoptosis (Albrekt, 2014). Cardenis et al (2014) have

also demonstrated that the inhibition of *Let7* can cause the expression of mesenchymal markers consistent with an epithelial/mesenchymal transition.

These phenotypes make the *Let7* family a prime candidate for an in vitro knockout study. Also, because the *Let7* sequences are scattered throughout the genome, it is not feasible to attempt their deletion by genome editing requiring site specific enzymes. However, transfection with multiple CRISPER/Cas9 plasmids has the potential to generate cells with many *Let7* loci deleted.

Given its clear tumor suppressor activity, I hypothesize that the loss of multiple *Let7* miRNA genes will cause increasingly tumor-like phenotypes in the keratinocytes (e.g. increased growth rates, decreased terminal differentiation). I also hypothesize that the redundancy of *Let7* in the genome indicates that this miRNA is necessary for normal cellular functioning. Based on this, I predict that the cells will have the ability to compensate for the loss of only a few *Let7* sites by up-regulating transcription at the remaining loci. This would mean that the initial few knockouts may show little to no change in phenotype.

Mouse keratinocytes are an adherent cell type originally harvested from mouse back skin and then immortalized. In vivo these cells make up most of the epidermis, making them a key aspect of the Yi lab's developmental model system – mouse skin and hair follicles. Additionally, mouse keratinocytes express 9 of the 11 mammalian *Let7* miRNA, making them ideal for my knockout study. These 9 miRNA are found in 7 loci in the genome. These are: *7a-1/7f-1/7d* on chromosome 13 (cluster 1), *7c-2/7b* on chr 15 (cluster 2), *7c-1* on chr 16, *7e* on chr 17, *7g* on chr 9, and *7i* on chr 10.

Because the Yi lab studies small RNA molecules and developmental mechanisms (which can overlap with cancer mechanisms), the capacity to knock out *Let7* in keratinocytes may be valuable to other members of the lab. Also, on a practical level, keratinocytes are good cell type for observing our hypothesized tumor-like phenotypes of *Let7* knockout. Their tendency to grow in a single layer until confluent allows for easy observation of increased cell proliferation, changes in cell morphology, or a decreased frequency of terminal differentiation.

After preliminary knockout of two *Let7* clusters was achieved, we began discussing the issue of off-target effects. Other lab members genotyped transfected cells for Cas9 several weeks after transfection, and found that the DNA was still present. This was concerning, as it implied that the plasmids had not been degraded, and may have been integrated into the genome. To avoid this possibly confounding

issue we began attempting transfection using only Cas9 RNA and sgRNAs, which successfully edited the genome when injected directly into fertilized mouse eggs.

While I was working on Let7 Dr. Garrido-Lecca, a post doctoral fellow in the Yi lab, began work on a Down Syndrome study. Since the genetic basis for the Down Syndrome phenotypes is not clear, Dr. Yi proposed that we apply the idea of transient transfection with CRISPR RNA to the Down Syndrome Project. Exogenous RNA should be rapidly degraded in the nuclei of cells, especially the miRNA which have neither a 5' cap nor a 3' tail. We hypothesized that this technique could be used to induce temporally brief transfections, leaving limited time for editing, and thereby producing single allele knockouts of chromosome 21 genes. This proof of principle would be incredibly valuable to the genetic study of Down Syndrome, as it would allow us to assess the expression of each chromosome 21 gene at its normal gene dosage, against a background of trisomy 21. The contribution each gene makes in Down Syndrome would become somewhat easier to assess.

Dr. Garrido-Lecca received lymphoblastoid cells which were collected from a proband with Down Syndrome and his family members, and then immortalized with Epstein-Barr Virus. Based on his preliminary analysis of total RNA sequencing and ribosome profiling, Dr. Garrido-Lecca identified the chromosome 21 gene for amyloid precursor protein (APP) as a target for further investigation. While APP is of primary concern in neurons, it is expressed in lymphoblastoids as well, and these cells appear capable of releasing cleaved amyloid into the blood. Somani et al. (2014) found evidence that increased levels of amyloid beta in blood plasma is important in the development of dementia in patients with Down Syndrome. We hoped to optimize specific transfection conditions which would allow for genome editing for different lengths of time. In this way, we were hoping to create Down Syndrome lymphoblastoid cells that had one, two, or three copies of APP knocked out. The number of copies present in cells that show heterozygous genotypes can be determined by qPCR. Should it be necessary to know which alleles were present in a particular clonal colony, fluorescent tags could be designed to indicate allelic identity when used in FISH. Alternately, primers could be found to amplify regions with allele specific SNPs, and amplicons could then be sequenced.

There are multiple ways in which APP protein might be over-expressed in DS cells beyond the expected level of 1.5x. Increased expression of chr21 genes may interact with one another in novel

ways, or they may alter regulation in other genomic regions, which feed back and change regulation on chr21. For example, transcription or translation rates might be either enhanced or de-repressed; mRNA transcripts might be actively stabilized by a molecule that is itself over-expressed, or the miRNAs that downregulate APP might be sponged away by a competing up-regulated mRNA target. Because of this, APP protein levels may not change in predictable ways as the alleles are knocked out.

Unfortunately, the lymphoblastoid cells proved highly resistant to chemical transfection as well as electroporation. Because our supply of them is limited, we chose to attempt transfection of another human cell line until my sgRNAs were verified as capable of modifying the genome. Human embryonic kidney cells (HEK293T) were chosen because they are durable in culture and easily transfectable. The T in the cell designation indicates that they possess SV40 large T antigen, allowing for episomal replication of plasmids containing the SV40 origin of replication. This is a way of identifying the cells, but we do not take advantage of this characteristic, as we do not want the plasmids to be freely replicated.

Methods:

Cell Culture

It should be noted that keratinocytes are not fully differentiated cells. In vivo they gradually migrate from the basement membrane outward, in layers. As they get closer to the surface of the skin they encounter an increasing concentration of Ca^{2+} ions, which aids in their eventual terminal differentiation into anucleated corneocytes. Because of this, keratinocytes must be cultured in low calcium media to prevent differentiation; we use low Ca^{2+} E media. Also, even in this media, once the cells become confluent they begin to form layers of cells and can eventually differentiate past the point at which they are useful for cell culture. That is, they no longer maintain a strong proliferation rate despite immortalization. Because of this, cells were passaged (split to 1/10) each time they became confluent.

Lymphoblastoid cells are suspension cells, and were grown in 20ml RPMI with 15% FBS and 1x pen/strep in flasks with self venting lids. They were passaged (split to 1/5) each time they reached 1 million cells per ml, determined by counting on a hemocytometer. If the cell media became at all cloudy this indicated contamination, and the culture was discarded.

Adherent HEK293T cultures were grown in DMEM with 10% FBS and 1x pen/strep. They were passaged (split 1/10) each time they became confluent, or when a noticeable number of dead cells were seen floating in the media. These cells can have a tendency to grow in piles instead of spreading out, and can become dense enough to deplete the media despite not covering the plate surface. They are terminally differentiated, however, so some overgrowth is not as much of a concern.

Also, these cells do not adhere tightly to the plate, and are easily peeled off. Care must be taken when washing cells or changing their media, and they need very little time in trypsin. When plating, count cells and then mix them thoroughly with the correct volume of media. Then add them to the plate with minimum disruption. I allow them to sit in the hood for a few minutes after plating to settle before moving them to the incubator, as they can easily become unevenly distributed on the plate.

All cells were incubated at 37 degrees C with 5% CO₂.

Design of CRISPR constructs

Endonuclease targets for murine (mm9) Let7 and human (hg19) APP were determined using the same process. Using the UCSC Genome Browser, target locations were determined. For the Let7 clusters, regions were selected which flanked the clusters without interfering with adjacent genes. For APP, target regions were chosen to flank the entire gene (300kb) as well as the gene's promoter region and first exon (5.6kb). This second deletion size was designed because I hypothesized that a smaller excision would be more easily repaired with NHEJ. These target ranges were fed 250 bp at a time into the optimization engine at crispr.mit.edu. This program locates viable DNA targets (20 bp, immediately upstream of a PAM sequence) and ranks them based on their predicted capacity for off-target effects. This is based on the number of predicted off-target sites, the number and position of mismatched bases at those sites, and whether or not the sites are within known gene coding regions. For my purposes I gave priority to targets with higher MIT scores (>90) over those that created a shorter deletion. No consideration was given to which strand the target occurred in, as this shouldn't matter when using wild type Cas9.

Once chosen, crRNA sequences were entered into NCBI's primer BLAST software and blasted against the appropriate genome to search for obvious extraneous targets. None were found for my Let7 crRNAs in the mouse genome or APP crRNAs in the human genome.

Cloning crRNA sequences into plasmids

Let7 Cloning

This project aimed to delete Let7 loci by NHEJ, using two sgRNAs flanking each region to be deleted. The first loci to be targeted for knockout were cluster 1 (C1) and cluster 2 (C2). A CRISPR construct plasmid was designed for each, which coded for Cas9/EGFP and contained two cluster specific sgRNA sequences (Diagram 1). To achieve this, Dr. Wang from the Yi lab designed two plasmids: K14 BbsI mutant Px458 (mutPx458), and U6 T.

MutPx458 includes ampicillin resistance, a U6 promoter controlled tracrRNA sequence, and a K14 promoter controlled NLS/Cas9/NLS/T2A/EGFP complex (Diagram 1). The nuclear localization sequences (NLS) ensure that the Cas9 proteins are retained in the nucleus, which maximizes genome editing. The T2a is a viral construct that allows for ribosomal “skipping”. This sequence contains codons for proline followed by glycine which allows for the ingress of translation release factor eRF1. This catalyzes the cleavage of the peptide from the tRNA which holds it, but without signaling to the ribosome that translation should be terminated. This then allows for EGFP to be translated immediately following cas9, and ensures that the two proteins will be expressed in equal molar amounts, while avoiding steric hindrance in the folding of either protein (Minskaia and Ryan, 2013). Additionally, since EGFP does not have an NLS it diffuses throughout the cytoplasm. This makes seeing EGFP+ cells somewhat easier when using a fluorescent microscope, but might slightly impede cell sorting (a cell with a more diffuse signal may not make the EGFP+ signal cutoff).

