ABLATION OF THE NEIL1 GENE ALTERS THE TRANSCRIPTIOME OF HTERT RPE-1 CELLS

A thesis in fulfillment of the Masters in Biology Degree

Bisi Otulana
Mentor: Professor David Scicchitano

Summary

Protection from DNA damage and mutagenesis within human cells is achieved by a collection of genome maintenance pathways. Important among them is base excision repair, which is initiated by class of enzymes called glycosylases that remove damaged bases from the genome. The NEIL1, NEIL2 and NEIL3 genes encode three specific glycosylases that typically remove oxidized bases from DNA. While these genes have been widely studied, there is still work to be done in understanding the role they play in cellular processes. This study employed CRISPR technology to generate a NEIL1 knockout (N1KO) in hTERT RPE-1 cells. The main goal of the study was to understand the impact of this specific gene on the cell's transcriptome. The results showed that knocking out the NEIL1 gene caused changes in at least 66 genes. Further analysis using Gene Ontology (GO) Terms suggested that the N1KO cells exhibit increased metabolism when compared to wildtype (WT) cells. This was reinforced by experiments using Agilent Seahorse technology to measure oxygen consumption rates that showed higher oxygen consumption in the N1KO cells.

Introduction

DNA damage can have serious consequences for an organism's health and survival. Lesions in DNA can disrupt normal biological processes that can injure the organism or even be fatal. There are two broad sources of damage—those derived endogenously and those derived exogenously. Exogenous sources include (1) chemical agents that damage the constituents of DNA, particularly the nitrogenous bases; (2) ionizing radiation (IR), which induces oxidative damage and DNA strand breaks; and (3) ultraviolet (UV) radiation, which also produces damaged bases (Hakem et al., 2008). Equally interesting and complex are the endogenous agents generated during normal metabolic processes that can induce hydrolysis, oxidation, and alkylation of DNA, producing breaks in the backbone of double helix.

Reactive oxygen species (ROS) are particularly germane to the work reported here. The main production of ROS occurs within the mitochondria when the electron transport chain functions for large-scale ATP production. At certain sites within the chain, electrons are passed to FeS carriers and flavin intermediates. These single electrons can react with oxygen instead of moving forward in the chain, which results in superoxide formation. Superoxide dismutase catalyzes the conversion of superoxide radical anions to hydrogen peroxide (Lambert et al., 2009). Catalase, in turn, catalyzes the breakdown of hydrogen peroxide into hydrogen and water. Nonetheless, superoxide anions that escape the action of SOD and other ROS can oxidize cellular constituents, including the nitrogenous bases in DNA. In some cases, these altered bases miscode during replication, resulting in mutations; in others, they block replication, which can lead to cell death (Rosenquist et al., 2003).

Mutations within genes can alter the RNA that is produced during transcription. When the mutations arise within a gene's transcription unit, the fundamental sequence of the resulting RNA

can be different from the WT. For mRNA, such alterations can result in changes to protein structure following translation. If mutations arise within a gene's regulatory elements, such as enhancers or promoters, functional regulation of the gene can be compromised (Cooper et al., 2009).

Clearly, the loss of genomic integrity due to DNA damage predisposes the organism to a variety of disorders. But cells have evolved a system of coordinated DNA-damage checkpoints and repair pathways that facilitate survival in the face of DNA damage (Chatterjee et al., 2017). Nucleotide excision repair (NER), mismatch repair (MMR), and base excision repair (BER) are among the pathway's cells use to repair damaged DNA. BER is relevant to the work described here and warrants special emphasis. BER is initiated by DNA glycosylases. These glycosylases, which are either monofunctional or bifunctional, are enzymes that hydrolyze the N-glycosidic bond of a purine or pyrimidine, releasing the damaged base. Monofunctional glycosylases remove their target bases, but the subsequent action of an abasic endonuclease is required to incise the DNA. Bifunctional DNA glycosylases possess inherent AP lyase activity that cleaves the sugar–phosphate bond following removal of the damaged base, introducing a nick in the DNA backbone 3′ to the lesion. The detailed repair by monofunctional and bifunctional glycosylases are shown in figure 1 along with the common DNA lesions shown in figure 2. The repair cycle is completed by the concerted action of DNA polymerases and DNA ligases (Rosenquist et al., 2003).

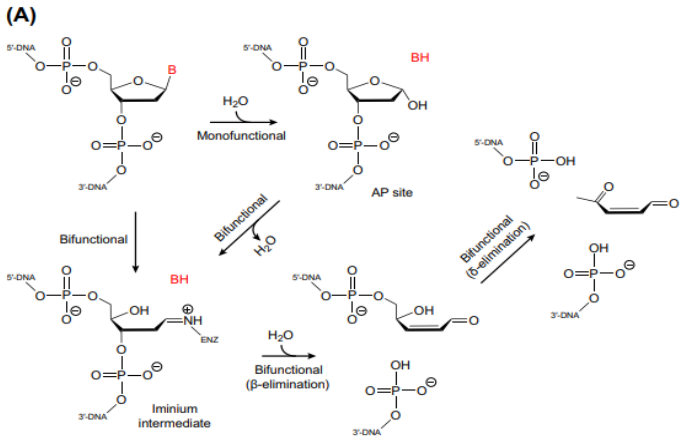


Figure 1: Base excision and strand incision reactions performed by monofunctional and bifunctional DNA glycosylases. (A) Monofunctional enzymes catalyze only base excision, wherein the glycosidic bond of the damage base is hydrolyzed, removing the nucleobase and creating an AP site. Bifunctional enzymes catalyze both base excision and strand incision (lyase activity). During the base excision step, Bifunctional glycosylases either directly form an iminium intermediate or initially hydrolyze the glycosidic bond to create an AP site before then converting the AP site to an iminium intermediate. Following base excision, all bifunctional DNA glycosylases incise the strand on the 3'-side of the AP site (β-elimination), generating a single-strand break with a 3'-phospho-α,β-unsaturated aldehyde (PUA) group and a 5'-phosphoryl group. Some bifunctional enzymes also subsequently incise the strand on the 5'-side of the PUA (δ-elimination), leaving a 3'-phosphate, which must also be removed prior to repair synthesis, requiring the phosphatase activity of a separate enzyme. Alternatively, if β-elimination occurs following strand incision by an AP endonuclease, a gap is generated with the 3'-hydroxyl and 5'-phosphoryl groups necessary for synthesis and ligation. (Mullin et al., 2019).

