1-2019

Targeted Resequencing of CRISPR Cas9 Mediated ICER Knockout in SK-MEL-24 Cells

Justin Wheelan
Montclair State University

ABSTRACT

Inducible cAMP Early Repressor (ICER) is a small transcription factor that originates from an intronic promoter within the cAMP Response Element Modulator (*CREM)* gene (Molina et al., 1993). ICER acts as a putative tumor suppressor by mediating cAMP antiproliferative activity by competitively binding to cAMP Response Elements (CRE's) and repressing transcription of genes involved in cell division (Mémin et al., 2002). ICER has been shown to be effectively absent in cancer and suspected to be targeted for proteasomal degradation via ubiquitination (Healey et al., 2013). ICER is most likely regulated via post-translational modifications as a result of mutations on *Ras/Raf* oncogenes (Healey et al., 2013). For example, it has been shown that in cancer, mutant activated GTP bound Ras protein continuously activates Braf, which in turn, results in over activation of the Mitogen Activated Protein Kinases, ERK1/ERK2 (Zhang and Liu, 2002). ICER is thought to be phosphorylated by ERK1/ERK2 and subject to proteasomal degradation; a result of ubiquitination in Ras/MAPK-mediated melanoma tumorigenesis (Healey et al., 2013). Melanoma cells rapidly apoptosed when transfected with a mutant form of ICER that did not contain any lysine residues. This was most likely because the transcriptional repressor ICER was unable to be ubiquitinated. While this highlights ICER's importance in cell division and growth, it is challenging to study the effects of ICER in melanoma if cells are unable to survive. One method to circumvent this issue is to create an inducible cell line, in which a mutant form of ICER with no lysine residues is under a promoter whose expression can be toggled on and off, dependent on the presence of a transactivator. This requires that mutated ICER be knocked-in to an alternative location in the human genome, and that wildtype ICER be

efficiently knocked-out in order to solely study the effects of the mutant. In this experiment, CRISPR Cas9 mediated genetic editing was used to knockout ICER, while mitigating off target effects on *CREM* gene, in an attempt to maintain otherwise normal cell physiology. The target loci of interest in this experiment was the Kozak consensus on an ICER specific promoter and the target of the guide RNA (gRNA) for Cas9 endonuclease activity. Amplicon Libraries of DNA extracted from cells transiently transfected with plasmid containing gRNA and Cas9-GFP cassette and empty vector expressing only EGFP (control) were generated with Nextera index adaptors. Paired-end sequencing on the Illumina Miseq provided sufficient coverage depth to determine how efficient the gRNA was at generating insertions/deletions (indels) or substitutions at the desired loci. Sequencing data between Experimental and Control was first reviewed using CRISPResso, an online bioinformatic tool that analyzes deep sequencing data. An additional bioinformatic analysis was designed and performed to corroborate CRISPResso results and identify any other low-level variants.

The goal of this experiment was to develop the workflow to identify possible indels/substitutions that resulted from CRISPR Cas9 induced genetic alteration. This was done with the expectation of identifying a possible knockout of ICER, while minimally affecting CREM. Although no variants identified suggest an ICER knockout, a scalable workflow is now in place to facilitate this stage of the experiment.

MONTCLAIR STATE UNIVERSITY

TARGETED RESEQUENCING OF CRISPR-CAS9 MEDIATED ICER KNOCKOUT IN SK-MEL-24 CELLS

by

Justin Wheelan

A Master's Thesis Submitted to the Faculty of

Montclair State University

In Partial Fulfillment of the Requirements

For the Degree of

Master of Science

January 2019

College/School College of Science and Mathematics

Department Biology and Molecular Biology

Thesis Committee:

[redacted]

Carlos A. Molina, Ph.D.
Thesis Sponsor

[redacted]

Robert Meredith, Ph.D.
Committee Member

[redacted]

Mitchell Sitnick Ph.D.
Committee Member

TARGETED RESEQUENCING OF CRISPR-CAS9 MEDIATED ICER KNOCKOUT IN SK-MEL-24 CELLS

A THESIS

Submitted in partial fulfillment of the requirements

For the degree of Master of Science

by:

JUSTIN WHEELAN

Montclair State University

Montclair, NJ

January 2019

Acknowledgments

My thesis sponsor Dr. Carlos Molina

NSF (DBI1725932 to Robert Meredith, John Gaynor, Sandra Adams, Chunguang Du, Kirsten Monsen)

My thesis advisors: Drs Robert Meredith and Mitchell Sitnick

My colleagues in the laboratory Angelo Cirinelli and Keith Lange

My wife, family and friends

TABLE OF CONTENTS

LIST OF FIGURES

Introduction

Inducible cAMP Early Repressor (ICER) is a small transcription factor that originates from an intronic promoter within the cAMP Response Element Modulator (*CREM)* gene (Molina et al., 1993). ICER acts as a putative tumor suppressor by mediating cAMP antiproliferative activity (Mémin et al., 2002). It has been demonstrated that ICER arrests cells at G1/S and G2/S checkpoints of the cell cycle and thus inhibits cell growth and division (Razavi et al., 1998). Cytosolic factors from the cAMP pathway phosphorylate cAMP Response Element Binding Protein (CREB) and CREM, activating transcription of CRE (cAMP Response Element) containing genes. ICER regulates cell division by competitively binding to CRE's and preventing CRE-mediated gene transcription. ICER, as an antagonist to CREB and CREM, prevents expression of proteins critical for cell division, such as cyclin A, cyclin D and c-fos (Mémin et al., 2011). ICER, itself, is a product of CRE's and therefore can regulate its own expression via cAMP driven gene expression (Yehia et al., 2001). In normal cell physiology, this autoregulation of a protein involved in mitosis is advantageous and helps maintain cell division regulation and homeostasis.

While ICER's importance in regulating cell division is apparent, it has also been shown that ICER is effectively absent in melanoma cells (Healey et al., 2013). This makes ICER a protein of interest with regard to cancer research and possible targeted therapeutic mechanisms. One explanation for ICER's apparent absence in cancer is that ICER is targeted for proteasomal degradation (via ubiquitination), a result of Ras/MAPK-mediated melanoma tumorigenesis (Healey et al., 2013). In cancer, common mutations

on the *RAS* gene for example, results in activated Ras protein that continuously activates Braf (Zhang and Liu, 2002). Overactive Braf activates the Mitogen Activated Protein Kinases (MAPK'S) Extracellular Signal-Related Kinases (ERK1/2) at much higher rate than in non-diseased cells (Zhang and Liu, 2002). One function of the proteins ERK1 and ERK2 is that they phosphorylate ICER at a critical Serine (Ser 41). In experiments performed in the lab of Dr. Molina, in which the Ser41 on ICER is mutated, the half-life of ICER is 4-5 hours longer than the wild types. This suggests that phosphorylation of this serine is required for ICER to be efficiently ubiquitinated and targeted for destruction. Further, ICER can also be phosphorylated by CDK1 at Ser 35, which results in a mono-ubiquitinated ICER, and re-localization of ICER to the cytoplasm (Mémin et al., 2011). This behavior is increased in cancer cells and suggests a rapid mechanism to deregulate ICER and promote cell division. Additionally, mono-ubiquitinated ICER could have an unknown secondary function as a result of this relocation. The idea that ICER's tumor suppressor capability is diminished in melanoma via post-translational modification is supported by static mRNA expression (Healey et al., 2013

It has been shown that ubiquitination occurs primarily on lysine residues of protein targets (Mattiroli and Sixma 2014). Current unpublished research at the time of this thesis from Dr. Molina's laboratory suggests that melanoma cells rapidly apoptosed when transiently transfected with a plasmid expressing a form of ICER in which all lysine residues were converted to arginine (NKO-ICER). This further supports the claim that ubiquitination is responsible for ICER degradation in these diseased cell states. While the mechanism for destruction of ICER is somewhat understood, what is less clear is how ICER relates to a tumorigenic phenotype. In order to study the effects of ICER on

cancer cell growth in more detail it is necessary to control NKO-ICER expression so cells remain viable and not immediately undergo apoptosis.