U6 T is smaller, primarily coding for ampicillin resistance and a U6 promoter controlled tracrRNA sequence. Target oligos were designed that allowed for cloning of one crRNA into each of these two plasmids. Then, using KpnI and XbaI restriction enzymes, the U6 promoter and sgRNA were cleaved from the U6T vector and spliced into mutPx458 to create one master plasmid (Diagram 2).

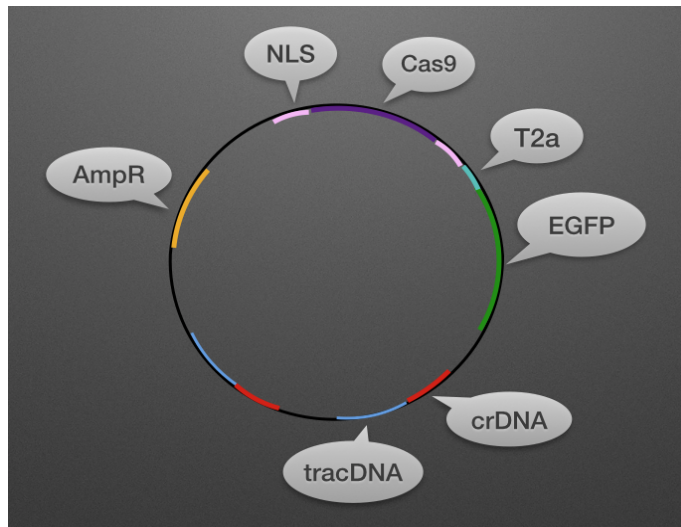


Diagram 1: Schematic of the mutPx458 master plasmid.

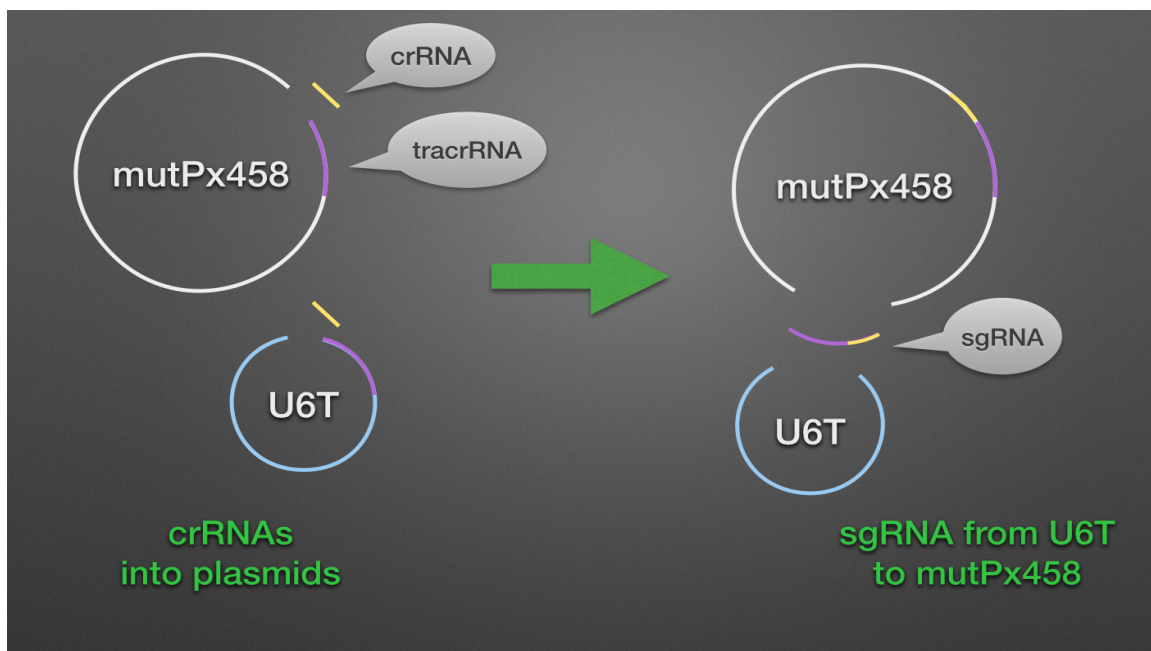
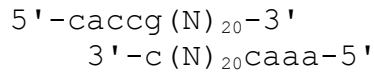


Diagram 2: Making a master plasmid for transfection. The mutPx458 plasmid also carries Cas9/EGFP which has been omitted in this diagram.

For each plasmid, crRNA oligos were inserted at BbsI restriction sites. BbsI is useful for this purpose because its recognition site is not its cleavage site(*). Instead, it leaves a nonspecific sticky end downstream of its recognition site:

$$\begin{array}{l}
 5' \dots \text{GAAGAC} (N)_2^* \dots 3' \\
 3' \dots \text{CTTCTG} (N)_6^* \dots 5'
 \end{array}$$

This enables the plasmid to be designed such that two adjacent BbsI sites cleave different sequences, allowing annealed crRNA oligos to be inserted into the plasmid immediately 5' to the tracrRNA with the correct directionality. This is necessary because the sense strand of the crRNA must be annealed to the sense strand of the tracrRNA. Cas9 cannot use the reverse complement of the target oligo, as it is unlikely to be immediately upstream of a PAM sequence. Our crRNA oligos were designed with the following format:



The U6 promoter, used for transcription of both sgRNAs, requires a leading G at the start site. Because of this, the G/C pair in the schematic are only included when the crRNA does not itself begin with a G.

Since each crRNA is inserted immediately 5' to a tracrRNA segment, they are transcribed as one chimeric piece. The entirety of the *S. pyogenes* tracr sequence is included on our plasmid, which was found by Sander and Joung (2014) to yield higher site specificity than transcripts with truncated tracrRNA sequences.

After the master plasmids were assembled they were each used to transform competent bacteria and plated on Ampicillin plates. Colonies that grew were then genotyped based on the plasmid they were transformed with. To do this we used the forward oligo for the left crRNA and the reverse oligo for the right crRNA as primers. Positive colonies were grown overnight in 5ml of LB+Amp and then processed via Mini-Prep to extract the plasmids. Plasmids were then sent for sequencing to confirm their inserts.

APP Cloning

In order to knock out APP I initially designed crRNA oligos to remove either the entirety of APP or a section of the gene, hopefully resulting in loss of functional protein either way. Our initial intention was to knock out APP by transfecting with only Cas9/EGFP mRNA and sgRNAs. This required that each crRNA be inserted into its own plasmid for clonal amplification. Putting two into a master plasmid was counter productive as it would yield extraneous amplification products.

Competent cells were transformed and plated as above. Colony PCR was done on each plasmid using the forward crRNA oligo as a forward primer and a reverse primer complimentary to the end of the tracrRNA sequence. Positive colonies for each of the four plasmids were grown up in 1 L of LB+Amp

overnight, and plasmids were extracted using a Maxi-prep kit. After collecting purified plasmids the sgRNAs were amplified from their plasmid templates in preparation for in vitro transcription.

We attempted a variety of methods to create a plasmid containing a T7 promoter in front of the NLS/Cas9/NLS/T2A/EGFP complex, in order to generate mRNA. While we could amplify the complex from mutPx458 as we did the sgRNAs, we wanted to avoid this if possible. The entire amplicon is over 5kb in length, and errors made in PCR amplification were a concern. Even using Phusion, a high fidelity polymerase, could introduce mutations, as it has no proof reading capacity. Instead, we wanted T7 to be on the plasmid with the amplicon, so that the entire sequence could be amplified in bacteria and then cut from the plasmid via restriction enzymes.

We did eventually attempt to PCR amplify the gene complex with a T7 promoter on a forward primer, and this appeared to be successful based on amplicon size when run on an agarose gel. However, I then noticed that mutPx458 was in fact carrying a T7 promoter already, allowing us to linearize the plasmid and use it directly for in vitro transcription.

In vitro transcription

To prepare for in vitro transcription of sgRNAs, the sgRNA templates were PCR amplified from their plasmids. Forward primers were used which included the T7 promoter. Because we wanted sense RNA, the primers were designed such that the T7 promoter was added to the 3' end of each sgRNA antisense strand.

For example:

T7 promoter plus sense crRNA used as forward primer:

5' TAATACGACTCACTATAGGATTTCACTTCGGACTTCGGT 3'

Universal reverse primer, complimentary to the end of the **tracrRNA** sequence:

3' GTGGCTCAGCCACGAAAAA 5'

Plasmid sgRNA sequence to be amplified:

5' ATTTCACTTCGGACTTCGGT.....CACCGAGTCGGTGCTTTTTT 3'
3' TAAAGTGAAGCCTGAAGCCA.....GTGGCTCAGCCACGAAAAA 5'

Priming:

5' ATTTCACTTCGGACTTCGGT CACCGAGTCGGTGCTTTTTT 3'
 <-- 3' GTGGCTCAGCCACGAAAAA 5'

5' TAATACGACTCACTATAGGATTTCACTTCGGACTTCGGT 3' -->
 3' TAAAGTGAAGCCTGAAGCCA GTGGCTCAGCCACGAAAAA 5'

Final dsDNA amplicon:


5' TAATACGACTCACTATAGGATTTCACTTCGGACTTCGGT CACCGAGTCGGTGCTTTTTT 3'
 3' ATTATGCTGAGTGATATCCTAAAGTGAAGCCTGAAGCCA GTGGCTCAGCCACGAAAAA 5'

Creation of APP sgRNAs was completed with MEGAShortScript from Ambion, which is specific for RNA transcripts 20-500 bases in length. Short transcripts can be difficult to amplify, as they require proportionally more transcription initiation events, and this is the rate limiting step for the polymerase enzymes. This kit has been designed to overcome these drawbacks and maximize RNA yield, in a presumably proprietary fashion. Because the T7 promoter is only appended to the 5' end of our dsDNA template, the polymerase complexes with the 3' end of the anti-sense strand, and only RNA matching the sense strand is synthesized. (Adding the T7 promoter to the other end of the template causes anti-sense RNA to be transcribed).

sgRNA PCR amplicon from above:

5' TAATACGACTCACTATAGGATTTCACTTCGGACTTCGGT CACCGAGTCGGTGCTTTTTT 3'
 3' ATTATGCTGAGTGATATCCTAAAGTGAAGCCTGAAGCCA GTGGCTCAGCCACGAAAAA 5'

Sense in vitro transcription:

5' UAAUACGACU  T7 RNA Pol
 ----->
 3' ATTATGCTGAGTGATATCCTAAAGTGAAGCCTGAAGCCA GTGGCTCAGCCACGAAAAA 5'

Samples of the sgRNA transcripts were then denatured and run on a 2% agarose gel to checked for correct size, length uniformity, and general concentration. When not denatured these sgRNAs are able to adopt their secondary conformation, which yields a band at about 200 bases. The sgRNAs were also run on a urea PAGE gel for finer granularity. Purification was carried out using phenol chloroform, and the RNA was subsequently ethanol precipitated. We found that these methods of cleanup gave much higher yields of sgRNA than did the MEGAclean Kit column filtration system.