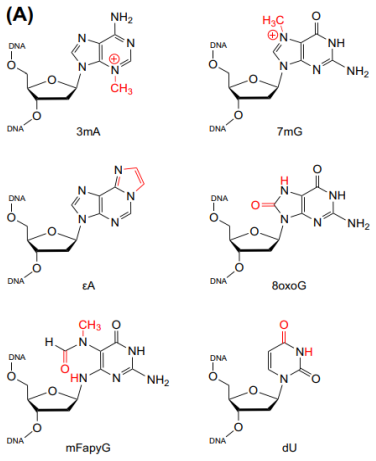


Figure 2: Common DNA lesions resulting from alkylation, oxidation, and deamination of nucleobases. 3mA,3-methyl-2'-deoxyadenosine;7mG, 7-methyl-2'-deoxyguanosine; 8oxoG, 8-oxo-2'-deoxyguanosine; εA, 1,N6-etheno-2'-deoxyadenosine; mFapyG, methyl derivative of N6-(2′-deoxyribosyl)-2,6-diamino-4-oxo-5-formamidopyrimidine (FapyG); dU, 2'-deoxyuridine (Mullins et al., 2019).

In humans, there are three important glycosylases whose function is to clear oxidized bases from DNA: They are encoded by the nei-like DNA glycosylase (NEIL) genes NEIL1, NEIL2, and NEIL3. While the NEIL1, NEIL2, and NEIL3 glycosylases share a common function, their specific catalytic processes and cell cycle dependency make each unique. A major difference is that NEIL1 acts preferentially on oxidized bases in double-strand DNA exhibiting a β,δ-lyase activity, while NEIL3 has an affinity for single-strand DNA, incising the DNA through a weak β-lyase activity (Albelazi et al., 2019). Additionally, NEIL1's repair activity acts on damage in DNA that is upstream to replication forks, whereas NEIL2's and NEIL3's activity function within open fork structures. Cell-cycle specificity among each of the genes is also present. NEIL1's and NEIL3's expression are induced in S/G2 phases of the cell cycle, linking them to replication associated repair. In contrast, NEIL2 is associated with transcription-coupled repair since it interacts with several proteins that are necessary for transcription, including RNA polymerase II (Mullins et al., 2019).

The differences among the functions of NEIL1, NEIL2, and NEIL3 are tied to the structural composition of the proteins. All three share an N-terminal DNA glycosylase domain and a helix-two-turn helix (H2TH) DNA binding domain displayed in Figure 3. NEIL 3 exclusively possesses an extended C-terminal domain that contains three additional zinc finger binding motifs. This extended region purportedly aids in the specific removal of DNA damage that occurs within the telomeres during S/G2 phase of actively dividing cells. The protection and repair of telomeres is of great importance as these G-rich regions are exceptionally vulnerable to reactive oxygen species. Interestingly, NEIL3 expression is repressed in non-dividing cells and restricted to cells with a high proliferative capacity, such as adult/embryonic stem cells and hematopoietic cells (Zhou et al., 2017).

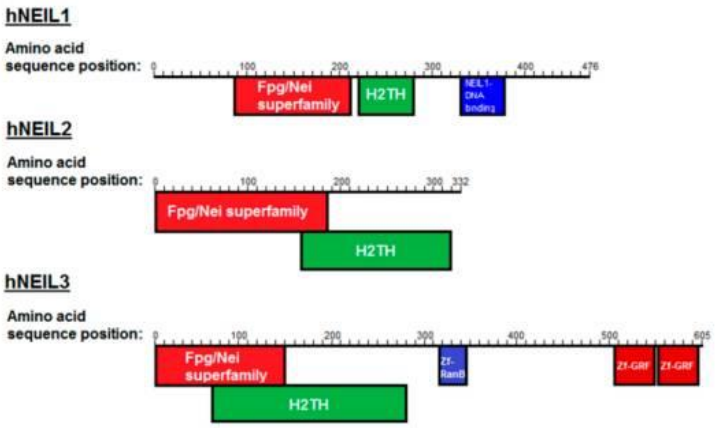


Figure 3: Conserved domains of NEIL1, NEIL2 and NEIL3 glycosylases (Albelazi et al., 2019).Although the NEIL family of enzymes has been well studied, specific details of their action are still under investigation. Knocking out NEIL genes impedes the BER of oxidized bases in DNA, which results in their accumulation . This suggests that cells that lack NEIL activity have a harder time repairing damaged DNA, which can lead to abnormal cell function and compromised survival. Studies have also discovered the relative importance of NEIL genes and their use in protecting mitochondrial DNA against oxidative damage (Han et al., 2019).

Considering the crucial role that NEIL genes play in genomic repair, it is vital to gain a deeper understanding of how they impact additional cellular pathways and networks. Knocking out these genes will provide the opportunity to observe the effects and infer the genes role in a particular cellular process. Additionally, creating knockouts can also provide insight into the causes of certain diseases and the potential for new therapeutic targets. For instance, mice harboring homozygous and heterozygous deletions of NEIL1 have an increased onset of severe obesity, dyslipidemia, and fatty liver disease leading to metabolic syndrome disorders. Additionally, mitochondrial DNA from homozygous N1KO mice show increased levels of steady-state DNA damage and deletions relative to WT controls (Vartanian et al., 2006). These data suggest an important role for NEIL1 in the prevention of the diseases associated with the metabolic syndrome. From these studies there is a central question that remains to be answered: What specific transcriptional networks are being altered from defects in these NEIL genes that lead to the

manifestation of such disorders and syndromes? As more studies are conducted on the role these glycosylases play in genome maintenance, it is important to understand further the effects of ablation these genes on the transcriptome. Consequently, this project set out to elucidate the NEIL proteins function on transcriptional networks by focusing on the functional consequences of impaired NEIL activity using the CRISPR CAS-9 system to ablate the gene. Given the wide scope of this project, special emphasis was exclusively placed on NEIL1. Immortalized human retinal pigment epithelial (hTERT RPE-1) cells were chosen as the model due to their ability to proliferate indefinitely and maintain intact telomeres via telomerase activity. Note too that these cells are not virally transformed. Furthermore, RNAseq was implemented to obtain a comprehensive analysis of the transcriptional effects caused by the knockouts.