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) and the CRISPR associated endonuclease 9 (Cas9), functions in nature as an adaptive immune system for bacteria and archaea (Deveau et al., 2008). Although these repeat sequences originally caught the attention of researchers in 1987, the function of the spaces between the repeats as a bacterial adaptive immune system was not recognized until 2007 (Horvath et al., 2010, Barrangou et al., 2007). In 2012, Jennifer Doudna and others published a landmark paper describing how to utilize CRISPR Cas9, as a programmable tool to permanently alter the genome (Hsu et al., 2014). Since this time, a flood of research has been done using the CRISPR Cas9 system to knockout genes in a variety of organisms, including cancer cells (Jafari et al., 2017).

Genome editing mediated by CRISPR Cas9 relies on two cellular repair mechanisms; Non-Homologous End Joining (NHEJ) and Homology Directed Repair (HDR) (Ran et al., 2013). The former is typically favored when generating a knockout because, as the cell attempts to repair the double-stranded break through NHEJ, mutations arise that can permanently alter DNA sequence, and thus protein structure/function (Ran et al., 2013). NHEJ is initiated by a protein named Ku that recognizes double stranded DNA breaks and attempts to form a bridge between the two ends to prevent degradation (Pastwa and Błasiak, 2003). Severed double stranded DNA must be blunted in order to be ligated together. This occurs by DNA synthesis via DNA polymerase or alternatively DNA overhangs are removed by exo-nuclease (Pastwa and Błasiak, 2003). The DNA blunting step is particularly prone to generating

insertions/deletions (indels) or Single Nucleotide Polymorphisms (SNP's). Once the severed ends of the DNA are blunted, DNA ligase attaches both ends of DNA (Pastwa and Błasiak, 2003). This error-prone process has been used in many different applications to generate efficient knockouts.

To facilitate future experiments using NKO-ICER, we devised a plan to insert NKO-ICER into the genome under the regulation of an inducible promoter. Specifically, the central component will be to knock-*in* NKO-ICER into what is known as a safe-harbor site. The safe-harbor site, also known as Adeno Associated Viral Integration Site (AAVS1), is a locus in the human genome that has been shown to be utilized in transgene integration (Tiyaboonchai et al., 2014). One advantage of a knock-in via AAVS1 transgene integration is a dramatic reduction in gene silencing or negative epigenetic effects of a traditional knock-in at the endogenous loci (Tiyaboonchai et al., 2014). However, it will not be sufficient to simply integrate NKO-ICER into the safe-harbor location because it would be constitutively expressed and result in rapid apoptosis of the cells. One solution offered by moving NKO-ICER to the safe-harbor location is that NKO-ICER can be regulated by an inducible tetracycline promoter, otherwise known as a *Tet-on* system (Takarabio Bio, US). This allows for cells to be treated with doxycycline (a tetracycline analog) to trigger expression of NKO-ICER in a controlled fashion. Similar methods have been demonstrated to induce expression of a transgene in human pluripotent stem cells *in-vitro*, but to our knowledge has not been demonstrated in tumorigenic cell types, let alone malignant melanomas.

One main challenge is that even with successful integration of NKO-ICER into the AAVS1 safe-harbor region under an inducible *Tet-on* promoter; wild-type ICER will

also be expressed endogenously. Thus, it would be difficult to discern the effects of NKO-ICER in isolation. Therefore, it is necessary to first create a cell line with wild-type ICER knocked out. To create a knockout cell line, it is required that a knockout be specific and effective. The premise of this experiment was to design a scalable workflow to address the specificity and effectiveness of an endogenous ICER knockout. One challenge of creating such a knockout is that ICER originates from an intronic promoter within the *CREM* gene. Thus, a major component of this study was in-depth analysis of the knockout genomic loci; necessary to rule out any disruption to the exonic region of *CREM* from NHEJ. This task was feasible using next-generation sequencing accomplished through what is called 'targeted re-sequencing'. The goal of targeted resequencing is to amplify and sequence a subset of the genome potentially thousands of times. The high resolution of targeted re-sequencing makes it is possible to theoretically identify low level indels/SNP's. In this experiment, two separate bioinformatic tools were used to review the targeted re-sequencing data, but did not reveal any noteworthy mutations. Future experiments should attempt to increase the scale and transfection efficiency of cells, as well as, utilize qualitative measures of success of genetic editing (such as the T7 endonuclease assay) prior to DNA sequencing.

Materials and Methods

Cell Culture, Transfection and DNA Extraction

A 20bp locus specific sequence for the guide RNA (gRNA) (5'-CTGTCTGCAGAAGCCCATTA-3') was designed using CHOPCHOP, an online web tool for genome editing (Montague et al., 2014). The gRNA was inserted into the all-in-one plasmid pCas9-GFP (Sigma Aldrich) in the experimental group. pCas9-GFP contained a U6 promoter for gRNA expression, and CMV for expression of Cas9-GFP (See Fig. 1). The control plasmid pEGFP-N1 contained EGFP under a CMV promoter (Fig. 2) Commercially available *Homo sapien* derived malignant melanoma cells, SK-MEL-24 (ATCC® HTB-71™) were cultured exactly according to ATCC instructions and transiently transfected using FuGENE® HD Transfection Reagent according to manufacturer's instruction at an 8:2 ratio of plasmid DNA to Fugene reagent on 35-mm dishes. Cells were lysed and DNA extracted using GeneArt™ Genomic Cleavage Detection kit (Invitrogen) according to manufacturer's protocol. DNA was analyzed via nanodrop spectrophotometer to evaluate DNA concentration and quality (A260/280 and A260/230 respectively).

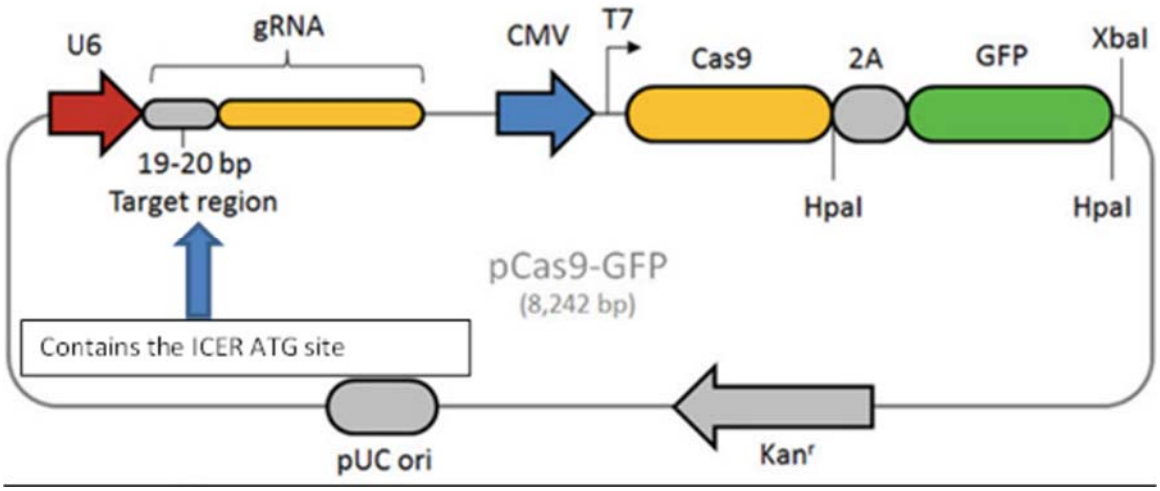


Fig. 1 Plasmid map highlighting position of gRNA and Cas9-GFP under different ubiquitous promoters. U6 is an RNA polymerase III promoter that is typically used in expression of shRNA and is useful in gRNA synthesis because it does not get polyadenylated.