Dr. Garrido-Lecca attempted to make NLS/Cas9/NLS/T2a/EGFP mRNA using the mMACHINE mMachine T7 Ultra kit from Ambion. This kit is used for in vitro transcription followed by addition of a 5' Anti-Reverse Cap Analog and a poly A tail. The ARCA ensures that the cap is added in the correct orientation to all transcripts. However, this kit is also designed for transcripts in the 0.3 – 5 kb size range, and our transcript is slightly larger than 5 kb, which may have been an issue. The initial RNA transcripts produced in this experiment were not a consistent size, and were smaller than 5kb. The kit suggested running the transcription with additional GTP, as this is a limiting reagent for larger transcripts and should produce a higher yield of full length product. However, increased GTP decreases the percentage of transcripts that are capped, as this reaction relies on a specific ratio of GTP to ARCA. Additional GTP did improve the length of our transcripts, but for some reason our mRNAs were not functional.

Lipid Transfection

Transfecting eukaryotic cells is a balancing act between effective transfection rates and levels of cell toxicity, and optimal protocols can vary widely between cell types. Also, even if cells are not killed outright, the ingress of foreign DNA as well as the presence of transfection reagents can induce cellular stress responses. These can include perturbation of the cell cycle, changes in metabolic signaling, and nonspecific activation of various genes. Because of this, once transfection is achieved, a control should be designed which exposes cells to the transfection process without the possibility of genome modification. This serves as a more accurate baseline when assessing phenotypic changes in experimental cells.

Also, protocols can vary widely in their suggested rates of confluency. For example, Biocompare (www.biocompare.com) notes that plasmid transfection is aided by cell division as this involves degradation of the nuclear envelope, which allows plasmids to be more easily incorporated into the nuclear space. Because of this, plasmid transfection should be done on cells that are actively dividing, and therefore at 60-80% confluency (for adherent cells). This is not as crucial for the transfection of mRNAs or sgRNAs, which only need enter the cytoplasm to be effectively processed. Biocompare suggests that transfection of these smaller molecules be done at higher confluency when using lipid reagents, in order to avoid the cytotoxic effects of many lipid complexes being delivered to each cell.

Alternately, protocols for lipid transfection reagents suggest confluencies ranging from 50 – 90% for adherent cells. Lipofectamine 2000 goes on to specify that small RNAs should be transfected when cells are 30-50% confluence, in order to minimize loss from cell overgrowth.

These contradictions illustrate an important concept: transfection protocols must be optimized for each cell type and each method of transfection. This involves transfecting at varying cell densities with different cell:reagent:nucleic acid ratios, or different cell:voltage:pulse ratios when electroporating.

Transfection for Let7

My two Let7 knock out plasmids (one for each cluster) were transfected separately into keratinocytes plated in their usual media (E low Ca^{2+}). The transfection was carried out using TransIT-LT1 cationic lipid reagent, mixed with plasmid at the suggested ratio of 3ul lipid to 1ug plasmid. Lipid/plasmid complexes were formed within antibiotic free P no calcium media.

An initial attempt at this procedure was done in 24 well plates seeded at 20,000 cells per well one day prior to transfection. The cells were high passage and divided quickly, so they were about 60% confluent at transfection. The lipid/plasmid complexes were added dropwise very gently, with a minimum of mixing. 24 hours post transfection we saw a few fluorescent cells and many dead cells. This may well be due to the fact that we used mini-prepped plasmids, which can contain endotoxins. For our second transfection attempt we used maxi-prepped plasmids, and 6 well plates seeded at 200,000 cells per plate. Again, efficiency appeared low 24 hours after transfection, though cell mortality seemed reduced.

Transfection for APP in Lymphoblastoid cells

TransIT-LT1 reagent is specifically designed for transfection with plasmids, and is not recommended for use with RNA; instead, we used Lipofectamine 2000 from Invitrogen. Nucleic acid/lipid complexes were formed in Optim-Mem low serum media, and the cells were plated in complete RPMI. Based on Dr. Wang's success with RNA transfection, we used 1 pmol of Cas9/EGFP and 32 pmol of each sgRNA per well. We plated 4 wells each of lymphoblastoid and HEK293T cells (as a putative positive control) at 80% confluence in a 24 well plate. For each type of cell we used 1, 1.5, 2 and 2.5 ul Lipofectamine 2000 reagent across the 4 wells.

Our next transfection attempt was done only on lymphoblastoid cells, using sgRNAs and maxi prepped

Px458. This is the non mutant plasmid, which codes for NLS/Cas9/NLS/T2A/EGFP, but is only about 9000 bases while the mutant plasmid is almost 12000 bases. The protocol suggests that initial transfection be done with varied amounts of reagent and 800 ng DNA plus 30 pMol of each sgRNA per 1ug of DNA. We used 1 ug plasmid and about 30 pMol (1 ug) of each sgRNA per well. 24 hours post transfection there were no EGFP+ cells, so we re-transfected with another 1ug plasmid per well and pipetted the cells up and down to thoroughly separate the cells and mix in the DNA. This again yielded no transfection.

We tried once more in lymphoblastoid cells using 1ug Px458 or 1ug newly synthesized Cas9/EGFP mRNA, but no sgRNA. Lipofectamine was varied across 4 wells as before. Again, this yielded no EGFP+ cells.

It is interesting to note that the lymphoblastoid cells seemed to experience low toxicity during these transfections. When counted with the hemocytometer, there was no noticeable change in the number of dead cells seen before and after treatments with Lipofectamine 2000.

Transfection for APP into HEK293T

In order to not waste lymphoblastoid cells I woke a fresh vial of HEK203T and plated 8 wells of a 24 well plate at 50,000 cells per well, in complete DMEM. As with the lymphoblastoid cells, I used Lipofectamine 2000 and Opti-Mem media to form the nucleic acid/lipid complexes. The cells were transfected at about 70% confluence after allowing them to grow for 24 hours.

In 4 wells I transfected with 1ug each of Px458 and the two sgRNAs, mixed with 1, 1.5, 2, or 2.5 ul of lipid. The same was done for the other 4 wells, except 1ug of Cas9/EGFP was substituted for the Px458 plasmid. The cells were allowed to grow for three days, and were then checked for EGFP. EGFP signal was seen in all of the plasmid wells, and none of the mRNA wells. By visual inspection, increasing Lipofectamine levels gave increasing fluorescence in the plasmid wells.

Since this was successful, and because I wanted many more cells for sorting and genotyping, I grew up and transfected 3 10cm dishes of HEK293T cells. With this volume of cells the protocol recommended 60ul of Lipofectamine 2000 and 24ug DNA per plate. I also used 24ug of each sgRNA per plate. A strong transfection rate (determined visually) was seen three days after transfection.

My next iteration of transfecting HEK293T cells was done to try a sequential experimental design. Additionally, more extensive Lipofectamine 2000 protocols were found and followed. Six 10cm dishes of cells were plated and allowed to reach 70-80% confluency. All 6 plates were then transfected with

px458, 3 using Lipofectamine 2000 and Opti-Mem medium, and the other 3 using TransIT-LT1 and Opti-Mem medium. Lipofectamine plates received 24ug of plasmid in 60ul lipid, while TransIT-LT1 plates received 15ug plasmid in 45ul lipid, per reagent protocols. The extended Lipofectamine instructions state that cells should be transfected without antibiotics in their growth media to avoid excessive cell death. For consistency, all 6 plates had their media switched to DMEM with 10% FBS and no antibiotics a few hours before transfection. Less than 24 hours after transfection one Lipofectamine 2000 plate had become contaminated and died; the other 5 plates showed EGFP+ cells. By 60 hours post transfection all 5 plates showed strong transfection rates, and differences between Lipofectamine plates and TransIT-LT1 plates were undetectable by eye. After sorting these cells for EGFP, a population of 100% EGFP+ cells will be re-plated and transfected with 600pmol of each sgRNAs using Lipofectamine 2000. This will be done when the cells are 30-50% confluent (vs. the 70-90% confluence suggested for plasmid transfection).

Cell Sorting

Cell sorting was done on a MoFlo XDP cell sorter. This machine can sort cells based on multiple parameters, including cell size and various levels of fluorescence. When sorting cells singly into wells it is a good idea to plate more cells than should be needed. Some cell types, including keratinocytes, do not grow well when plated individually. Also, it is possible for the machine to appear to be sorting cells into a 96 well plate when it is not actually doing so.

Let7 in Keratinocytes

Approximately 3 million cells were sorted for each Let7 cluster, 24 hours after transfection with maxi prepped master plasmids.

The cells were trypsinized for ten minutes, quenched with E low calcium media, filtered through a 40 micron filter, spun down and resuspended in 1ml of PBS + 3% FBS. 1ul Hoescht DNA counterstain was added before cells were put in a sterile, capped flow tube on ice.

For each cluster three 96 well plates were prepared for single cell sorting by adding 200ul of E low calcium media to each well.

Sorting was gated for size, then Hoescht expression, and finally for high EGFP expression, and cells were sorted singly into the 96 well plates.

APP in HEK293T

Two 10cm plates of cells treated with 24ug each of px458 and the sgRNAs in Lipofectamine were sorted four days post transfection. To maximize sorting efficiency it was recommended that I sort 4-5 million cells per ml of liquid. Based on LifeTechnology's estimate that 8.8 million cells per dish constitutes confluence in a 10cm plate, I approximated the number of cells present in each dish by estimating the percent confluence of each plate and multiplying that by 8.8 million. I then trypsinized the cells, spun them down and resuspended them in less media than I thought necessary for a concentration of 4.5 million cells per ml. Cells were counted using the hemocytometer and the volume of media was adjusted to give the desired concentration. Samples were filtered through 40 micron filters into sterile flow tubes on ice. For cell collection I prepared three 96 well plates for each sample, with 200 ul complete media in each well. Additionally, a sterile flow tube containing 1 ml of complete media was made for each cell sample, and placed on ice.