Methods

Tissue culture

All hTERT RPE-1 cells and cell lines were grown in Dulbecco's modified eagle medium F-12 (DMEM F-12) supplemented with 10% fetal bovine serum (FBS), 100 units/mL Pen-Strep, and 0.01 mg/mL hygromycin at 37 °C in a humidified 5% CO_2 atmosphere. Cells were passaged at 70% to 90% confluency from T-25 flasks using 500 µL of 0.25% trypsin. After the addition of trypsin, cells were placed in a 37 °C incubator for two minutes. To inactivate the trypsin, 1 mL of medium was added to the flask, and 50,000 cells were then seeded to new T-25 well plates containing 5 mL of medium.

Transfection

The two NEIL1 sgRNA plasmids used for CRISPR were provided by Irene Yang. The two gDNA sequences in Table 1 were added into two separate Cas9 plasmids using restriction enzymes, oligonucleotide insertion, and ligation. The mapped plasmid can be seen in Figure 4, and the resulting plasmids were verified by sequencing.

Table 1. gDNA sequences for each gRNA.

gRNA	Length (bp)	Sequence
NEIL1.3R	20	GGCTGAAGCTGAGATGCGGT
NEIL1.4R	20	AAGAGCCGGACATGCCGAAG

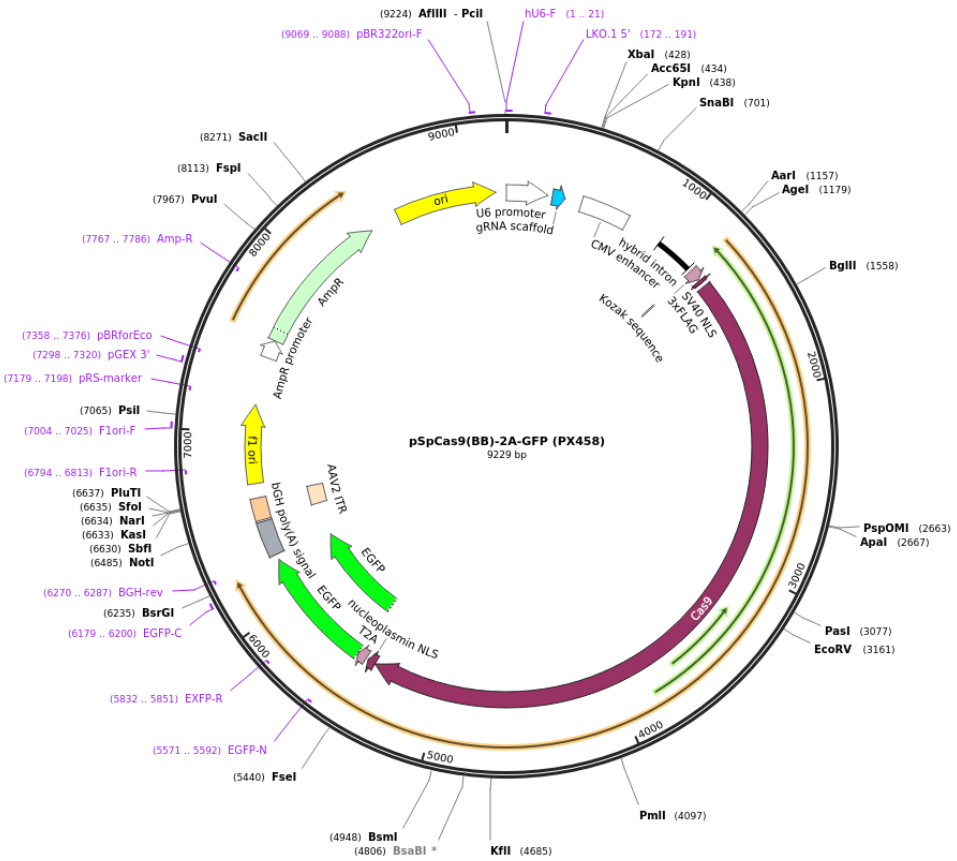


Figure 4: Cas9 from S. pyogenes with 2A-EGFP, and cloning backbone for sgRNA (Addgene plasmid # 48138). The U6 promoter makes small RNAs and regulates the synthesis of sgRNA. The CBh promoter controls how much Cas9 protein is made. The GFP is used as a reporter gene in plasmid (Nerys-Junior et al., 2018).

To accomplish transfection with the sgRNA plasmids, 50,000 RPE-1 WT cells were plated in 60 mm dishes and incubated for 24 hours. After 24 hours, transfection medium was created by diluting 3.75 µL of lipofectamine 3000 Reagent in 250 µL of Opti-MEM medium. A total of 5 µg of DNA from each plasmid was diluted with 250 µL of Opti-MEM medium. These two solutions were combined and mixed and incubated for 10 to 15 minutes at room temperature. After incubation, the entire volume of the DNA transfection reagent complexes was added directly to the plate containing the WT cells. The cells treated with transfection medium and DNA were incubated at 37 °C in a CO_2 incubator for another 24 hours.

Fluorescence-activated cell sorting (FACS)

Suitable 96-well plates were labelled and prepared for FACs single-cell sorting. To each well, 10 µL of medium previously collected from confluent cells were pipetted. This medium was used because it contained the necessary growth factors that facilitate cell division and proliferation.

After 24 hours, the transfected cells were removed from the incubator, and the medium was aspirated. Each well was washed with 1 mL of PBS, and the washings were repeated. 500 µL of 1x trypsin was added to each well, and the plates were incubated 2 to 5 min to detach the cells. After incubation, 2 mL of medium was added to the plate to inactivate the trypsin. The cells were transferred to a collection tube and pelleted in a centrifuge at a speed of 1,000 rcf for 3 to 5 min. The supernatant was aspirated and the cell pellet was resuspend by adding 2 mL of new medium. The cells were then sorted based on GFP signal into the 96 well plates using the FACSAria single cell sorter. The GFP signal indicated a successful transfection as the CRISPR plasmid contains a GFP region as shown in Figure 4.