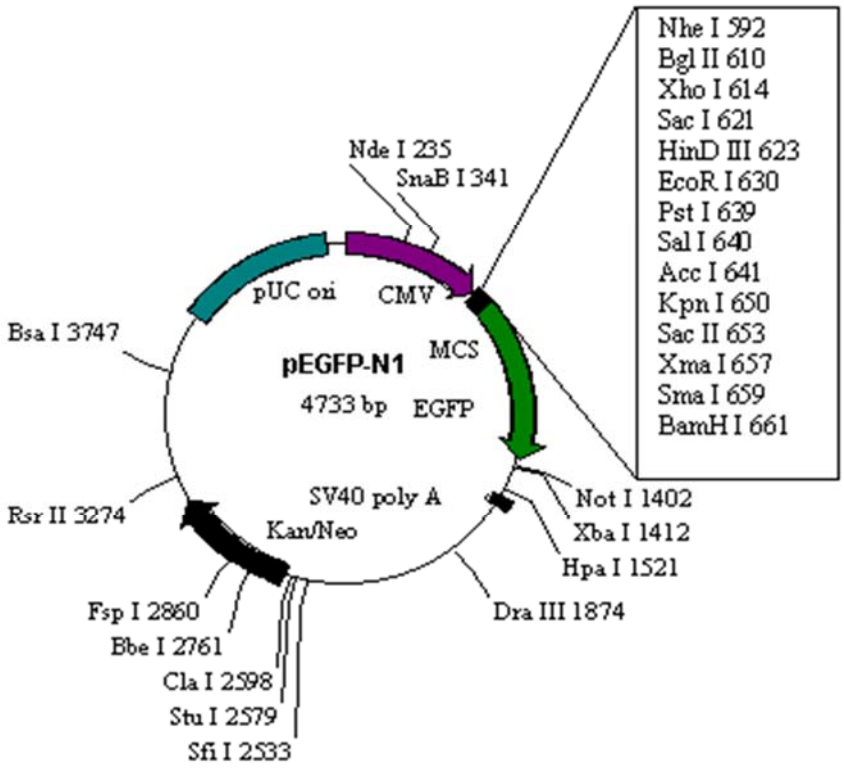


Fig. 2 Plasmid Map of pEGFP-N1. One interesting thing to note, is that the EGFP is under a CMV promoter similar to experimental plasmid.

T7 endonuclease Assay

T7 endonuclease assay was performed according to GeneArt™ Genomic Cleavage Detection kit (Invitrogen) protocol. Briefly, DNA was amplified using a Veriti 96-well Thermal Cycler (Thermo Fisher Scientific) with primers (5'-CCTGTGACAAAGCAAATTGATG-3' and 5'-AGGATTAGTGCCTCAGTCAAG-3') that created an off-center amplicon, relative to predicted cleavage site by Cas9 endonuclease. The total size of the predicted amplicon homohybrid was 408bp and the expected sizes of the two fragments were 151bp and 257bp.1uL of PCR product was added to 1uL of 10x Detection buffer and brought to a total volume of 10uL with DNase/RNase free water. DNA was denatured and allowed to reanneal. Reannealing creates heteroduplex that T7 Endonuclease recognizes and cleaves at those locations. T7 and water was added to heteroduplex mix and allowed to incubate at 37 ℃ for either 30 minutes or 45 minutes (see fig8). After incubation, mix was immediately added to lanes of 2% E-Gel® EX Gel (Invitrogen) as well as Hi-Lo 1 KB DNA Ladder (Bionexus) and allowed to run for E-Gel® iBase™ Power System (Invitrogen) for 30 minutes on low voltage setting, as per manufacturer's instructions.

Library Preparation

GTTGAACTGTGGTAGAGGAAACAAGACAGTTCTGTCTGCAGAAGCCCATTAT
GGCTGTAACTGGAGATGACACAGGTAAGAATGTTAAAGAGGGGTTTTCAGTT
AATTGTGCAGATTGTTTTGAAGTTTAGGAAGTATTCAGGAACATCTGAGTGTT
TCAGAAAGTGTTACTCTCCTAGTCACTTAGGTGTAAGACTTTTTTTGAAATAT
ACATCTATATATTCAGCTCACTTTGTTAGGGCATCTTAGTGTGATTGTTTC

Fig. 3 Predicted Amplicon for deep sequencing. ICER specific sequences for primers are highlighted in yellow. 20 bp gRNA sequence is highlighted in blue, adjacent to PAM site 'TGG'. The A immediately 5' of the PAM site is the first base of the start codon (red font) of ICER, and hence the target of locus specific DNA cleavage by CRISPR Cas9.

Gene specific primers (Forward 5'-GTTGAACTGTGGTAGAGGAAAC-3' and Reverse 5'-GAAACAATCACACTAAGATGCC-3') for first stage PCR were designed manually and confirmed to have sufficient melting temperature and little potential for complementarity by primer-Blast (NCBI). Additional sequences (Forward overhang 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3' and Reverse primer overhang 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3') were concatenated to the oligos listed above for downstream indexing, as per 16S metagenomic sequencing protocol (Illumina, USA). Primer design also ensured that the anticipated cleavage site would be incorporated into amplicon (See Fig. 3 in red) and allowed for sufficient space on either end to allow for detection of larger indels. Primers were synthesized by Eurofins, USA and resuspended with TE Buffer at 100 μM. PCR master mix for initial PCR was created according to KAPA HiFi HotStart Ready Mix PCR Kit (KAPA biosystems) reagent protocol. High-Fidelity Taq polymerase is required to reduce risk of introducing a mutation and preserving integrity of the DNA during PCR. 14μL of Master mix was kept on ice and 11μL of Genomic DNA from extraction protocol above was combined in a 0.2mL PCR safe nuclease-free tube and low-cycle amplified according to KAPA Hifi PCR Kit protocol (see table 1).

After PCR, 45uL of room temperature AMPure XP beads (Beckman Coulter) were added directly to PCR reaction. Samples were placed on an Eppendorf thermomixer at 20℃ and shaking at 1,800 rpm for two minutes, then sat for 10 minutes without shaking. Tubes were placed on DynaMag magnetic separation rack for two minutes to allow beads to move to one side of the magnet. After two minutes, supernatant was carefully removed without disturbing the bead pellet. 200μL of freshly prepared 80%

Ethanol was then added to each well, and tubes were moved back and forth on the Dynamag magnet, so beads would move from one side to the other through the Ethanol about 10 times in order to wash the beads. All Ethanol was removed, and a second Ethanol wash was performed. Beads were allowed to dry for about two minutes, then 30µL of TE buffer was added to each well to elute DNA. Tubes were placed back on Eppendorf thermomixer at 20℃ and shaking at 1,800 rpm for two minutes and sat without shaking for three minutes. After three minutes, tubes were placed back on Dynamag magnet for two minutes. Supernatant collected was cleaned up DNA product.

In preparation for paired-end DNA sequencing, in which all samples are pooled together into a single tube, unique adaptors must first be attached to the end of each amplicon for downstream sample identification. Illumina Nextera XT indexes, which can produce up to 24 unique combinations of i5 and i7 indexes (see Fig.4) were used in this experiment. Creation of a matrix (as shown in Fig. 4) ensures minimal overlap between index tags and reduces risk of index read failure. In a clean 0.2mL PCR safe nuclease free tube, 5µL of cleaned up Amplified DNA with overhang sequences was combined with 5µL of i5 and i7 indexes, 10µL of Nuclease free-water and 25µL of KAPA Hi-Fidelity Hot start Taq Polymerase Ready Mix. Short 8 cycle PCR was performed to add on indexes to each amplicon (see table 1).

Round 1 PCR			
Step	**Temperature**	**Duration**	**Cycles**
Initial Denaturation	95℃	3 min	1
Denaturation	98℃	20 sec	18
Annealing	65℃	15 sec	
Extension	72℃	30 sec	
Final Extension	72℃	1 min	1
Round 2 PCR			
Step	**Temperature**	**Duration**	**Cycles**
Initial Denaturation	95℃	15 mins	1
Denaturation	95℃	15 sec	8
Annealing	65℃	30 sec	
Extension	72℃	30 sec	
Final Extension	72℃	5 mins	1
qPCR			
Step	**Temperature**	**Duration**	**Cycles**
Initial Denaturation	95℃	5 mins	1
Denaturation	95℃	30 sec	35
Annealing	60℃	45 sec	

Table 1. Settings for first and second round PCR, as well as settings for library quantification via qPCR

	N701	N702	N703	N704	N705	N706
S517	Sample 1	0	0	0	0	0
S502	0	Sample 2	0	0	0	0
S503	0	0	Sample 3	0	Sample 5	0
S504	0	0	0	Sample 4	0	Sample 6

Fig. 4 Matrix for unique selection of Nextera indexes to be ligated to amplicon. In paired-end sequencing, little overlap between identical indexes is ideal to reduce the risk of downstream index read failure. This matrix allows for not only unique combinations of indexes but ensures diversity in first base of index called.