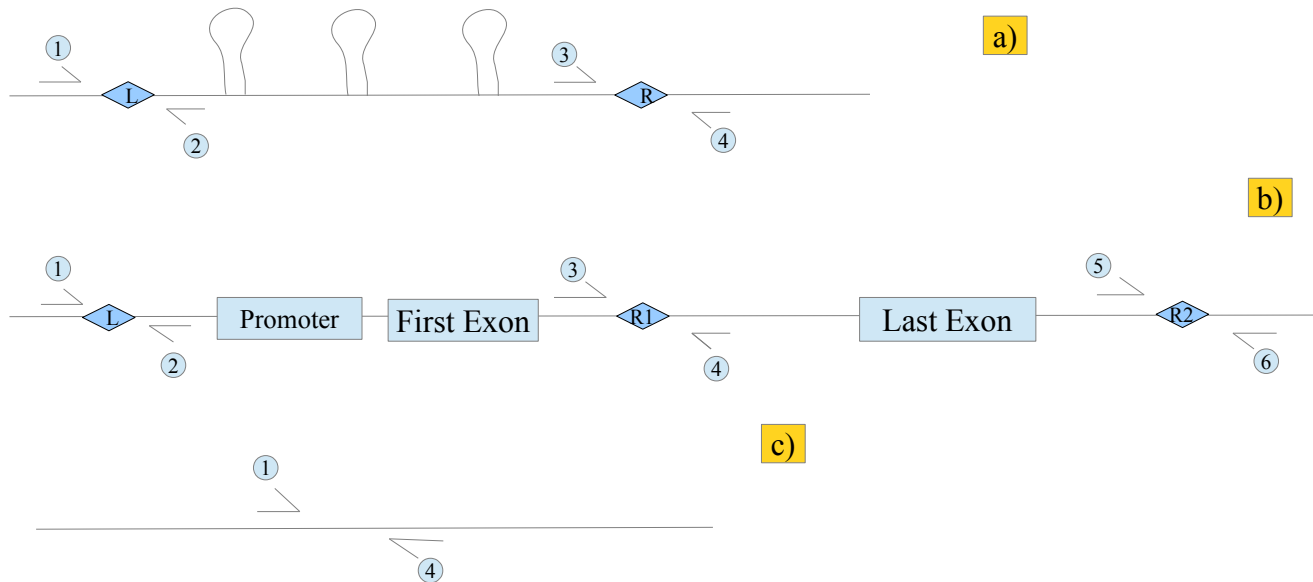
For each original 10 cm plate I collected 3 96 well plates of singly sorted cells, and 880 thousand cells in a collection vial. Sorting was gated based on size and high EGFP expression.


Knockout Genotyping

To acquire genomic DNA for genotyping cells are collected and mixed with 9ul of ear punch buffer and 1ul proteinase K. Cells are lysed for 30 minutes at 55C, and then diluted with 20ul of nuclease free water. The samples are then placed at 95C for 5 minutes to heat inactivate the enzyme. When genotyping was inconclusive on samples prepared in this way, the DNA was ethanol precipitated to remove salts and other impurities, then redissolved in nuclease free water. DNA concentration was then assessed by NanoDrop.

In order to assess transfected cells for gene knockouts I designed primers to flank each putative sgRNA target site (Diagram 3). To do this I used primer designing software from the University of Massachusetts Medical School (http://biotools.umassmed.edu/bioapps/primer3_www.cgi). For each genome modification I effectively had three primer pairs: “left”, which flanks the sgRNA site at the 5' end of the desired deletion (1 and 2 below), “right”, which flanks the sgRNA site at the 3' end of the deletion (3 and 4 below) and then the “knockout pair” consisting of the left forward primer and right reverse primer (1 and 4). For both the Let7 and APP designs this knockout pair prime sites that are more than 1000 bp apart in an intact genome. However, they are only 100-200 bp apart if NHEJ was carried out successfully using DSBs at the two sgRNA sites. For this PCR reaction I use taq

polymerase and the elongation time is set to 30 seconds. This means that an intact genome should yield no amplicon from the knockout pair, while a modified genome should yield the short amplicon.



SgRNA target 

Primer 


Let7 hairpin 

Diagram 3: Schematics of genotyping primers flanking sgRNA target sites within the genome. a) Let7 cluster 1 schematic b) APP schematic. When knockout pairs (a: 1 and 4, b: 1 and 4 or 1 and 6) give an amplicon it implies that the deletion was successful, since PCR conditions do not allow for the larger non KO length to be amplified. c) Schematic of what knockout allele for a)

Ideally, using these three pairs of primers should yield the cells' genotype for the knockout (KO) locus. If a left and right amplicon are seen with no KO band it implies that no genome modification has taken place. Additionally, the left and right bands serve as a partial positive control, showing that the DNA is intact, the PCR mix and conditions were functional and all four primers are capable of annealing. Should genotyping come back with all three amplicons present, it implies that the cells have had one allele modified while the other is intact. And if only a KO band is present it implies that the cells are homozygous for the genome modification.

Unfortunately this does not always work perfectly, and since there is no true positive control for the KO band it can be hard to know if the sgRNAs failed to modify the genome or if the KO primer pair don't work well together, or require different conditions. To assess these options it is essential to have a non template control (NTC).

These PCR products could be seen with the best resolution when run on a 2% agarose gel with

ethidium bromide at 125 volts.

Let7 Genotyping

The primer sets I designed for the cluster 1 and cluster 2 Let7 knockouts worked as intended. When used on unsorted samples of transfected keratinocytes left, right, and KO amplicons could all be seen. While there were a few non-specific bands at matching lengths in the experimental and NTC reactions, they did not interfere with the interpretation of the gels.

APP Genotyping

When I genotyped the transfected, unsorted HEK293T cells, inappropriate bands were seen in both the left sgRNA NTC and the KO NTC PCR products. To check for systemic contamination I repeated the amplification, with all new reagents, very low levels of primers (0.25ul of each 10uM primer per 20ul reaction) and clean, concentrated genomic DNA from another source. When this showed very low amplification for the left and right primer pairs, it was clear that the primers needed optimization. To do this, I ran PCR reactions with and without DNA template using a gradient of annealing temperatures (55 – 60 C) . I then ran these samples on 2% agarose gels to assess the strength of amplification and to check for nonspecific bands in the NTC samples.

When this was not sufficient to achieve strong amplification of the left and right primer pairs, I modified the PCR reactions with either 1ul of 25mM MgCl₂ or 1ul DMSO. I ran these reactions with a wider annealing temperature gradient (43 – 57 C). This clarified the PCR recipe and annealing temperature of the KO primer pair, though samples still showed some smearing on the agarose gel. To attempt to clarify this I re-ran the PCR reaction using a NEB 5x premixed reagent containing taq, dNTPs, buffer, MgCl₂, KCl and “stabilizers”. Per the reagent protocol, the PCR elongation temperature was dropped from 72C to 68C. This reaction gave much sharper bands, and was used for all subsequent APP genotyping.

For future use, I also designed a new, KO specific pair of primers based on the anticipated DNA sequence at the site of a successful APP knockout. Since these primers were designed to work together, they may provide clear, corroborating genotyping for future APP KO studies.

Electroporation

Trisomy 21 lymphoblastoid cells were transfected with 1ug each of px458 and the APP sgRNAs using

a Life Technologies Neon Transfection System. Transfection efficiency and cell viability can change depending on the size of the molecules being transfected; larger plasmids are less effectively integrated into cells. capable of delivering nucleic acids in very small volumes of cells. Specific protocols have been developed for many cell lines, but none of these exactly match our immortalized yet non clonal samples. We chose to try the transfection protocols for HL 60 cells and Jurkat cells. These were chosen because both cell lines are human suspension lymphoblasts derived from blood with media requirements similar to those of our cells. Prior to transfection, cells were washed with PBS and resuspended in R buffer. The Neon tube itself was filled with buffer E2.

Using a 100ul transfection tip, HL60 calls for one 1325 volt pulse at a width of 35 ms. They are concentrated to 5 million cells per ml prior to electroporation, and have an average transfection efficiency of 55% and average viability of 70%.

The Jurkat cell protocol specifies electroporation at a density of 20 million cells per ml. In a 100ul tip they receive three 1350 volt pulses at a width of 10ms each. Average transfection efficiency is 93.7% and average post transfection cell viability is 97.8%

Post transfection, each batch of cells was suspended in 4ml of prewarmed, antibiotic free media in a 6 well plate.

Three days post electroporation neither cell sample was EGFP positive. However, this is apparently not unusual when using a non-optimized cell type. The Neon system has an involved optimization protocol which will be tried in the future.

Results and Discussion

Let7 in Keratinocytes

C1 colonies proved more difficult to amplify than did the C2 colonies. PCR was attempted multiple times before the correct annealing and elongation times were established. However, C2 plasmids returned poorer sequencing results than did C1 plasmids. The cells transfected with mini-prepped versions of these plasmids showed poor transfection rates, based on visual assessment. However, the presence of EGFP+ cells indicated that transfection was possible. Transfection rates with maxi-prepped plasmids were stronger, but still under 10%. Cells were sorted based on size matching control keratinocytes, low Hoescht fluorescence, low pulse width, and then high EGFP vs RFP. For C1 2.5% of the cells passed the criteria for all of these gates, while 6% of C2 cells passed. This would likely have

improved if the cells had been left for longer than 24 hours before being sorted (Fig 1).