Parafilm was used to cover the 96-well plates, and the cells were incubated for a minimum of two weeks in a 37 °C CO_2 incubator to allow the cells to become confluent. Once the single cell grows and divides to confluency, the colony is propagated further in a 48-well plate, then a 24-well plate, and finally a 12-well plate, respectively. When the cells were confluent, the genomic DNA was extracted using the QiAmp DNA Mini Prep kit (Qiagen Catalog #: 51304).

Polymerase chain reaction (PCR)

All PCR reagents, which included 2X Phusion High fidelity PCR master mix (New England Biolabs Inc.), 10 µM forward and reverse primers, and genomic DNA, were thawed on ice.

Sequences for the forward and reverse primers appear in Table 2. The volumes of each component used appear in Table 3. After they were mixed, the samples were placed on ice.

Table 2. DNA sequences for the forward and reverse primers used in the PCR reaction.

Primer	Length (bp)	Sequence
NEIL1.1F	19	CTGTCCGGGTAGGGGATTG
NEIL1.1R	21	CGCGACTATGTTCATCACACC

Table 3. Polymerase chain reaction reagents and final concentrations.

Component	50 µl Reaction	Final Concentration
10 µM Forward Primer	2.5 µL	0.5 µM
10 µM Reverse Primer	2.5 µL	0.5 µM
2X Phusion Master Mix	25 µL	1X
genomic DNA	~5 µL(< 250ng)	
Nuclease-free water	to 50 µL	

The mixtures were placed in a Bio-Rad T100 thermocycler, and the reactions were carried out using the settings in Table 4.

Table 4. Time and temperature settings for thermocycler.

STEP	TEMP	TIME
Initial Denaturation	98°C	30 seconds
30 Cycles	98°C	10 seconds
	45-72°C	30 seconds
	72°C	30 seconds
Final Extension	72°C	10 minutes
Hold	4°C	

Gel electrophoresis

Agarose gels (0.8%) were used throughout the duration of the experiments unless stated otherwise. In brief, 0.5 g of agarose was dissolved in 50 mL of 1X Tris-borate-EDTA (TBE) (pH = 8.0). The solution was heated until the agarose dissolved. The solution was left to cool slightly for two minutes. After cooling, 5 µL of 10 mg/mL ethidium bromide was added. The solution was poured into the gel box with an appropriate comb and left to solidify. While the solution solidified, 2 µL of 6X MassRuler Loading dye (ThermoFisher Scientific) was added to 10 µL of sample and mixed thoroughly. Once the gel completely solidified, 5 µL of 103 ng/µL MassRuler DNA ladder (ThermoFisher Scientific) and 10 µL of the loading dye DNA mix were pipetted into the gel. Electrophoresis was run in 1X TBE at the desired voltage and time, which was typically 65V for 2 hrs.

Whole genome sequencing

WT and N1KO cells were grown in replicates of 4 in 12-well plates until 70% confluency. The DNA was extracted from the cells and purified using the QiAmp DNA Mini Prep kit (Qiagen Catalog #: 51304). Whole genome sequencing was performed on the DNA samples using the xGen DNA Library Prep EZ UNI kit (IDT, Catalog #: 412119677). Once the DNA libraries were created, quality control metrics were performed using a KAPA Library Quantification Kit (Roche) and Agilent's 4200 Tapestation to check for quality, base pair size and quantification of the libraries. The libraries that passed the quality control parameters were sequenced with the NovaSeq 6000 (XP Workflow - SP 2x150 - 300 Cycle v1.5 configuration). Following sequencing, the results were processed by Nizar Drou, Senior Research Scientist in the Bioinformatics Core of the Center for Genomics and Systems Biology at NYU Abu Dhabi. The resulting VCF files were analyzed

against the Genome Reference Consortium Human Build 38 (GRCh38). A potential off target list for the two guide RNAs used was compared with the KO samples.

RNAseq DESeq analysis

WT and KO cells were grown in replicates of 4 in 48-well plates until 70% confluency. The RNA was extracted from these cells using an RNeasy Plus kit (Qiagen, Catalog#: 74134). After RNA was isolated, an RNA 6000 Nano Kit (Agilent, Catalog #: 5067-1511) was used to assess the quality of the extracted RNA. After determining the quality for each RNA sample, the RNA sequencing libraries were created using the TruSeq Stranded Total RNA Library Prep Gold Kit (Illumina, Catalog#: 20020598). Once the RNA libraries were created, quality control metrics were performed using a KAPA Library Quantification Kit (Roche) and Agilent's 4200 Tapestation to check for quality, base pair size, and quantification of the libraries. Pooled RNAseq libraries were submitted and sequenced by NYUs's genomics core using the NovaSeq 6000 XP (Workflow - SP 2x150 - 300 Cycle v1.5 configuration).

Following sequencing, the results were again processed by Nizar Drou. The resulting BAM files were analyzed in Integrative Genomics Viewer (IGV) against the reference genome GRCh38. Differential expression analysis was performed using DESeq2. Finally, Gene Ontology (GO) Term analysis was performed to identify biologic function, process, or location of impacted genes.

Results

N1KO of hTERT RPE1 cells was generated using CRISPR9 technology.

Cells expressing GFP were plated in 96 well plates using Fluorescence-activated cell sorting (FACS). Only two colonies successfully grew withing the 96 well plate: NEIL1 colony 1 and colony 2. The PCR product for the N1KO colony 1 sample traveled lower on the gel when compared to the WT sample that had an amplicon size of approximately 660 base pairs (Figure 5).

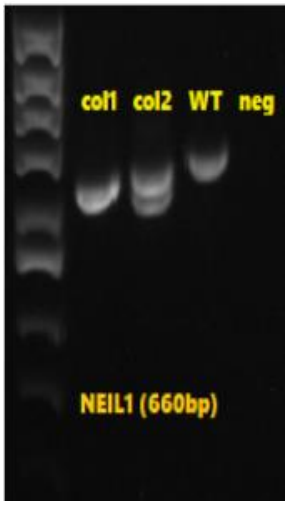


Figure 5: Gel electrophoresis result confirming NEIL1 col1 traveled down further than WT. This colony was presumed to have a successfully homozygous deletion for the region and were then sequenced using Sanger sequencing to determine the exact base pairs that were deleted. Alternatively, colony 2 produced 2 bands. This signified that colony 2 contained heterozygosity within the deleted region, producing an inadequate deletion for this study.