Sample #	Sample_Name	Sample_Well	I7_Index_ID	index	I5_Index_ID	index2
1	Experimental_1	A01	N701	TAAGGCGA	N517	GCGTAAGA
2	Control_1	B01	N702	CGTACTAG	N502	CTCTCTAT
3	Experimental_2	C01	N703	AGGCAGAA	N503	TATCCTCT
4	Control_2	D01	N704	TCCTGAGC	N504	AGAGTAGA
5	Experimental_3	E01	N705	GGACTCCT	N503	TATCCTCT
6	Control_3	F01	N706	TAGGCATG	N504	AGAGTAGA

Table. 2 Sample ID table showing same name with corresponding index tags.

After indexing PCR, 45µL AMPure XP beads were added to 30µL of PCR product in a clean 8 tube strip and additional bead cleanup was performed as previously described. Elution tube containing final libraries were labelled in a 1.5mL Eppendorf tube and stored in -20℃ prior to sequencing.

Library Quantification

Final libraries were quantified via real-time PCR (qPCR) according to KAPA Library Quantification Kit Illumina® Platforms instructions. Briefly, 'ROX high' passive reference dye was added to KAPA SYBR FAST qPCR Master Mix. 12.4µL of Master mix was added to MicroAmp™ Fast Optical 96-Well Reaction Plate (Applied Biosystems) with 3.6uL of Nuclease-free water and 4µL of each sample (ran in triplicate) diluted 1:10,000 along with a no template control (NTC) and PhiX as a positive control whose concentration was known to be 10nM. Real-time PCR was performed on StepOnePlus™ Real-Time PCR System programmed for an initial denature for 5 minutes at 95℃ and then 35 cycles of 95℃ for 30 seconds and 45 seconds at 60℃ for annealing/extension and data acquisition (see table. 1). No melt curve was performed during this QC. Mean Ct value for each of the six controls was plotted on a line with the log (known concentration) with an R^2 value ~0.999. From this line, the Y-intercept and slope were derived and mean Ct for each sample was put into equation to determine each sample concentration in nM. Final concentration was determined using the average size of amplicon derived from Tapestation.

Sample size was quantified via Agilent TapeStation 2200. 2µL of final libraries was combined with 2µL of High Sensitivity buffer (Agilent) and ran using High Sensitivity DNA Tape (Agilent). Regions were set from 200 to 700 bp.

Pooling and Sequencing

All samples were normalized to 2nM based on concentrations derived from library quantification via qPCR. 5µL from each diluted library was pooled together into a single 1.5mL Eppendorf tube. 20µL of the pooled amplicon library was added to a 0.2mL PCR tube. In another tube, 20µL of 2nM PhiX was added. PhiX is commonly used in next-generation sequencing experiments to diversify libraries. Both tubes were denatured at 95ºC for five minutes, then tubes were placed on ice for five minutes. 20µL of freshly prepared 0.1M NaOH was added to each tube (reducing concentration of each pool to 1nM. Final loading concentration was empirically determined based on previous sequencing runs and corresponding cluster densities and concluded to be 5pM. From the diluted library tube, 20% of the final volume was removed and discarded and replaced with diluted denatured PhiX control as a 20% spike in. 600uL of final denatured pool was added to Miseq Reagent Kit v3 (600-cycle) and performed 200x8x8x200 sequencing under amplicon chemistry.

Bioinformatic analysis

Initial analysis of CRISPR Cas9 mediated knockout of ICER was performed using CRISPResso, an online computation pipeline for detection of SNP's and indels on target regions by direct comparison of raw paired-end FASTQ files and predicted amplicon (Pinello et al., 2016). Read 1 and Read 2 FASTQ files were retrieved from Basespace sequence hub (Illumina, USA), and uploaded onto CRISPResso online interface along with predicted amplicon (see Fig. 3) and gRNA sequence (5'-CTGTCTGCAGAAGCCCATTA-3') with no window surrounding predicted cleavage

site, minimum average read and base quality of phred ≥30, and no exclusions from either side of predicted cut site. No trimming of adaptor sequences was required because this function was performed via Basespace Sequence Hub.

Additional bioinformatic analysis using bowtie and SAMtools was performed to confirm CRISPResso results and to investigate the output data more thoroughly. Ultimately these bioinformatic tools enable creation of human genome indexed .BAM files; a file type compatible with Integrative Genome Browser or IGV (Broad Institute) for visualization of sequence data and variant calling/confirmation. Briefly, Sequence Alignment to reference human genome GRCh38 was created with bowtie (Johns Hopkins University) running from Bioconda using terminal on MacOS 10.12.6. Output file from reference genome alignment was .SAM file, and subsequently converted to .BAM file which is much more efficient for storage because it converts sequence to a binary file. Reads from each file were filtered for quality based on phred score (Q30). Only bases with >=Q30 were kept, minimizing background. BAM file outputs from Read 1 and Read 2 were then merged together using SAMtools to overlap based on previously specified sequencing parameters (200x8x8x200). Merged files were then sorted according to FASTA for GRCh38 reference human genome.