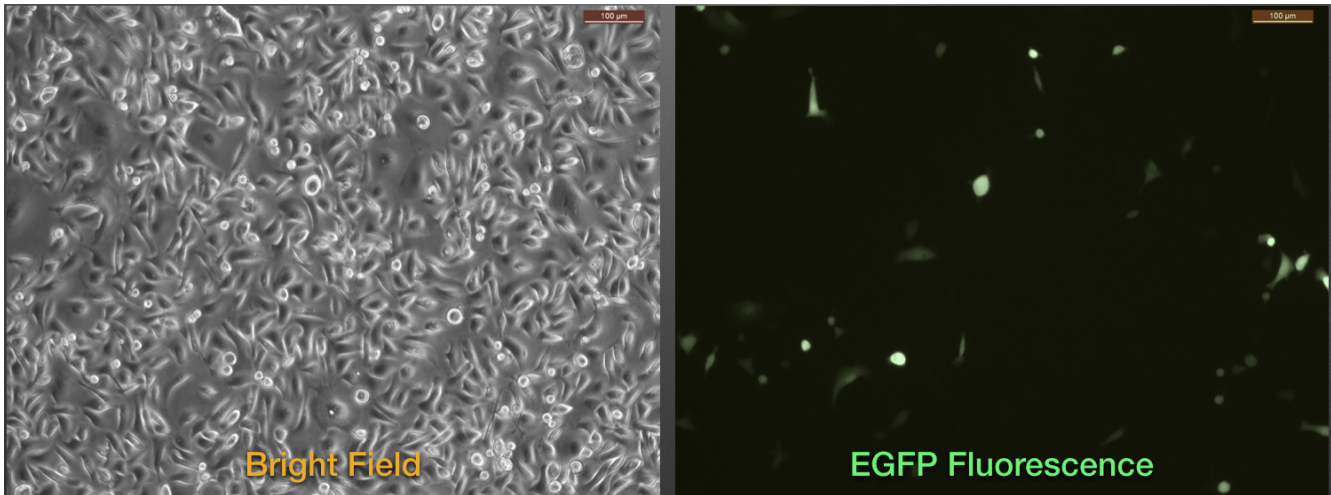


Figure1: P29 Keratinocytes, 24 hours after transfection with C1 CRISPR-Cas9 plasmid construct. Based on flow cytometry, 2.5% of these cells showed high enough fluorescence to be considered EGFP+.

For each Let7 cluster cells were sorted singly into three 96 well plates. Eight days after sorting 3/288 C1 cells and 2/288 C2 cells had produced colonies. These were moved to a 24 well plate and allowed to grow for another 7 days (Figure 2). At this point one C2 colony had become contaminated; the three C1 colonies were collected for genotyping, and the remaining C2 colony was allowed to continue to grow. However, it too died before genotyping. There are several reasons why so few clonal colonies formed. Sorting may have killed some of the cells, such that they were plated dead. If that were the case, successive dilutions instead of MoFlo sorting may have improved survival. They may have also struggled to grow in isolation, and may have thrived more readily had they been sorted into conditioned media from feeder cells. The 3 C1 clonal colonies that survived post sorting were genotyped, and found to have no observable genome modifications using the plasmid specific KO primers (Figure 3). No C2 colonies were genotyped, as one became contaminated and the other did not past what is seen in figure 2 before dying.

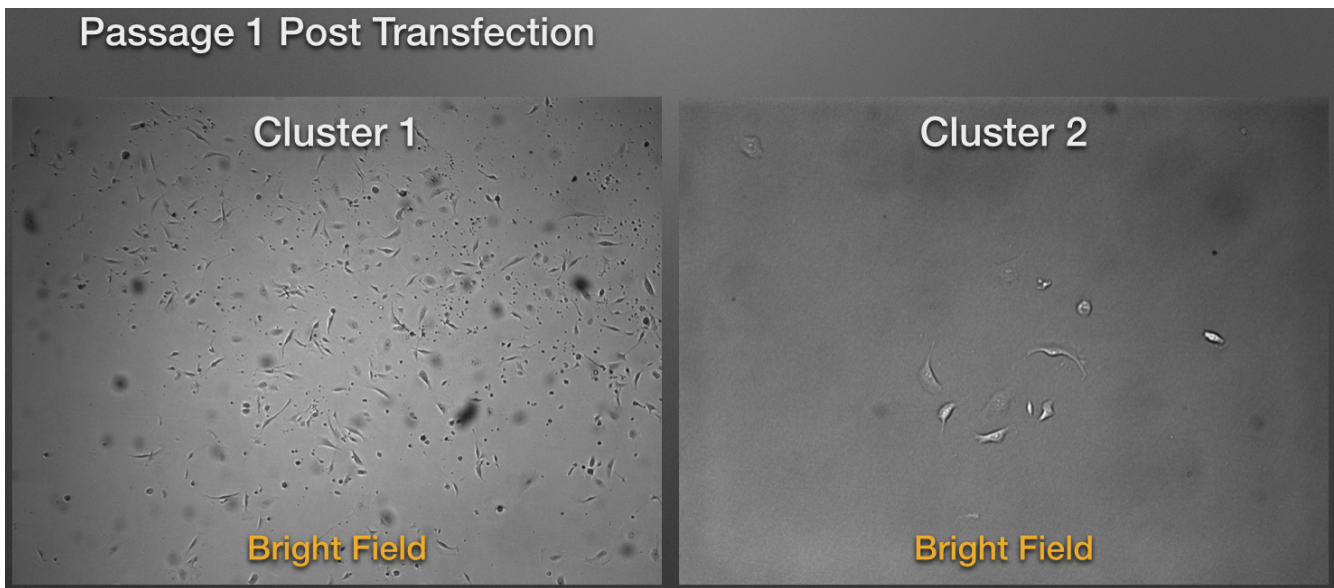


Figure 2: C1 and C2 clonal colonies. All three C1 colonies looked this strong, and were passaged once before this picture. Both C2 colonies looked as weak as this picture shows, and did not progress past this point before dying.

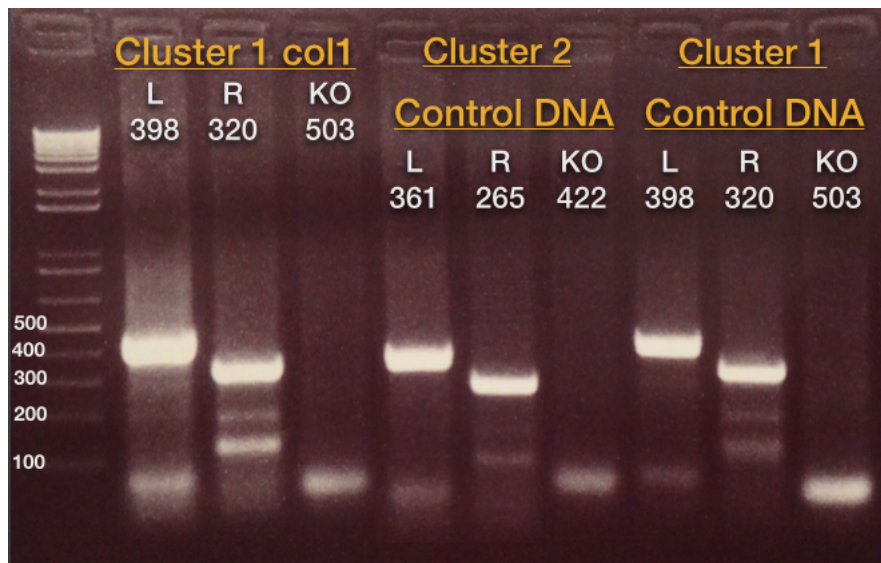


Figure 3: Clonal colony PCR on Let7 cluster 1 colonies. All three resembled the shown genotyping, with no KO band. C2 colonies were not genotyped before they died, but clean genomic DNA was used to check the PCR primer pairs, yielding positive controls for the L and R bands and negative controls for the KO bands.

Transfection was attempted twice more, and cells were genotyped without sorting. The first round of transfection was done in cells that were p28. These yielded no discernible KO band 24 or 48 hours post transfection. However, when p17 cells were woken and transfected were visible in the genotyping done on non-sorted cell lysate. Cell samples were taken from each population at 48 and 67 hours post transfection. C1 cells show a very faint band at 48 hours and a stronger band at 67 hours. C2 cells show distinct bands at both time points (Figure 4).

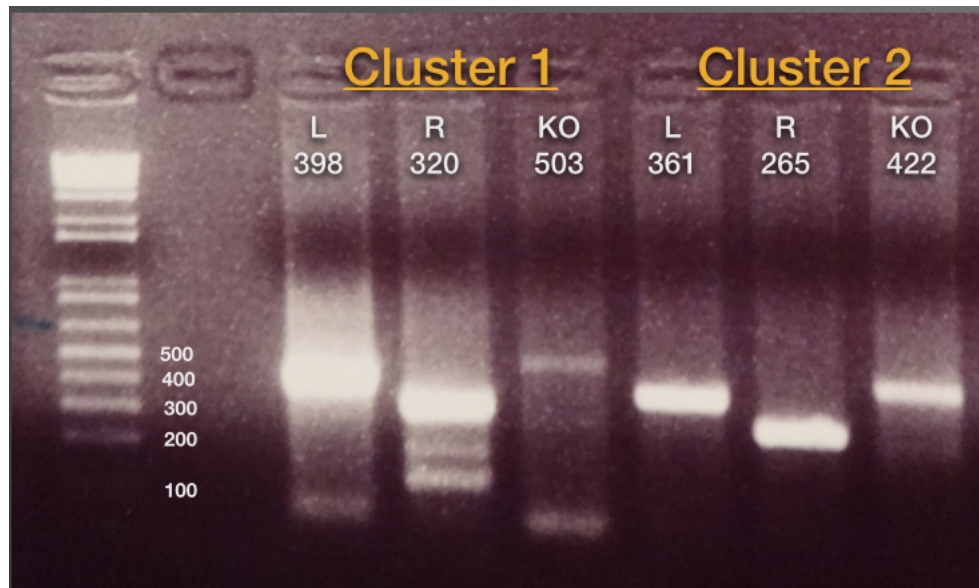


Figure 4: Unsorted keratinocytes, transfected with C1 or C2 plasmid, were allowed to grow for 67 hours before being collected and genotyped. Left and Right bands are prominent, but KO bands for each cluster are clear as well. This indicates that the CRISPR constructs for Let7 cluster 1 and cluster 2 were functional.

These results were promising, and when this project is continued my experimental design will be better. I will transfect a control cell population with a plasmid carrying only human promoters, to have a background of any phenotypic effects the process of transfection might cause. Experimental wells will include: transfection with C1 and C2 plasmids concurrently; C1 alone; C2 alone; and Cas9 plus all four sgRNAs. The single plasmid wells will be sorted and EGFP+ cells will be genotyped and replated before transfection with the alternate plasmid. Each condition will be attempted with both Transit-LT1 and Lipofectamine 2000, and Opti-MEM media will be used for all complex formation. Cells will be left for a minimum of 67 hours before they are sorted or genotyped.