Sanger sequencing confirmed a 22 base pair deletion in the N1KO.

After confirming the deletion obtained from the Gel, Sanger sequencing was used to determine the exact DNA deletion sequence of the N1KO samples. The resulting FASTA file obtained from Sanger sequencing was blasted to the FASTA sequence of the WT samples using NCBI. The query showed a 22 base pair deletion present in the DNA of the NEIL1 colony 1 KO. This confirmed

that the CRISPR Cas9 complex was successful in deleting a part of the sequence. Further analysis was conducted, and the protein sequence was generated from the N1KO genomic sequence. Premature stop codons in the exonic regions of the NEIL1 colony 1 KO protein signaled that this cells would produce a shorter, nonfunctional protein (Figure 6).

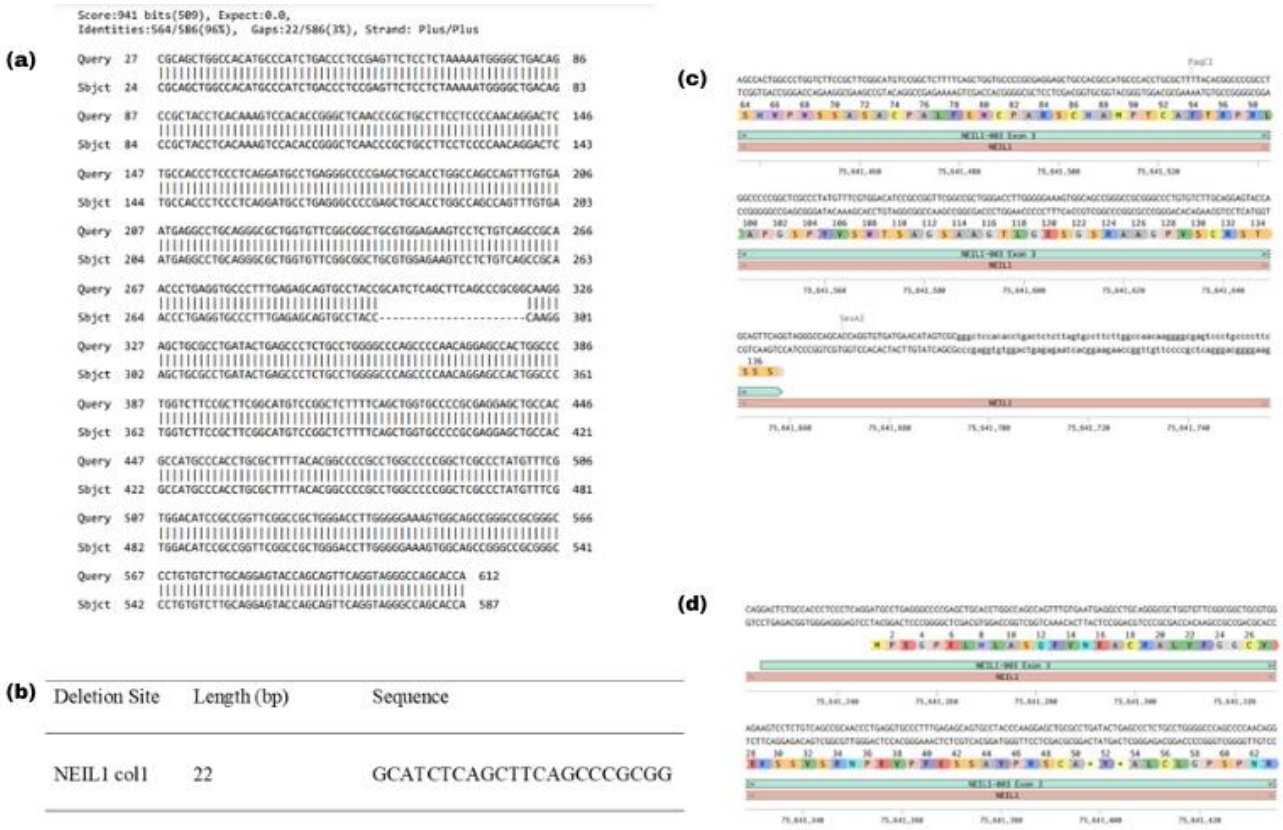


Figure 6: **(a)** Blast results between the WT RPE-1 and N1KO samples. **(b)** 22 base pair deletion sequence that is present within the NEIL1 colony 1 KO. **(c)** WT protein sequence (d) NEIL1 colony 1 protein sequence with premature stop codon.

<u>Whole genome sequencing and off target analysis showed that there are no off targets present in within the NEIL1 colony 1 KO.</u>

A list of the possible known off targets within the exonic regions for the 2 gRNA's were generated using a software offered by IDT (Table 5). After comparing this list with the WT and KO results,

there appeared to be no genomic variants present for the expected off target regions within the KO colony. The CRISPR Cas-9 and gRNA system was effective and precise, as demonstrated by the successful generation of a homozygous KO for the NEIL1 region.

Table 5. List of possible off targets for the 2 gRNA's used to generate the N1KO.