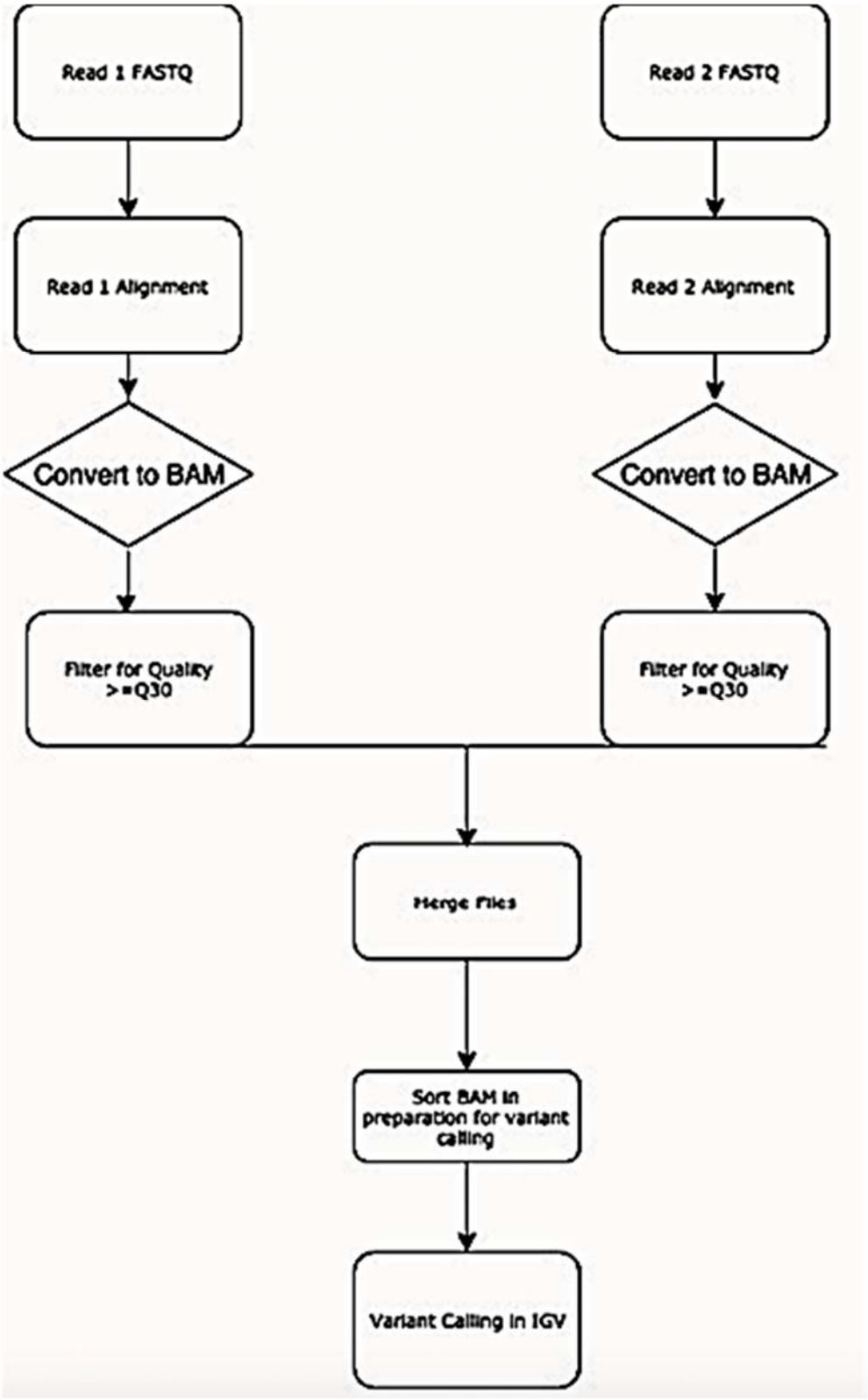


Fig. 5 Flow chart describing general workflow for manipulating sequencing output file in preparation for variant calling. Final output from this bioinformatic pipeline is an indexed .BAM file that can be directly interpreted by IGV to display corresponding basecalls.

Results

Transfection efficiency

Both control and experimental samples expressed GFP see figures 6 through 7. Expression of GFP suggests that gRNA and Cas9 were also expressed because they originate from the same plasmid. Transfection efficiency for both groups was between 20 and 40%.

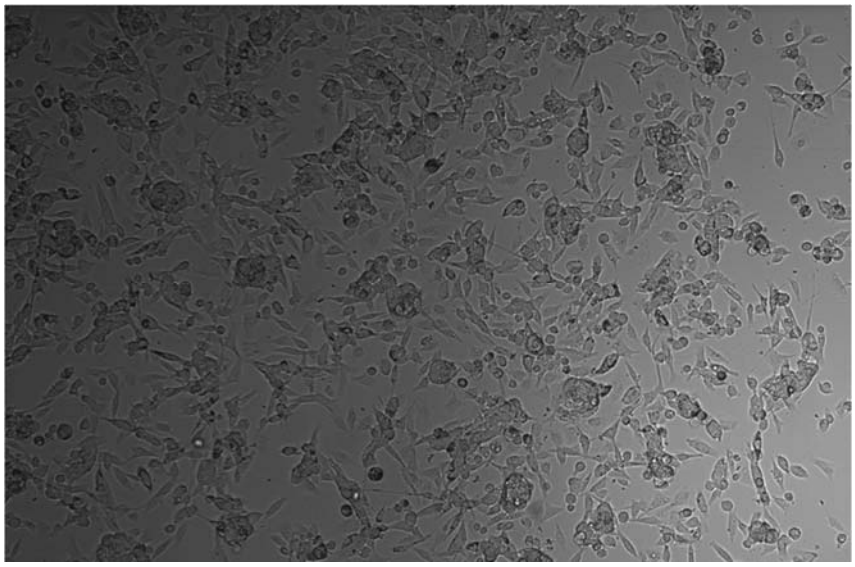


Fig. 6a Brightfield Control SK-MEL-24 cells at 10x magnification.

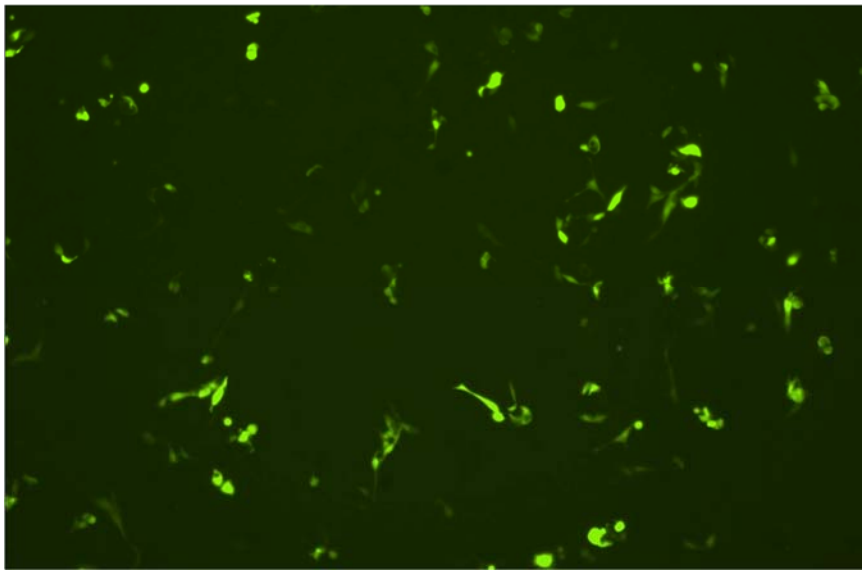


Fig. 6b FITC Control SK-MEL-24 under same magnification as brightfield. Transfection efficiency of cells was determined to be roughly 30%.

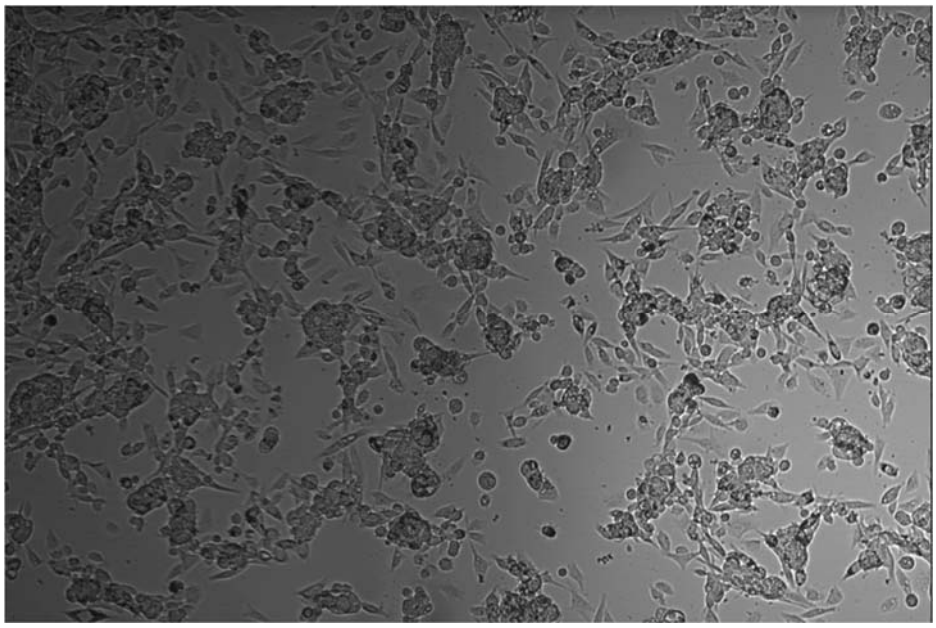


Fig. 7a Brightfield Experimental SK-MEL-24 cells at 10x magnification.