APP in Lymphoblastoid Cells

As mentioned previously, the human lymphoblastoid cells used in this study were taken from four members of a family participating in Down Syndrome research. The proband is a 3 year old boy; family members include his 5 year old brother and their parents. Having cells from an entire family allows us to see the gene and protein expression in trisomy 21 cells against a familial background. This makes it easier to delineate those phenotypes that are part of the normal variation between individuals from those that play a role in the symptoms of Down Syndrome. The cells were collected from blood

samples and immortalized with Epstein Barr Virus (EBV), in order for them to survive long term in culture. This infection causes primary B lymphocytes in G_0 to enter the cell cycle and proliferate indefinitely. For this to take place, the cells must express proteins from the viral genome, which exists as a multi-copy plasmid in the cells. These proteins include nuclear antigens and latent membrane proteins. Overall, immortalization causes these cells to phenotypically resemble activated B cells, which should not confound our experiments.

Given that the lymphoblastoid cells possess an extra copy of each chr21 gene, a baseline hypothesis would be that each of these genes should be expressed at 1.5x the rate seen in chromosomally normal individuals. This does not take into consideration gene interactions or mechanisms that might exist to regulate gene dosage, but it is a place to start. Dr. Garrido-Lecca performed small RNA sequencing, mRNA/miRNA expression profiling and ribosome profiling on the lymphoblastoid cells from all 4 family members. In comparing the proband to his brother or father, expression levels of multiple protein coding genes and regulatory RNAs were discrepant. These differences were scattered throughout the genome, but were most notable in chr21 genes. One gene whose mRNA production was strikingly up-regulated in the proband is amyloid precursor protein (APP). This mRNA was present at 7x the expected level, though this result was attenuated by the fact that the father also showed higher than normal levels of APP mRNA. This result has been seen in other DS studies as well (Ness et al, 2012); Somani et al (2014) found that people with DS had increased plasma levels of amyloid-beta peptides. Additionally, they found that a segment of repeats in the 7th intron of APP was capable of modifying the age of AD onset in patients with DS, and the progression of their dementia. The prevalence of different APP splice variants may also play a roll in this pathogenesis. Zigman and Lott (2007) also discuss a case study in which a patient with partial trisomy 21 was disomic for APP, and developed no symptoms of dementia or AD.

Based on this background, my goal in working with the trisomic lymphoblastoid cells was twofold. First, APP appeared to be a desirable target for a CRISPR/Cas9 knockout study, given its high expression in DS and in lymphoblast cells, and its possibly causal role in AD. However, a full knockout of APP would not actually be as informative as a partial knockout. We know the function of APP in normal cells, and we know the phenotype of immortalized cells with three copies of APP. What we don't know is how much of that phenotype is due to APP over expression, and what is causing that overexpression. Because of this, knocking out one copy of APP would be most informative. This is not

likely to happen using a master plasmid, as they can persist for days in the nucleus and would likely result in the complete removal of APP alleles, if not additional off-target effects. Instead, we hoped that inducing transient transfection with RNA would produce less than a full knockout. This coincided with our desire to use RNA transfection to avoid off-target cleavage. Unlike a plasmid which can be stable in cell nuclei for multiple weeks, RNA is degraded fairly rapidly in the cytoplasm. This is especially true for the short sgRNAs which have neither 5' caps nor poly A tails. Limited data is available regarding the half-life of exogenous RNA in eukaryotic cytoplasm, but some estimates are as short as 12 to 18 hours. Since there is less chance of off-target DSBs when using short lived RNA, optimization can focus on the extent of on-target editing that can occur before degradation. This might mean changing the amounts of sgRNA or mRNA transfected at once, or transfecting the same cells multiple times.

Transfection into the lymphoblastoid cells using mRNA and sgRNAs failed to result in EGFP production, as did subsequent transfection with a Cas9/EGFP plasmid and sgRNAs. We did this because the plasmid had been verified as functional in other cellular systems, and our mRNA had not. This failure, along with a visually undetectable level of cell death, indicated that the cells were rather impervious to Lipofectamine 2000.

We next tried to introduce nucleic acid to the cells via electroporation with the Neon syringe system. We again used the Cas9/EGFP plasmid as we were unable to verify the functionality of our mRNA. We also reasoned that even if the plasmid lingered in the cell, the introduction of additional DSBs would be impossible after the sgRNAs had been degraded. We chose electroporation protocols optimized for cell types similar to ours, as detailed above. This also resulted in no transfection and little noticeable cell death.

Our next step will be to attempt to optimize a Neon protocol for our specific cells.

APP in HEK293T

Because transfection of the lymphoblastoid cells was not progressing, and creation of novel sgRNA is a time intensive procedure, I wanted to attempt transfection in a different human cell type to verify the genome editing capacity of my sgRNAs. For these tests we chose to use HEK293T cells which are sturdy, fully differentiated, easy to culture and amenable to chemical transfection. These cells would also allow us to test the functionality of our mRNA, since the verified Cas9/EGFP plasmid should

readily transfect into HEK293T.

These cells are adenovirus transformed and have a karyotype that is far from normal. They are described as hypo triploid, having a mode of 64 chromosomes, with two or more copies of each chromosome, typically including 3 copies of X. These idiosyncrasies do not confound my experiment, and may even be more useful than originally thought, should my culture end up having trisomy 21. If they do, any knockout signal has the potential to be stronger than in typical cells, with three possible knockout alleles.

An initial transfection was set up comparing three levels of Lipofectamine 2000 across wells transfected with Cas9/EGFP plasmid, and wells transfected with Cas9/EGFP mRNA. Three days post transfection this assay clearly showed that our mRNA was not functional, as all three plasmid wells had EGFP production, which increased proportionally with the volume of Lipofectamine 2000 used. None of the mRNA wells showed EGFP (Figure 5).

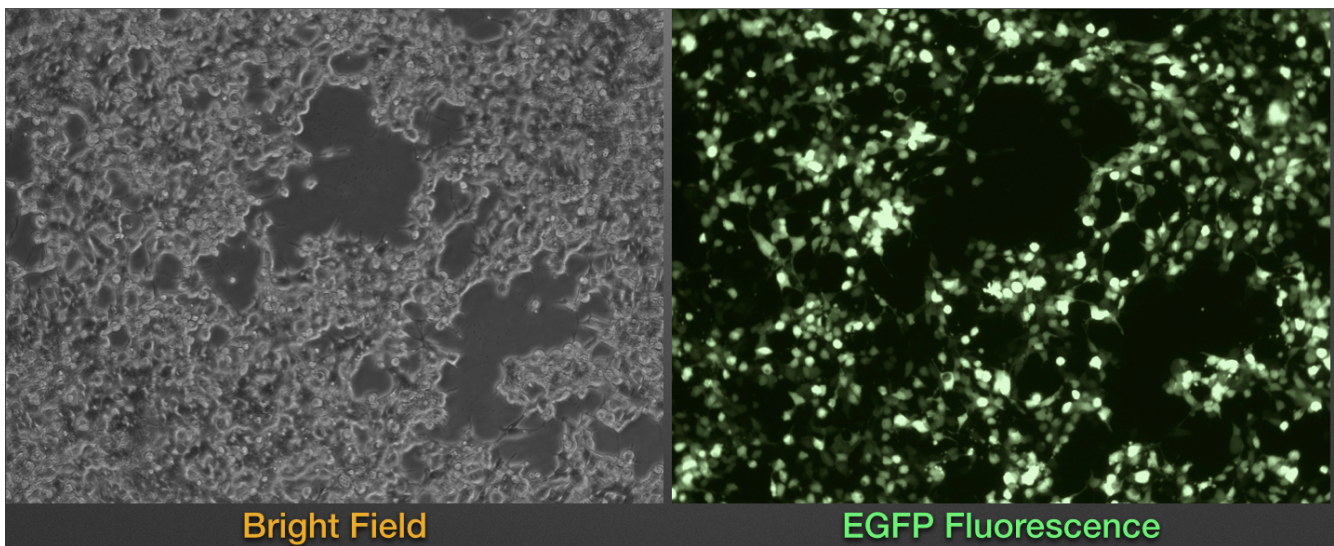


Figure 5: HEK293T cells, 60 hours post transfection with Cas9 plasmid and sgRNAs in Lipofectamine 2000. Strong transfection efficiency is seen.

Plasmid transfected cells were collected without sorting and genotyping was performed. Results appeared to show the presence of all three bands, implying that editing had occurred in some cells. However, many nonspecific bands were seen (Figure 6). A larger scale transfection was carried out in 10cm dishes, using the same DNA/RNA ratio, and highest level of Lipofectamine 2000, scaled up. 72 hours post transfection, a high percentage of each plate was fluorescent. Cells were almost confluent and were split in half. 24 hours later, the cells were gated for size and high EGFP expression. 18% of

the cells met these criteria; they were sorted individually into 6 96 well plates, and en masse into a collection tube. Two collection tubes, each with 880k cells, were collected. These cells were lysed and used for genotyping. The single cell plates were allowed to grow for six days and then were scanned for colonies. A total of 37 clonal colonies grew. Despite the positive readouts on the cell sorter, only three plates appear to have received cells, as the other three were devoid of colonies.