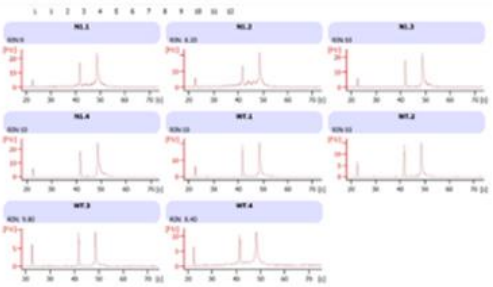
NEIl 1 gene w/ gRNA off targets						
gRNA Name	gRNA sequence	Gene	Locus	off target Sequence	PAM	# of mismatches
NEIL1.3 R	GGCTGAAGCTGAGATGCGGT	NEIL1	chr15:-75349038	N/A	AGG	N/A
		TCERG1	chr5:-146447241	AGCAGAAGTCTGAGATGCGGT	TGG	3
		NUP214	chr9:-131233522	GGAAGAAGCTGAGGATGCGGT	TGG	3
		CNEP1R1	chr16:+50025260	CGCGGAAGCTGCGATGCGGA	CAG	4
		UNC13C	chr15:+53916857	CACTGAAGCTGGGATGCTGT	CAG	4
		LOC105373027	chr22:-38079997	GGCTGCA-CTGAGATGGGGT	GAG	3
		GJB4	chr1:-34763382	GGAGGAAGCTGGGATGAGGT	GGG	4
		LOC388780	chr20:-2236470	GACGGAAGCGGAGATGCTGT	CAG	4
		RNLS	chr10:+88312168	AGCTGAAGCTGAGATGGGAT	AAG	3
		NF2	chr22:+29683713	GGCTG-AGCAGAGATGTGGT	CAG	3
NEIL1.4 R	AAGAGCCGGACATGCCGAAG	NEIL1	chr15:-75349138	N/A	CGG	N/A
		LOC729737	chr1:+135070	AGGAGCTGGACCTGCCGAAG	TGG	3
		LOC100132062	chr5:-181327877	AGGAGCTGGACCTGCCGAAG	TGG	3
		LINC01002	chr19:+197202	AGGAGCTGGACCTGCCGAAG	TGG	3
		LINC00999	chr10:-38451835	AGGAGCTGGACCTGCCGAAG	CGG	3
		LINC01001	chr11:+127284	AGGAGCTGGACCTGCCGAAG	TGG	3
		LOC100132287	chr1:+491053	AGGAGCTGGACCTGCCGAAG	TGG	3
		LOC100133331	chr1:+726056	AGGAGCTGGACCTGCCGAAG	TGG	3
		AP1G1	chr16:-71729521	AACAGAAGGAAATGCCGAAG	CAG	4
		FZD5	chr2:+207767748	TGGAGCTGGCCATGCCGAAG	AAG	4
		FZD10	chr12:-130163895	GCGAGCTGGCCATGCCGAAG	TAG	4
		FZD8	chr10:+35640195	TGGAGCTGGCCATGCCGAAG	AAG	4
		FZD7	chr2:-202035681	TGGAGCTGGCCATGCCGAAG	AAG	4
		FZD9	chr7:-73434972	GCGAGCTGGCCATGCCGAAG	TAG	4
		LINC00265	chr7:-39794427	AGGAGCTGGGCCTGCCGAAG	TGG	4
		LOC729218	chr4:-118634809	AGGAGCTGGACCTGCTGAAG	TGG	4
		LINC01000	chr7:-128654800	AGTAGCTGGACCTGCCGAAG	TGG	4
		GTF2IP20	chr1:+223951559	AGTAGCTGGACCTGCCGAAG	TGG	4
		EPS15L1	chr19:+16355480	GAGAGACGGACCTGCAGAAG	TGG	4
		HIVEP3	chr1:-41848312	AAGAGCCAGAGAAGCCAAAG	AAG	4
		ATXN7L2	chr1:+109489475	GAGAGCCTGACTTGCTGAAG	GAG	4
		TCHH	chr1:-152111107	AAGAGCCG--CATCCCGAAG	TGG	3
		ACSL6	chr5:-131966472	AAGATCCAGACAGGACGAAG	GAG	4

DESeq2 analysis showed that 66 genes were differentially expressed in the N1KO cells when compared to the wildtype cells.

After ensuring the accuracy and efficiency of our gRNA system, RNA sequencing was performed to capture the effects of the Knockout on the transcriptome. Extracted RNA from 4 replicates of WT and N1KO cells were tested for quality using the Bioanlyzer. All the RNA samples had RIN

scores above 7 which denoted high quality, passing the quality control metric for use in Illumina sequencing kits (Figure 7a). The Illumina libraries were tested for quality using Agilent tapestation high sensitivity reagents which showed that 4 out of the 8 libraries had low primer and adapter contamination (Figure 7b). The other 4 samples that appeared to have high primer and adapter contamination were purified using ampure beads and rerun resulting in high levels of purity. One of the NEIL1 samples (N1.4) however appeared to have no DNA whatsoever. This may have been attributed to the samples being washed away during the 2nd ampure purification step. This samples were not used for further quality control. Kappa qPCR performed on the libraries also suggested that enough DNA was present for sequencing- 3nM of DNA for each library.

(a)



Sample Name	Result Label	Conc ug/uL	.5ug desired for library prep
N1.1	RIN:9	0.171	2.9uL
N1.2	RIN: 8.20	0.196	2.5uL
N1.3	RIN:10	0.109	4.58uL
N1.4	RIN:10	0.117	4.27uL
WT.1	RIN:10	0.099	5.05uL
WT.2	RIN:10	0.061	8.19uL
WT.3	RIN: 9.80	0.043	11.6uL
WT.4	RIN: 8.40	0.101	4.95uL

(b)

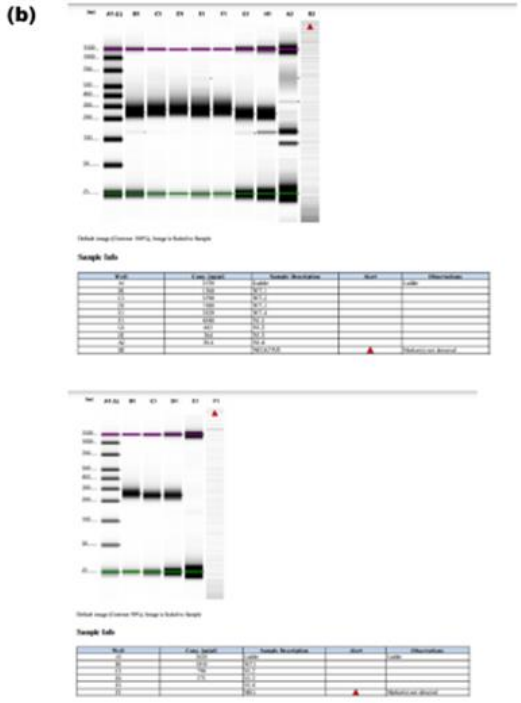


Figure 7. (a) RIN scores for WT and N1KO cell lines (b) Tape station results for completed WT and N1KO libraries

A 22 base pair deletion was present in the transcripts for the NEIL1 gene (Figure 8). This deletion corresponded to the deletion site found using sanger sequence analysis.