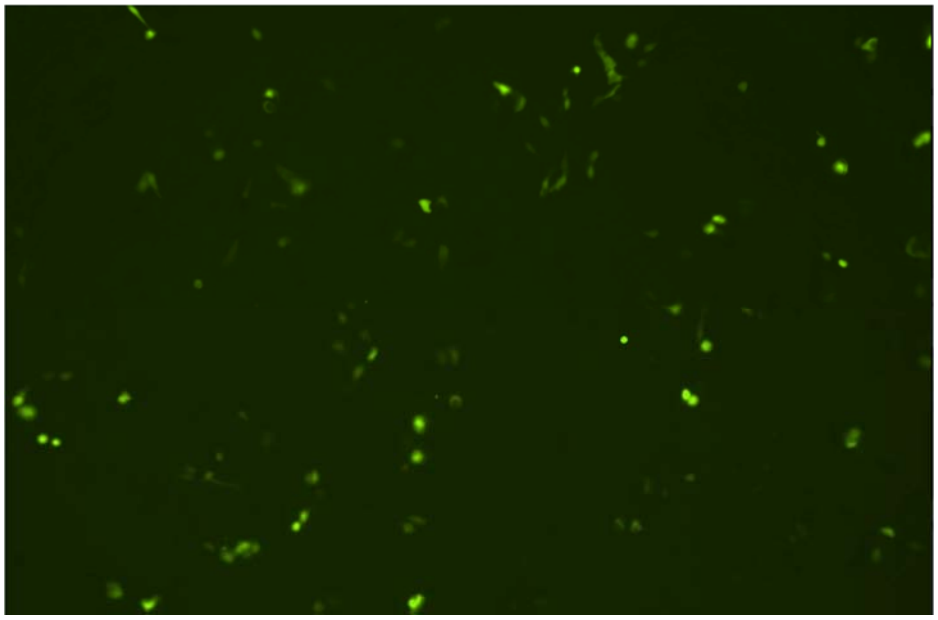
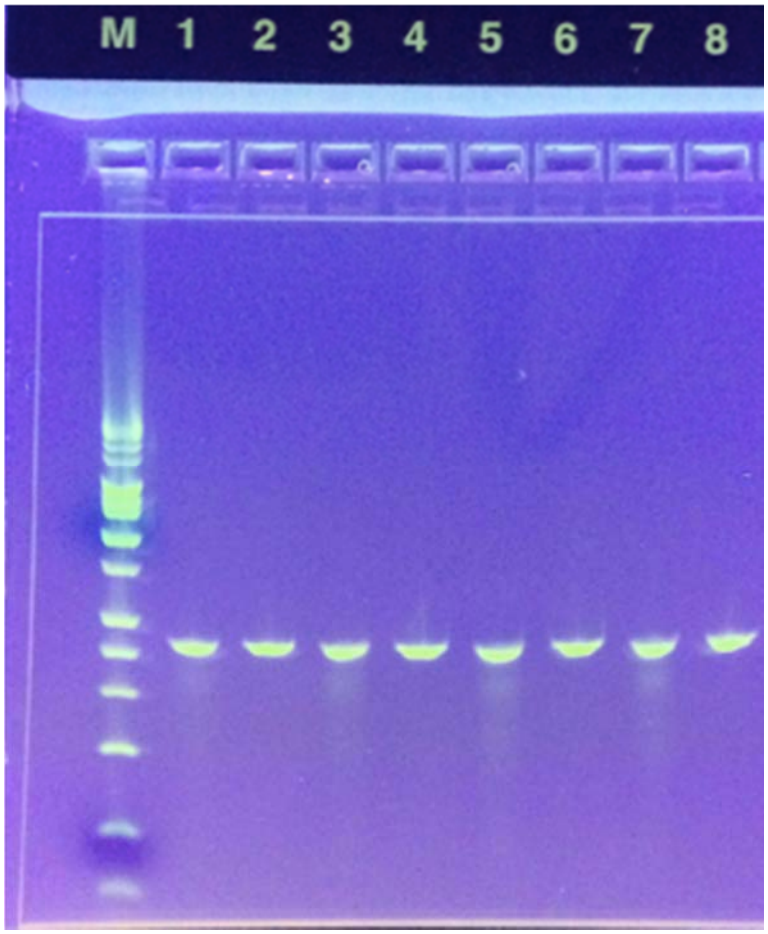


Fig. 7b FITC Experimental SK-MEL-24 cells under same magnification as brightfield. Transfection efficiency of cells was determined to be roughly 30%.

T7 Endonuclease Assay

No distinct bands smaller than initial PCR product were identified from T7 Endonuclease assay (see Fig. 8), which suggests no genomic cleavage by CRISPR Cas9. Although the sensitivity of this assay is limited, this can be used as a good predictor of experimental success. A time course of DNA incubation with T7 endonuclease was performed to rule out non-specific cleavage by T7. Even after only 30 minutes of incubation, the control sample (Fig. 8 lane 3) (that should not create heteroduplexes) was recognized and appears to be cleaved by T7, as evident by a smear on the lane when compared to control without T7 in lane 4 (Fig. 8). Thus, with this high level of background, T7 endonuclease may require much higher transfection efficiency for efficient identification of indels/substitutions.



Lane	Group	Time
MWM	MWM	MWM
1	Experimental+	30 Mins
2	Experimental-	30 Mins
3	Control+	30 Mins
4	Control-	30 Mins
5	Experimental+	45 Mins
6	Experimental-	45 Mins
7	Control+	45 Mins
8	Control-	45 Mins

Fig. 8 2% E-gel from T7 endonuclease assay. Description of lanes are found in Table below. To determine if there was any non-specific cleavage after extended incubation with T7 endonuclease, each group was exposed +/- enzyme and also with varying degree of time (30 or 45 minute incubation).

Library Quantification

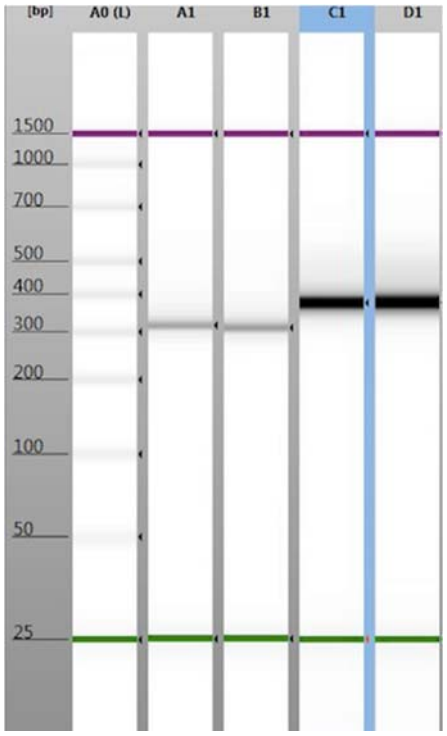


Fig. 9 Highsensitivity TapeStation QC. From left: High Sensitivity Ladder, Experimental after first round PCR, Control after first round PCR, Experimental after round 2 PCR, Control after Round 2 PCR. Increase in base pair size suggests successful addition of index adaptors to amplicon after second round PCR. Final amplicon size was ~377bp.

Bioinformatic analysis

Fig. 10 Graph from basespace showing %reads per sample with corresponding unique index assignments. This graph was primarily used to determine which set of control and experimental samples to be used for direct bioinformatic comparison. Index 5 and Index 6 corresponding to Experimental 3 and Control 3, respectively contain comparable number of reads.

LANE	READ	CLUSTER PF (%)	%≥Q30	YIELD	ERROR RATE%	READS PF	DENSITY	TILES	LEGACY PHAS/PREPHAS (%)	INTENSITY	COMMENTS	STATUS
1	1	95.34 ±0.84	91.82	1.86 Gbp	2.43 ±0.22	9,346,079	380 ±9	38	0.177 / 0.254	73 ±8		QC Passed
	2(I)		64.49	65.42 Mbp	0.00 ±0.00				0.000 / 0.000	146 ±11		
	3(I)		88.11	65.42 Mbp	0.00 ±0.00				0.000 / 0.000	638 ±58		
	4		91.16	1.86 Gbp	1.90 ±0.15				0.283 / 0.257	86 ±10		

Fig. 11 Chart showing Quality metrics from Basespace. While clusters passing filter rate was quite high 95.3, the cluster density of run was on the low end ~ $380K/mm^2$

.