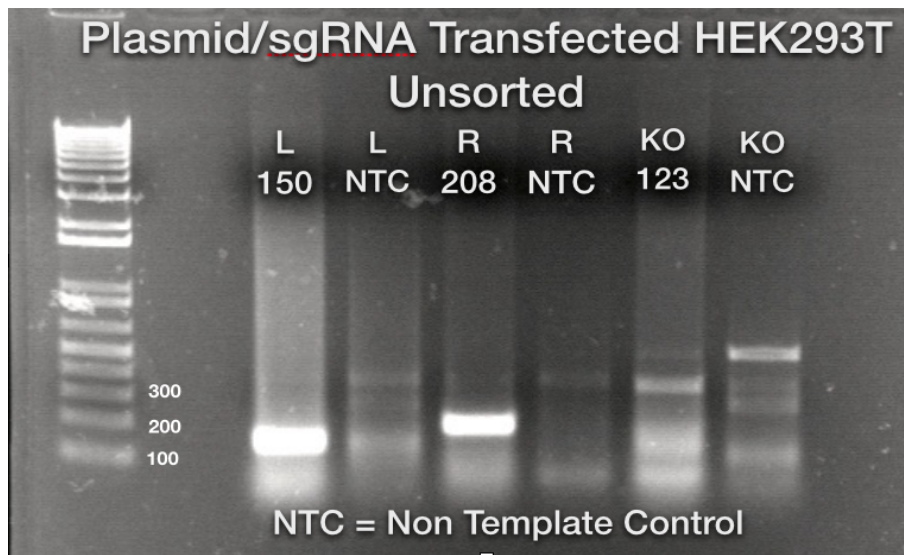


Figure 6: APP genotyping PCR with primers to check L, R and KO fragments in transfected, unsorted HEK 293 T cells. Odd amplification bands in KO NTC were worrisome.

These colonies showed large variations in their growth rates. Some were confluent at the 6 day point, while others never covered more than 1/3 of their well surface in a 96 well plate. Colonies were plated in 24 and then 12 well plates as they grew, and samples of each colony were taken for genotyping. Regardless of initial growth rates, colonies grew more slowly as they were plated into larger wells, after their second plating colonies began to die. Some may have been killed because the media was not changed soon enough to remove the trypsin, but even colonies replated without trypsin (as the cells adhered only lightly to the plates) eventually died. Three weeks after single cell sorting, all clonal colonies had died.

Genotyping on the EGFP⁺ cell lysate appeared to show strong bands from all three PCR reactions, indicating that some cells had in fact been successfully edited (Figure 7). These PCR bands were extracted from the agarose, reamplified and sent for sequencing to verify their identity. However,

sequencing failed as there was no priming. I then took the remaining amplicon and digested it using the restriction enzyme BseI, one of the only enzymes able to cut the 123 bp KO amplicon. This should have produced fragments of lengths 94 and 29, and cleavage does appear to have happened correctly (Figure 8).

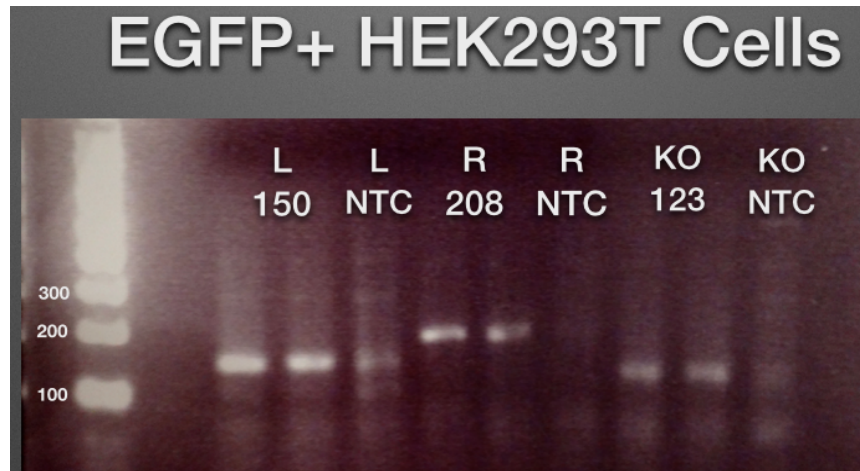


Figure 7: APP Genotyping on two samples of EGFP+ sorted cells appears to show an intact KO band. Expected lengths are listed on the figure.

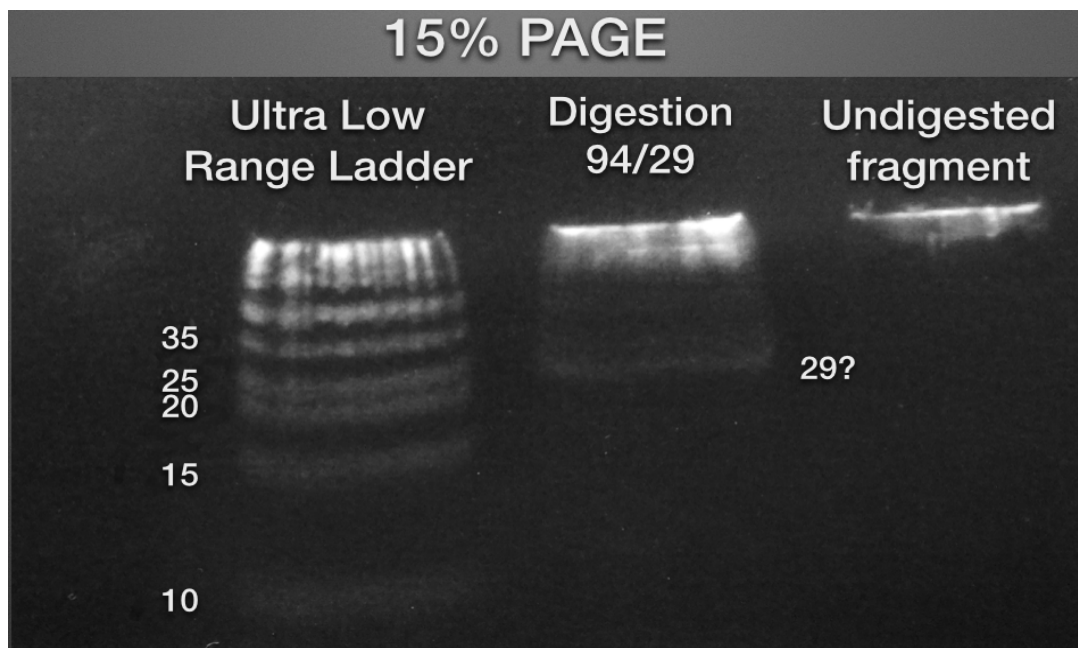


Figure 8: Digestion of putative APP KO band with BseI enzyme. Restriction enzyme should give bands of 94 and 29 base pairs, and appears to have worked.

When clonal colonies were genotyped, all colonies came up positive for all three bands, and the non template control reactions showed what appeared to be very strong amplicons as well (Figure 9). This indicated that there was nonspecific binding occurring with the primers, or that contamination had occurred somewhere. After I optimized the PCR protocols for these primers, as described above (Figures 10a, 10b, 10c), I genotyped the clonal colonies again. This time, none showed KO amplicons (Figure 11). However, some showed no left or right amplicons either, indicating that the DNA was either very low concentration, degraded, or PCR was being inhibited by salts in the cell lysate. This same issue occurred when I re genotyped the non-clonal EGFP+ cell lysate with the optimized PCR protocol. I then purified the lysate with phenol chloroform and precipitated the DNA with ethanol. I will redo the PCR on this sample in the future. It may be the case that the crRNA targets I have chosen are inaccessible due to chromatin density or other epigenetic factors, and new crRNA targets may need to be designed.

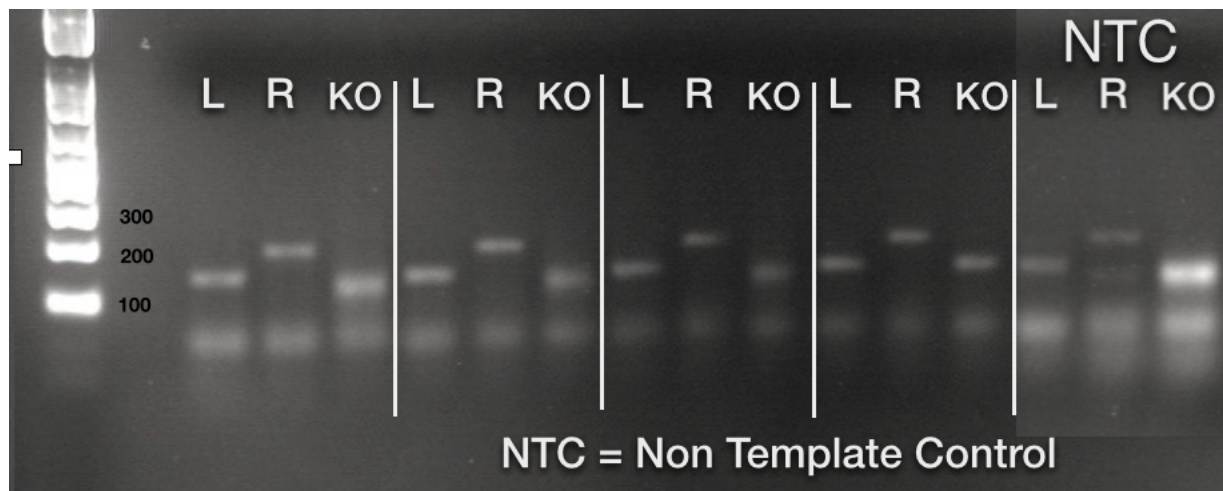


Figure 9: HEK293T APP genotyping of clonal colonies. Right three lanes are L1sht, R2, and KO NTCs, which all show bands. This is also 35 PCR cycles, and should show better primer depletion. The fact that the colonies all show the same genotype is suspicious, and the NTC bands indicate contamination or poorly optimized primer conditions.

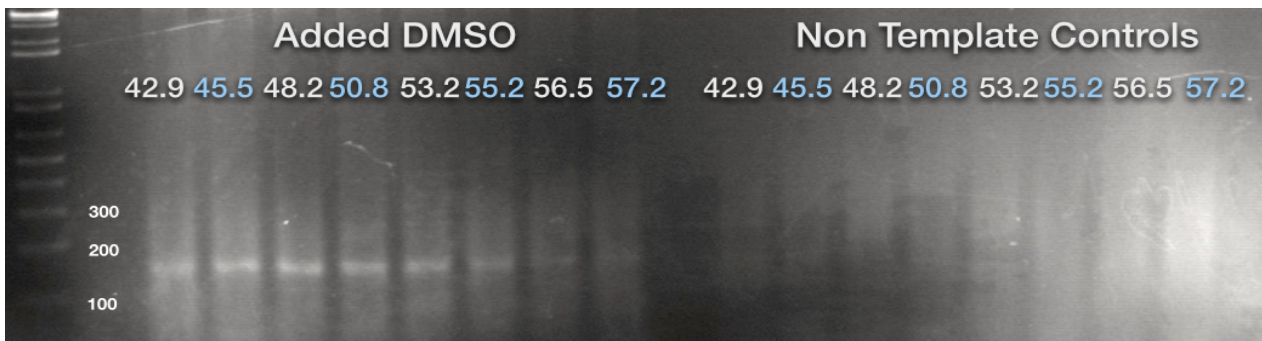


Figure 10a: Optimization of L1sht PCR as it was the more troublesome one of the two pairs when used on clean control DNA. These 25ul reactions using IPSc DNA each had 1ul DMSO added. 8 lanes of varied annealing temperature, from 42.9 to 57.2 degrees. 8 additional lanes of NTC on the same temperature gradient.