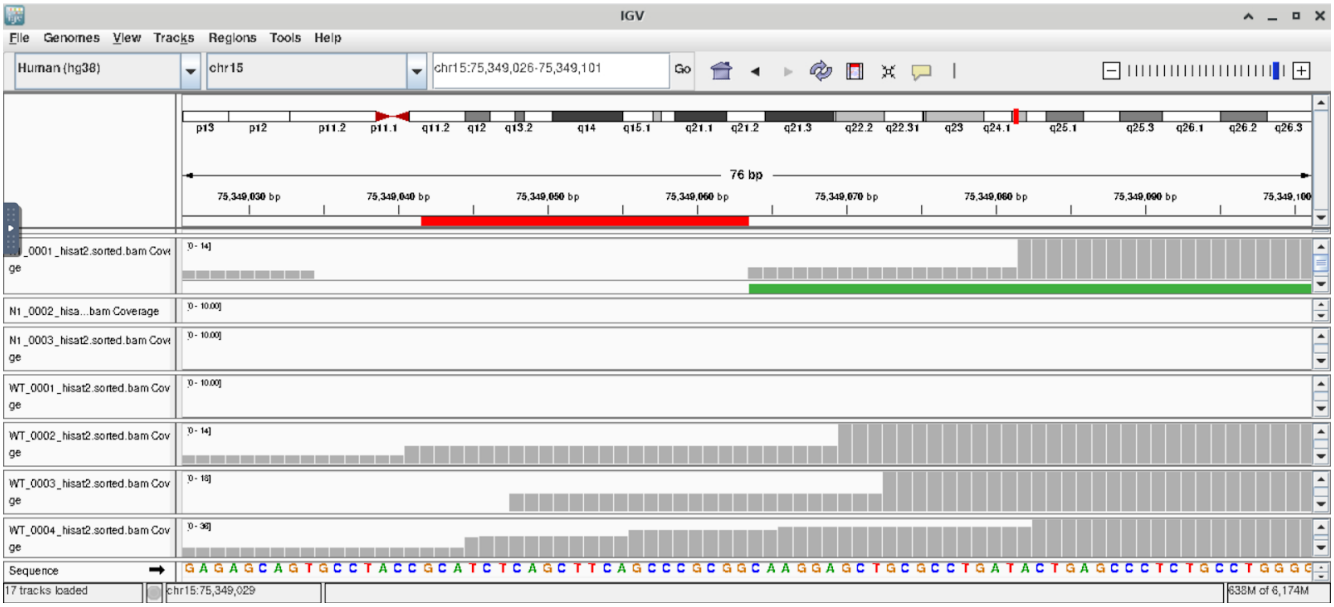


Figure 8: IGV for WT and N1KO RNAseq transcripts.

DESeq analysis showed 66 genes were differentially expressed (Table 6). Of those 66 genes, 23 were upregulated and 44 were down regulated. After analyzing the entire list of genes using GO terms, many of the upregulated genes were related to metabolism (Table 7). The downregulated genes appear in Table 8.

Table 6. Differentially expressed genes

Pvalue	0.5	
$\log_2$FC	≥ 1	≤ -1
Number of Genes	23	43

Table 7. GO Terms from upregulated genes

	GOBPID	Term
1	GO:0044238	primary metabolic process
2	GO:0071704	organic substance metabolic process
3	GO:0006807	nitrogen compound metabolic process
4	GO:0008152	metabolic process

Table 8. GO Terms from downregulated genes

	GOBPID	Term
1	GO:0009725	response to hormone
2	GO:0061061	muscle structure development
3	GO:0001655	urogenital system development
4	GO:0006810	transport
5	GO:0051234	establishment of localization
6	GO:0010033	response to organic substance
7	GO:0007165	signal transduction
8	GO:0003002	regionalization
9	GO:0003012	muscle system process
10	GO:0007389	pattern specification process

Higher levels of oxygen consumption were present in the N1KO cells compared to the WT.

Since GO term analysis showed an upregulation for metabolism in the N1KO cells, mitochondrial function within these cells was assessed with the Agilent Seahorse XF system. Cells were shipped to NYU's Abu Dhabi facility, and Agilent's XF Substrate Oxidation Stress Test kit (Agilent, Catalog#: 103672(3)(4)-100 & 103693-100) was used to measure oxygen consumption. The work was performed by Haifa Alkhalifa. Sample results are shown in Figure 9. The basal respiration for the N1KO cells was roughly 25% higher than the WT. Treatment with FCCP, a drug used to stimulate the respiratory chain to operate at maximum capacity, showed that the reserve capacity for the KO cells was 40% higher than the WT. Collectively, these data indicate a significant increase in oxygen consumption in the N1KO cells relative to WT.

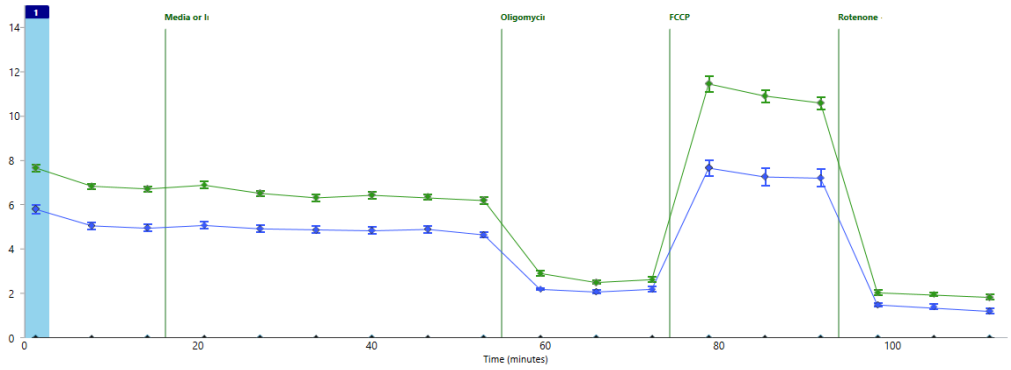


Figure 9: Seahorse results for oxygen consumption in N1KO and WT cells.

Discussion

This study used CRISPR Cas-9 technology to knockout the NEIL1 gene from hTERT RPE-1 cells. The KO was verified by a 22 base pair deletion within exon 2, and the DNA sequencing showed no off-targets were generated from the use of the two gRNAs. The in silico characterization of protein for the N1KO showed premature stop codons present in the amino acid sequence. This suggests that the KO would affect the normal function of the NEIL1 gene by altering both mRNA transcripts and amino acid sequences. It is important to note that any differences in the KO phenotype compared to the WT phenotype appear to be solely due to the absence of the NEIL1 gene.