	CYCLES	YIELD	PROJECTED YIELD	ALIGNED (%)	ERROR RATE (%)	INTENSITY CYCLE 1	%≥Q30
Read 1	200	1.86 Gbp	1.86 Gbp	24.01	2.43	73	91.82
Read 2 (I)	8	65.42 Mbp	65.42 Mbp	0.00	0.00	146	64.49
Read 3 (I)	8	65.42 Mbp	65.42 Mbp	0.00	0.00	638	88.11
Read 4	200	1.86 Gbp	1.86 Gbp	25.97	1.90	86	91.16
Non-Index Reads Total	400	3.72 Gbp	3.72 Gbp	24.99	2.16	80	91.49
Totals	**416**	**3.85 Gbp**	**3.85 Gbp**	**24.99**	**2.16**	**236**	**90.97**

Fig. 12 Chart showing Yield metrics from Basespace. Of note, the total yield 3.85giga bases and 24.99 aligned. This number of alignments should be similar to percent of PhiX spike in, which was 20%. Of the 3.85giga base yield, greater than 90% were >=Q30.

CRISPResso analysis did not reveal any significant changes in experimental sequence that would suggest successful editing via CRISPR Cas9. One single C->T substitution appears in 0.2% of all sequence reads according to CRISPResso output (see Fig. 16), but this does not represent a notable change in sequence.

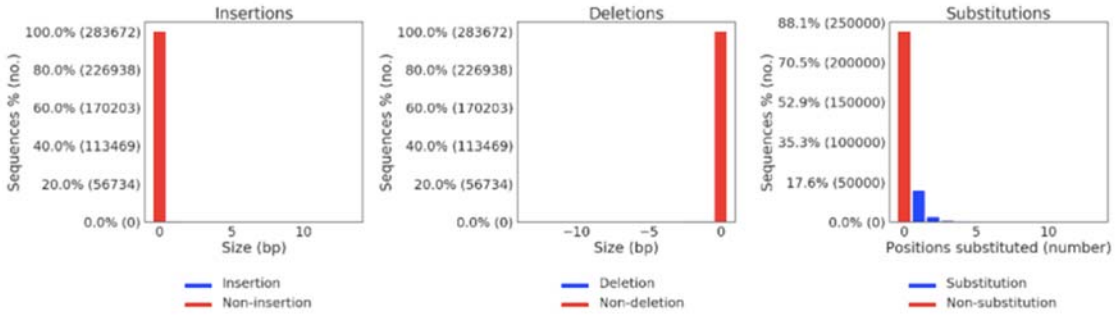


Fig. 13 Control 3 showing the frequency and distribution of insertions, deletions, and substitutions.

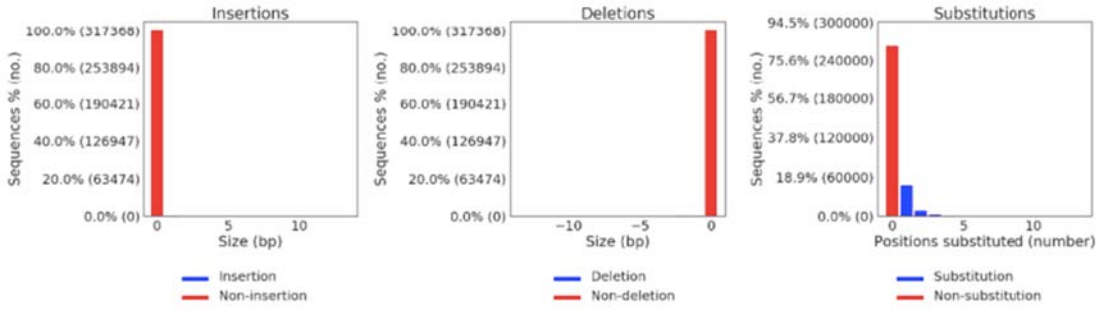


Fig.14 Experimental 3 showing the frequency and distribution of insertions, deletions, and substitutions.

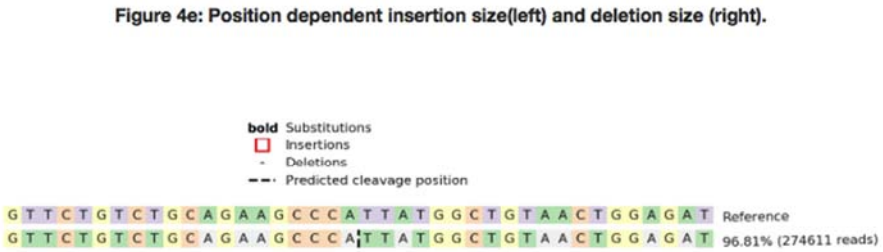


Fig. 15 Results from CRISPResso showing number of reads that align directly to the reference from Control sample and any substitutions/indels within a window from predicted cleavage site.

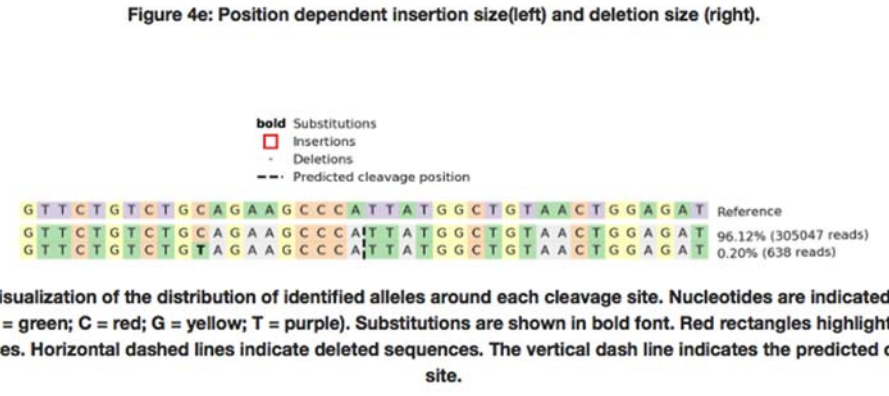


Fig. 16 Results from CRISPResso showing number of reads that align directly to the reference from Experimental sample and any substitutions/indels within a window from predicted cleavage site.

A custom bioinformatic pipeline using Bowtie and SAMtools was created in order to identify whether any sequences were missed by CRISPResso's algorithm in determining how efficient CRISPR Cas9 was at mediating genetic alteration. These analyses were performed as described above s to render the paired-end FASTQ files in a suitable file format for visualization using IGV. Indexed samples 5 and 6 were directly compared using IGV as shown in Fig. 17 with the Experimental sample on top and

Control sample on the bottom. CRISPResso identified the same low-level variant in experimental sample but failed to display the level of detail achieved by IGV. In addition, the exact location of this substitution varies between the two results. CRISPResso displays this substitution as C->T, but when mapped to human genome reference GRCh38, IGV displays this substitution as A->T (See Fig. 17). Other lower level variants were identified with IGV that were not shown by CRISPResso, but none were considered statistically significant or near the start codon. This suggests that although more of these low level variants were identified in experimental samples compared to the control, they are most likely PCR artifacts and not the result of CRISPR Cas9 mediated genetic alteration.

