Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Kristin Leigh Limpose Date

Overexpression of the base excision repair glycosylase, NTHL1, causes cellular transformation

By

Kristin L. Limpose
Doctor of Philosophy

Graduate Division of Biological and Biomedical Sciences
Cancer Biology

Paul W. Doetsch
Advisor

Anita H. Corbett
Committee Member

Sumin Kang
Committee Member

David Yu
Committee Member

Wei Zhou
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Overexpression of the base excision repair glycosylase, NTHL1, causes cellular transformation

By

Kristin L. Limpose
B.S., Clemson University, 2011

Advisor: Paul W. Doetsch, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
In partial fulfillment of the requirements for the degree of
Doctor of Philosophy
In Cancer Biology
2017

Abstract

Overexpression of the base excision repair glycosylase, NTHL1, causes cellular transformation

By Kristin L. Limpose

Base excision repair (BER), which is initiated by the DNA *N*-glycosylase proteins, is the frontline for repairing potentially mutagenic DNA base damage. The NTHL1 glycosylase, which excises DNA damage caused by reactive oxygen species, is thought to be a tumor suppressor. In addition to *NTHL1* loss of function mutations, our analysis and data mining of cancer genomic datasets in cBioPortal reveal that *NTHL1* frequently undergoes amplification or mRNA upregulation in certain cancer types. The contribution that NTHL1 overexpression could have to cancer has not previously been explored. We report that NTHL1 protein levels are elevated in a panel of lung cancer cell lines. To address the functional consequences of NTHL1 overexpression, transient NTHL1 overexpression was employed. Both NTHL1 and a catalytically-dead NTHL1 (CATmut) induce DNA damage and genomic instability in non-transformed human bronchial epithelial cells (HBEC) when overexpressed. Strikingly, overexpression of either NTHL1 or CATmut causes replication stress signaling and a decrease in homologous recombination (HR) activity. HBEC cells that overexpress NTHL1 and CATmut acquire the ability to grow in soft agar and exhibit loss of contact inhibition, suggesting that a catalytic-independent mechanism of NTHL1 contributes to the acquisition of cancer phenotypes. Our results demonstrate that dysregulation of a base excision repair protein, NTHL1, can induce genomic instability, interfere with HR repair activity, and result in the acquisition of cellular transformation markers.

Overexpression of the base excision repair glycosylase, NTHL1, causes cellular transformation

By

Kristin L. Limpose
B.S., Clemson University, 2011

Advisor: Paul W. Doetsch, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
In partial fulfillment of the requirements for the degree of
Doctor of Philosophy
Cancer Biology
2017

Acknowledgements

I would like to start by thanking the numerous members of the Doetsch and Corbett labs, past and present. Your scientific expertise knows no limits, and I am unbelievably grateful to have been imbued in your knowledge. Without your guidance, I would not be the scientist I am today. Beyond the science, you have extended me kindness and friendship, without which, I don't know how I would have survived this process. To Erica and Natasha, you have become my science family, and I cannot detail how much I have learned from you. To try to do so here could not possibly do your mentorship justice. To Paul and Anita, my science "dad and mom," thank you for teaching me how to crawl and walk as a scientist. I am a better scientist, writer, and person for having gone through this process with you. To everyone in the Corbett lab, thank you for your friendships and listening ears! PS, I owe you guys about 100 pounds of coffee from over the last six years….

Thank you to the Cancer Biology Program, Laney Graduate School, and the GDBBS for taking a chance on me six years ago. As part of the inaugural class of Cancer Biology, I am humbled and grateful every day that I was able to attend Emory University. Through Emory, I have undoubtedly met some of the finest and brightest faculty and students that I now consider friends and mentors. I look forward to continuous learning from these people, and seeing what the scientific future has in store for all of us! I especially have to thank my dissertation committee for all of their patience though the years: Paul, Anita, David, Sumin, and Wei. Your leadership and direction molded my project into a success story.

To my friends in Cancer Biology and all my friends across the GDBBS, I could not have made it through these years without you. Thank you for always being there. We always pushed each other forward.

Lastly, I have to thank my family and loved ones, especially my parents Terri and John Limpose. You are a constant source of love and support. Thank you for always pushing me forward, even when I was hesitant. Thank you for showing me what a strong work ethic looked like; I strive to emulate your hard work each day. I know it's been a long six years but we did it! I could not have asked for a better support system.