RNA-seq results revealed that 66 genes were affected by the loss of NEIL1, with some being upregulated and others downregulated. The most affected upregulated pathways were related to metabolism, which is consistent with a previous study in which mice lacking the NEIL1 gene displayed increased metabolic syndrome (Vartanian et al., 2006). It is most likely that the absence of the NEIL1 gene leads to an increase in oxidative stress due to the overproduction of reactive oxygen species, although the underlying mechanism of this change remains elusive. This would affect the mitochondrial DNA and respiratory chain, leading to changes in metabolic processes. This is consistent with the known role of NEIL1 in repairing DNA damage caused by oxidative stress, and points to the possibility that the mitochondria within these cells are compromised. It is possible that ROS production is increased in the N1KO cells, but this remains to be tested experimentally. Oxygen consumption studies using the Agilent Seahorse system suggest that the N1KO cells exhibit enhanced ATP production compared to the WT cells as exemplified by the increased oxygen consumption rate.

Future Studies

While this study was successful in confirming NEIL1's relationship to mitochondria function, further biological tests on the N1KO need to be done to quantify the relative differences in oxygen consumption and mitochondria function when compared to the WT cell line. Additionally, since the mitochondria plays a large role in creating ROS that damage DNA, characterizing the mitochondria is of great interest.

The study used RNA-seq and pathway analysis to determine the transcription profiles of the entire N1KO and WT cell lines. Previous single-cell RNA-seq studies conducted by a lab member, Irene Yang, on NEIL3 KO cell lines showed that subsets of NEIL3 KO cells had different transcription profiles. This raises two possibilities: (1) the entire aggregate of N1KO cells exhibit the same transcription profiles or (2) there may be subgroups of these cells that exhibit different transcription profiles. If the latter possibility is correct, it is important to determine the number and composition of these subgroups, as well as the genes that are over- or under-expressed within each subgroup. To further investigate this, single-cell RNA-seq and pathway analysis for the N1KO and WT cells will need to be conducted. This can provide a more detailed understanding of the changes in gene expression that result from knocking out the NEIL1 gene. Additionally, these results can also shed light on possible subpopulations with distinct properties within the N1KO cells that might be important in disease development or progression.

References

Albelazi, M.S., Martin, P.R., Mohammed, S., Mutti, L., Parsons, J.L., and Elder, R.H. (2019). The Biochemical Role of the Human NEIL1 and NEIL3 DNA Glycosylases on Model DNA Replication Forks. Genes (Basel) *10*. 10.3390/genes10040315.

Chatterjee, N., and Walker, G.C. (2017). Mechanisms of DNA damage, repair, and mutagenesis. Environ Mol Mutagen *58*, 235-263. 10.1002/em.22087.

Cooper, T.A., Wan, L., and Dreyfuss, G. (2009). RNA and disease. Cell *136*, 777-793. 10.1016/j.cell.2009.02.011.

Divakaruni, A.S., Paradyse, A., Ferrick, D.A., Murphy, A.N., and Jastroch, M. (2014). Analysis and interpretation of microplate-based oxygen consumption and pH data. Methods Enzymol *547*, 309-354. 10.1016/B978-0-12-801415-8.00016-3.

Hakem, R. (2008). DNA-damage repair; the good, the bad, and the ugly. EMBO J *27*, 589-605. 10.1038/emboj.2008.15.

Han, D., Schomacher, L., Schule, K.M., Mallick, M., Musheev, M.U., Karaulanov, E., Krebs, L., von Seggern, A., and Niehrs, C. (2019). NEIL1 and NEIL2 DNA glycosylases protect neural crest development against mitochondrial oxidative stress. Elife *8*. 10.7554/eLife.49044.

Lambert, A.J., and Brand, M.D. (2009). Reactive oxygen species production by mitochondria. Methods Mol Biol *554*, 165-181. 10.1007/978-1-59745-521-3_11.

Mullins, E.A., Rodriguez, A.A., Bradley, N.P., and Eichman, B.F. (2019). Emerging Roles of DNA Glycosylases and the Base Excision Repair Pathway. Trends Biochem Sci *44*, 765-781. 10.1016/j.tibs.2019.04.006.

Nerys-Junior, A., Braga-Dias, L.P., Pezzuto, P., Cotta-de-Almeida, V., and Tanuri, A. (2018). Comparison of the editing patterns and editing efficiencies of TALEN and CRISPR-Cas9 when targeting the human CCR5 gene. Genet Mol Biol *41*, 167-179. 10.1590/1678-4685-GMB-2017-0065.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat Protoc *8*, 2281-2308. 10.1038/nprot.2013.143.

Rosenquist, T.A., Zaika, E., Fernandes, A.S., Zharkov, D.O., Miller, H., and Grollman, A.P. (2003). The novel DNA glycosylase, NEIL1, protects mammalian cells from radiation-mediated cell death. DNA Repair (Amst) *2*, 581-591. 10.1016/s1568-7864(03)00025-9.

Sobanski, T., Rose, M., Suraweera, A., O'Byrne, K., Richard, D.J., and Bolderson, E. (2021). Cell Metabolism and DNA Repair Pathways: Implications for Cancer Therapy. Front Cell Dev Biol *9*, 633305. 10.3389/fcell.2021.633305.

Vartanian, V., Lowell, B., Minko, I.G., Wood, T.G., Ceci, J.D., George, S., Ballinger, S.W., Corless, C.L., McCullough, A.K., and Lloyd, R.S. (2006). The metabolic syndrome resulting from a knockout of the NEIL1 DNA glycosylase. Proc Natl Acad Sci U S A *103*, 1864-1869. 10.1073/pnas.0507444103.

Zhou, J., Chan, J., Lambele, M., Yusufzai, T., Stumpff, J., Opresko, P.L., Thali, M., and Wallace, S.S. (2017). NEIL3 Repairs Telomere Damage during S Phase to Secure Chromosome Segregation at Mitosis. Cell Rep *20*, 2044-2056. 10.1016/j.celrep.2017.08.020.