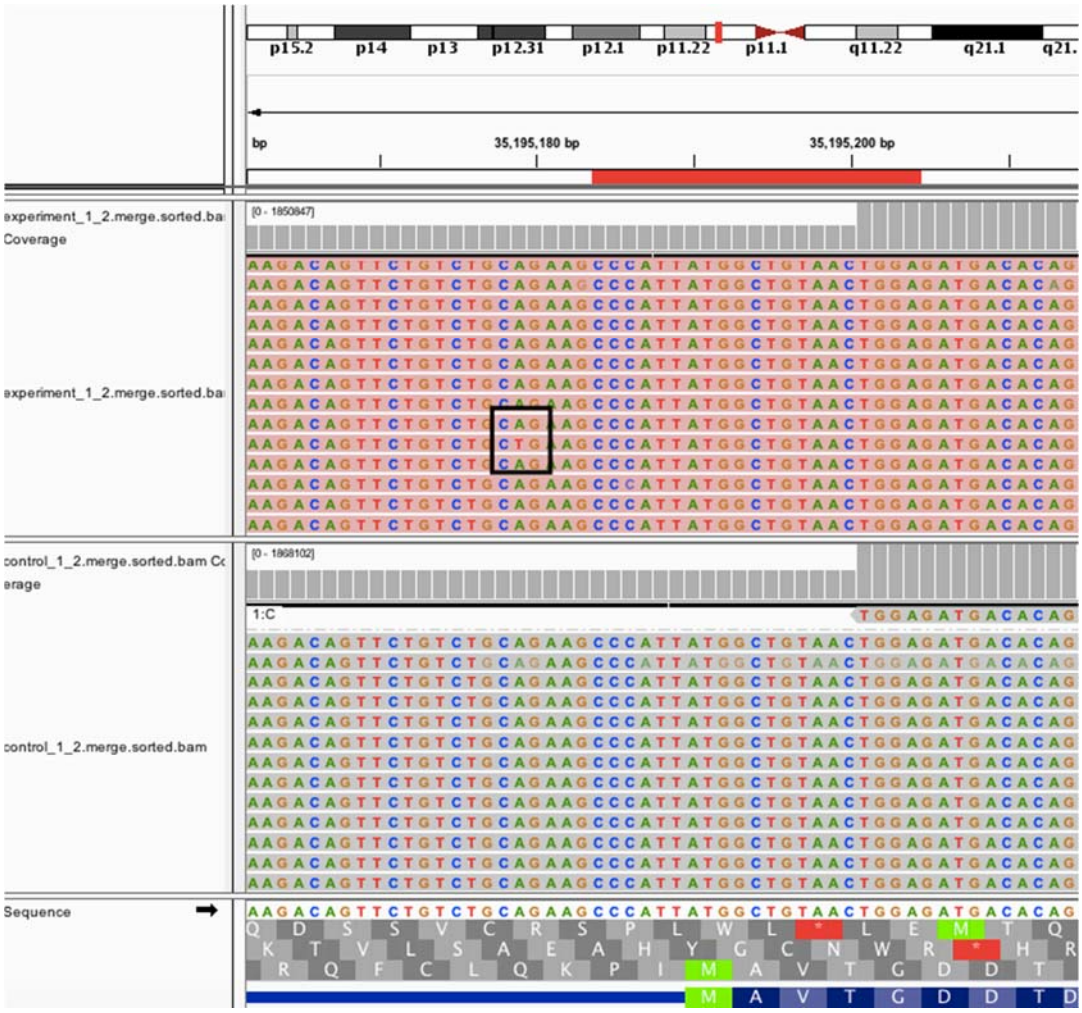


Fig. 17 Screenshot of Integrative Genome Browser (IGV) corroborating CRISPResso result of low-level variant although when mapped to human genome, it appears this variant is mapped one bp off compared to CRISPResso output.

Discussion

Inducible cAMP Early Repressor (ICER) is a putative tumor suppressor that is actively targeted for proteasomal degradation, most likely through ubiquitination. Melanoma cells rapidly apoptosed when transfected with a mutant form of ICER that did not contain any lysine residues. This suggests that in diseased cells, such as malignant melanoma (and other cancers), that the cell recognizes ICER as an imminent threat to cell proliferation. Unfortunately, due to apparent rapid apoptosis, protecting ICER from degradation or subcellular relocation alone is insufficient as a model system. This issue compounds the need for an inducible expression system, in which expression of ICER can be regulated so that further investigation can identify how ICER plays a role in cancer growth. To accomplish this task, wild type ICER should be effectively knocked out to be able to characterize the role of NKO-ICER exclusively. In this thesis, an ICER knockout was attempted using transient transfection with an all in one plasmid containing a gRNA designed to direct Cas9 to the start codon of ICER. While transfection efficiency of the cells appeared promising, targeted resequencing of this locus many times did not suggest any endonucleolytic activity by Cas9. While no notable differences were identified between the experimental and control group in this particular experiment, a successful workflow for carrying out and analyzing targeted re-sequencing data of CRISPR Cas9 mediated genetic alteration was developed. In doing so, it was identified that sequencing results can differ depending on human genome reference. For example, CRISPResso a bioinformatics tool that compares sequence data directly to a reference sequence could be different from a more in-depth analysis which maps sequence data to the coordinates of a reference genome. In this experiment, this subtle difference of one

nucleotide that was present a minority of the time did not change the outcome, but it is not impossible that this difference could be significant in other instances.

Future experiments would take into consideration number of clone's transfected and screened. It may be required to isolate many different clones and screen hundreds to identify a genotype sufficient for our purposes of creating a knockout cell line. One advantage is that this workflow is easily scalable and so more in-depth screening is certainly possible. It is plausible that this particular gRNA sequence is not compatible with this cell type. In this case, it may be required to explore different Cas9 species that originate from within different bacterial strains. The Cas9 in this experiment was of the most widely used variety, *Streptococcus pyogenes,* but different Cas9 species would recognize different PAM sites, providing more flexibility in terms of gRNA design and thus Cas9 endonuclease activity. In addition, it has not escaped my knowledge that more sophisticated bioinformatic tools exist and could be used in this application. Further work will be done to incorporate such tools into this workflow to rapidly and accurately identify indels and substitutions from CRISPR Cas9.

Bibliography

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., ... & Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 315(5819), 1709-1712.

Deveau, H., Barrangou, R., Garneau, J. E., Labonté, J., Fremaux, C., Boyaval, P., ... & Moineau, S. (2008). Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *Journal of bacteriology*, *190*(4), 1390-1400.

Healey, M., Crow, M. S., & Molina, C. A. (2013). Ras-induced melanoma transformation is associated with the proteasomal degradation of the transcriptional repressor ICER. *Molecular carcinogenesis*, *52*(9), 692-704.

Horvath, P., & Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. *Science*, 327(5962), 167-170.

Hsu, P. D., Lander, E. S., & Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, 157(6), 1262-1278.

Jafari, N., Kim, H., Park, R., Li, L., Jang, M., Morris, A. J., ... & Huang, C. (2017). CRISPR-Cas9 mediated NOX4 knockout inhibits cell proliferation and invasion in HeLa cells. *PloS one*, *12*(1), e0170327.

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*, *10*(3), R25.

Mattiroli, F., & Sixma, T. K. (2014). Lysine-targeting specificity in ubiquitin and ubiquitin-like modification pathways. *Nature structural & molecular biology*, *21*(4), 308.

Mémin, E., Yehia, G., Razavi, R., & Molina, C. A. (2002). ICER reverses tumorigenesis of rat prostate tumor cells without affecting cell growth. *The Prostate*, *53*(3), 225-231.

Mémin, E., Genzale, M., Crow, M., & Molina, C. A. (2011). Evidence that phosphorylation by the mitotic kinase Cdk1 promotes ICER monoubiquitination and nuclear delocalization. *Experimental cell research*, *317*(17), 2490-2502.

Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M., & Valen, E. (2014). CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic acids research*, *42*(W1), W401-W407.

Pastwa, E., & Błasiak, J. (2003). Non-homologous DNA end joining. Acta Biochimica Polonica, 50(4), 891-908.

Pinello, L., Canver, M. C., Hoban, M. D., Orkin, S. H., Kohn, D. B., Bauer, D. E., & Yuan, G. C. (2016). Analyzing CRISPR genome-editing experiments with CRISPResso. *Nature biotechnology*, *34*(7), 695.

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature protocols*, *8*(11), 2281.

Razavi, R., Ramos, J. C., Yehia, G., Schlotter, F., & Molina, C. A. (1998). ICER-IIγ is a tumor suppressor that mediates the antiproliferative activity of cAMP. Oncogene, 17(23), 3015.

Tiyaboonchai, A., Mac, H., Shamsedeen, R., Mills, J. A., Kishore, S., French, D. L., & Gadue, P. (2014). Utilization of the AAVS1 safe harbor locus for hematopoietic specific transgene expression and gene knockdown in human ES cells. *Stem cell research*, *12*(3), 630-637.

Yehia, G., Schlotter, F., Razavi, R., Alessandrini, A., & Molina, C. A. (2001). MAP kinase phosphorylates and targets inducible cAMP early repressor to ubiquitin-mediated destruction. *Journal of Biological Chemistry*.

Zhang, W., & Liu, H. T. (2002). MAPK signal pathways in the regulation of cell proliferation in mammalian cells. *Cell research*, 12(1), 9.