Figure 10b: Continued L1sht PCR optimization, with 1ul $MgCl_2$ per 25 ul reaction. Same temperature gradient as above. Chose to focus on 2nd and 3rd annealing temperatures (45.5 and 48.2) as they have strong amplification but no discernible NTC band.

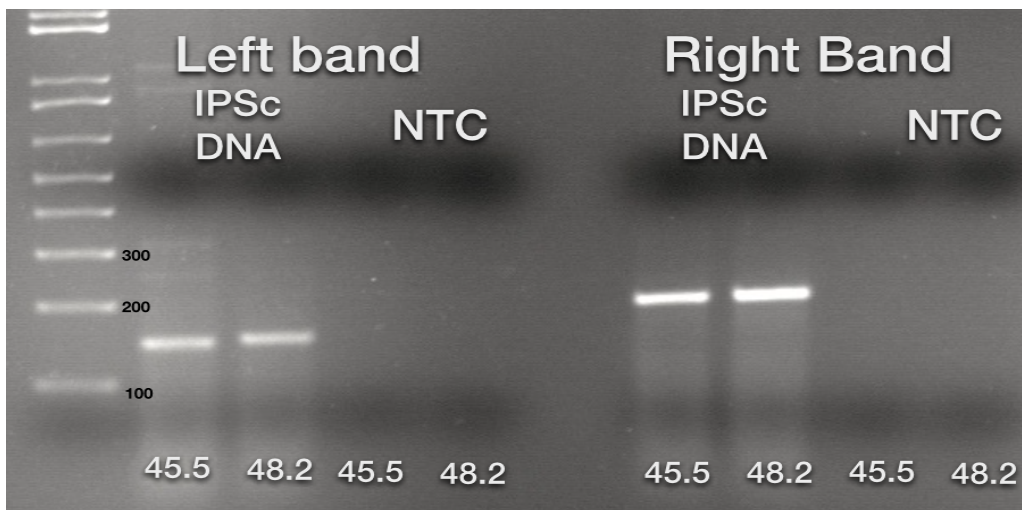


Figure 10c: Further optimization of APP KO primers using 5x taq mix containing $MgCl_2$, KCl, and stabilizers, in an attempt to decrease the smearing seen in previous optimizations. L1sht at 45.5 and 48.2, NTCs for those temps, R2 and NTCs at those temps. Elongation time dropped to 68 degrees from 72, per reagent protocol, and final elongation step was 5 minutes instead of 1.



Figure 11: Post optimization, 4 clonal HEK293T colonies genotyped with APP KO primers. Each colony is L1sht, R2, and KO bands, three positive control reactions (IPSc DNA) and three NTC reactions. The positive amplicon in the intact IPSc DNA is of unknown origin, but it is useful as a positive KO primer control. None of the colonies show that editing has occurred.

Conclusions

For easily transfected cell types, using one large plasmid carrying Cas9 and both sgRNAs is advantageous, since the reporter protein indicates the presence of all three CRISPR molecules in the cell. However, its constitutive expression makes the plasmid problematic, as it may be retained in the nucleus for weeks and wreak off-target havoc. Transfecting cells with RNA may be the best option for avoiding off-target editing events. Also, inserting nucleic acid into a difficult to transfect cell type may be aided by using smaller molecules. For these reasons, investing time in generating functional NLS/Cas9/NLS/T2a/EGFP may be a worthwhile task. If this process does not end up being reliable, retaining this molecule on the smallest plasmid possible and transfecting it along with sgRNAs seems like a reasonable alternative. Either way, the experimental timeline probably needs to be tweaked.

One concern is transfection efficiency when all three molecules are added to the cells at once. Even if each molecule has a 50% transfection rate, which is quite high, on average only one in eight cells would receive all three nucleic acids and have a chance to undergo editing. This fraction drops precipitously if transfection rates are lower, and is decreased further in single cell sorting since not all EGFP+ cells will pass the criteria for the sorting machine. In my HEK293T cells 18% of the cells

passed the sorter's criteria. Assuming this means even 25% of the cells were EGFP+, that implies that on average one in sixty-four cells would have taken up all three components. This is an issue for either form of Cas9-EGFP. When transfection is done with a plasmid and two sgRNAs, a temporal issue is introduced as well. The sgRNAs are short with no cap and no poly A tail, which means they are susceptible to rapid degradation. The plasmid on the other hand must enter the nucleus for transcription of Cas9/EGFP and its mRNA must be exported and translated before the protein can complex with the sgRNAs. Because of this time delay in the production of Cas9, much if not all of the sgRNAs might be degraded before they can be used, meaning that even if all three molecules enter one cell, there may still be no opportunity for genomic editing.

To address both issues, my future experiments will involve two transfections separated by cell sorting. Cas9/EGFP plasmid or mRNA will be introduced to the cell and sorting will be done after EGFP signal is robust. EGFP+ cells will be replated, allowed to grow to the appropriate density, and then transfected with sgRNAs. Even at a 20% transfection rate genome editing would then be possible in about 1 in 25 cells. If successful, this strategy could be optimized temporally for the best editing rates.

Once I can reliably modify the genomes of experimental cells and am ready to assess changes in cellular phenotype related to gene knockout, all subsequent experiments should include a transfection-effect control population. Since transfection of any type is stressful to the cells, this will necessarily change their transcription and translation profile independent of any genome editing that may occur. Therefore, I will have a population of cells that is transfected with matching types and sizes of nucleic acid which are incapable of genomic editing. For example, I will transfect murine keratinocytes with plasmid carrying human promoters and sgRNAs that target the human genome.

Both of these CRISPR projects are promising, and will be continued. Additional modifications to my experimental protocols may also prove useful. Because the lymphoblastoid cells have proven intractable, we may move on to attempting transfection on other trisomy 21 cells. Another option is to use viral insertion of the RNAs into cells.

References

- Albrekt, Ann-Sofie, et al. "Skin sensitizers differentially regulate signaling pathways in MUTZ-3 cells in relation to their individual potency." *BMC Pharmacology and Toxicology* 15.1 (2014): 5.
- Bire, Solenne, et al. "Exogenous mRNA delivery and bioavailability in gene transfer mediated by piggyBac transposition." *BMC biotechnology* 13.1 (2013): 75.
- Cardenas, Christian L. Lino, Naftali Kaminski, and Daniel J. Kass. "Micromanaging microRNAs: using murine models to study microRNAs in lung fibrosis." *Drug Discovery Today: Disease Models* 10.3 (2013): e145-e151.
- Cho, Seung Woo, et al. "Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease." *Nature biotechnology* 31.3 (2013): 230-232.
- Cong, Le, et al. "Multiplex genome engineering using CRISPR/Cas systems." *Science* 339.6121 (2013): 819-823.
- Fu, Yanfang, et al. "Improving CRISPR-Cas nuclease specificity using truncated guide RNAs." *Nature biotechnology* 32.3 (2014): 279-284.
- Kim, Tae Kyung, and James H. Eberwine. "Mammalian cell transfection: the present and the future." *Analytical and bioanalytical chemistry* 397.8 (2010): 3173-3178.
- Jinnin, Masatoshi. "Various applications of microRNAs in skin diseases." *Journal of dermatological science* 74.1 (2014): 3-8.
- Johnson, Charles D., et al. "The let-7 microRNA represses cell proliferation pathways in human cells." *Cancer research* 67.16 (2007): 7713-7722.
- Johnson, Steven M., et al. "RAS Is Regulated by the let-7 MicroRNA Family." *Cell* 120.5 (2005): 635-647.
- Hsu, Patrick D., et al. "DNA targeting specificity of RNA-guided Cas9 nucleases." *Nature biotechnology* 31.9 (2013): 827-832.
- Matsumoto, Akira, and Yoshisada Fujiwara. "Aberrant proteolysis of the β -amyloid precursor protein in familial Alzheimer's disease lymphoblastoid cells." *European Journal of Biochemistry* 217.1 (1993): 21-27.
- Minskaia, Ekaterina, and Martin D. Ryan. "Protein coexpression using FMDV 2A: effect of "linker" residues." *BioMed research international* 2013 (2013).
- Ness, Seth, et al. "Down's syndrome and Alzheimer's disease: towards secondary prevention." *Nature Reviews Drug Discovery* 11.9 (2012): 655-656.
- Nižetić, Dean, and Jürgen Groet. "Tumorigenesis in Down's syndrome: big lessons from a small chromosome." *Nature Reviews Cancer* 12.10 (2012): 721-732.
- Phua, Kyle KL, Kam W. Leong, and Smita K. Nair. "Transfection efficiency and transgene expression kinetics of mRNA delivered in naked and nanoparticle format." *Journal of Controlled Release* 166.3 (2013): 227-233.
- Ran, F. Ann, et al. "Genome engineering using the CRISPR-Cas9 system." *Nature protocols* 8.11 (2013): 2281-2308.
- Sander, Jeffrey D., and J. Keith Joung. "CRISPR-Cas systems for editing, regulating and targeting genomes." *Nature*

biotechnology 32.4 (2014): 347-355.

Somani, Alyma A., et al. "APP INTRON 7 polymorphism and plasma amyloid-beta peptide levels can predict the onset of dementia in down's syndrome." *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 10.4 (2014): P786.

Wilcock, Donna M., and W. S. Griffin. "Down's syndrome, neuroinflammation, and Alzheimer neuropathogenesis." *J Neuroinflammation* 10 (2013): 84.

Wu, Xuebing, et al. "Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells." *Nature biotechnology* (2014).

Zigman, Warren B., and Ira T. Lott. "Alzheimer's disease in Down syndrome: neurobiology and risk." *Mental retardation and developmental disabilities research reviews* 13.3 (2007): 237-246.