Table of Contents

List of Figures

Abbreviations

In order of appearance

DNA: deoxyribonucleic acid

ROS: reactive oxygen species

UV: ultraviolet light

8-oxoG: 8-oxoguanine

APOBEC: Apolipoprotein B MRNA Editing Enzyme Catalytic Subunit

BER: base excision repair

AP: apurinic/apyrimidinic

APEX1/APE1: apurinic/apyrimidinic endonuclease I

UNG1/2: uracil DNA glycosylase

SMUG: single-strand-selective monofunctional uracil-DNA glycosylase 1

TDG: thymine DNA glycosylase

MBD4: methyl-CpG binding domain 4, DNA glycosylase

MPG: N-methylpurine DNA glycosylase

MUTYH: MutY DNA glycosylase

POLB: DNA polymerase β

LIGIII: DNA ligase III

FEN1: flap endonuclease I

LIG1: DNA ligase I

SP-BER: short patch base excision repair

LP-BER: long patch base excision repair

NTHL1/Ntg1/Ntg2/mNTH1: nth endonuclease like III

OGG1: 8-oxoguanine DNA glycosylase

NEIL1: nei like DNA glycosylase 1

NEIL2: nei like DNA glycosylase 2

NEIL3: nei like DNA glycosylase 3

dRP: deoxyribose phosphate

PUA: phospho-α,β-unsaturated aldehyde

ATP: adenosine triphosphate

PARP: poly(ADP-Ribose) polymerase

NAD^+: nicotinamide adenine dinucleotide

NER: nucleotide excision repair

XPG: ERCC excision repair 5, endonuclease

CSB: cockayne syndrome B

TSC2: tuberous sclerosis complex 2

SLC9A3R2: sodium/hydrogen exchanger 3 kinase A regulatory protein

mRNA: messenger ribonucleic acid

NCBI: national center for biotechnology information

MTS: mitochondrial targeting sequence

NLS: nuclear localization sequence

HhH: helix-hairpin-helix

4Fe-4S: iron-sulfur cluster

Tg: thymine glycol

YB-1: Y-box binding protein 1

Tg:A: thymine glycol paired with adenine

Tg:G: thymine glycol paired with guanine

COS-7: African green monkey kidney cells

H_2O_2: hydrogen peroxide

MYCN: myelocytomatosis viral related oncogene, neuroblastoma derived

EGFR: epidermal growth factor receptor

ERBB2: erb-b2 receptor tyrosine kinase 2

HR: homologous recombination

XPC: xeroderma pigmentosum, complementation group C

GG-NER: global genome nucleotide excision repair

XPA: xeroderma pigmentosum, complementation group A

XPB: xeroderma pigmentosum, complementation group B

FapyG: 2,6-diamino-4-hydroxy-5-formamidopyrimidine

FapyA: 4,6-diamino-5-formamidopyrimidine

TC-NER: transcription coupled nucleotide excision repair

DSBs: double strand breaks

NHEJ: non-homologous end joining

c-NHEJ: classical non-homologous end joining

KU70/80: X-ray repair cross complementing 6/5

alt-NHEJ/alt-EJ: alternative nonhomologous end joining

RNAi: RNA interference

FANCA: fanconi anemia complementation group A

BRCA2: breast cancer gene 2

PALB2: partner and localizer of BRCA2

Rad52: RecA homolog 52

RPA: replication protein A

CATmut: lysine to glutamine mutation at position 220 of NTHL1 (K220Q)

HBEC: human bronchial epithelial cells

PNK: polynucleotide kinase phosphatase

NSCLC: non-small cell lung cancer

CPT: camptothecin

HU: hydroxyurea

SDS: sodium dodecyl sulfate

ATR: ataxia telangiectasia and Rad3 related kinase

KRAS: V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog

GFP: green fluorescence protein

HRP: horseradish peroxidase

DR-U2OS: direct repeat osteosarcoma cell line

CSK: cytoskeletal buffer

γH2AX: phosphor serine 139 histone 2A variant x

53BP1: tumor protein p53 binding protein 1

FACS: fluorescence-activated cell sorting

qRT-PCR: quantitative real-time polymerase chain reaction

NT: non-transfected control

Vector: empty vector control

DR-GFP: direct repeat green fluorescence protein assay

MN: micronucleus

CHK1: cell cycle checkpoint kinase 1

CHK2: cell cycle checkpoint kinase 2

RAD51: RecA homolog 51

IR: ionizing radiation

MMR: mismatch repair

MSH6: MutS homolog 6

MSH3: MutS homolog 3

STRING: functional protein association network

PTMs: post-translational modification

TCGA: the cancer genome atlas

Chapter 1: Introduction

1.1 Oxidative Stress

Section 1.1 is excerpted from "DNA Damage", authored by Kristin Limpose, Anita H. Corbett, and Paul W. Doetsch, Encyclopedia of Life Sciences (online), (Sep 2014). In: eLS. John Wiley & Sons Ltd. http://www.els.net. Permissions attached.

[doi: 10.1002/9780470015902.a0000557.pub3]. Based in part on the previous version of this eLS article 'DNA Damage' (2009) by Bryn S Moore, Lydia P Morris and Paul W Doetsch.

KL, AHC, and PWD contributed to updating, writing and editing the manuscript.

Under cellular physiological conditions of temperature, pH and metabolic activity, DNA is susceptible to spontaneous chemical alterations of the bases, sugar, and phosphate components. Each of these DNA bases are reactive centers that can undergo spontaneous hydrolysis, oxidation and reaction with electrophiles present in the cell.

1.1.1 Causes of Reactive Oxygen Species

Reactive oxygen species (ROS) found in the forms of singlet oxygen, hydrogen peroxide, superoxide anion radicals and Fenton reaction generated hydroxyl radicals, are a main source of DNA damage to the nuclear and mitochondrial genomes [1]. Importantly, cellular exposure to ROS occurs through both endogenous and exogenous means. Leakage of ROS from the mitochondria during oxidative phosphorylation occurs when oxygen is incompletely reduced to water, and is one of the main cellular sources of ROS [2, 3]. Other endogenous sources of ROS include peroxisomal metabolism [4], nitric oxide synthesis [5], and inflammatory responses [6, 7]. Cellular exposure to exogenous agents including ultraviolet light (UV), ionizing radiation, and certain chemicals all produce ROS that damage DNA. Genomes are therefore exposed to substantial oxidative assaults, and estimates of 'oxidative hits' per cell per day are between several hundred to tens of thousands [8].

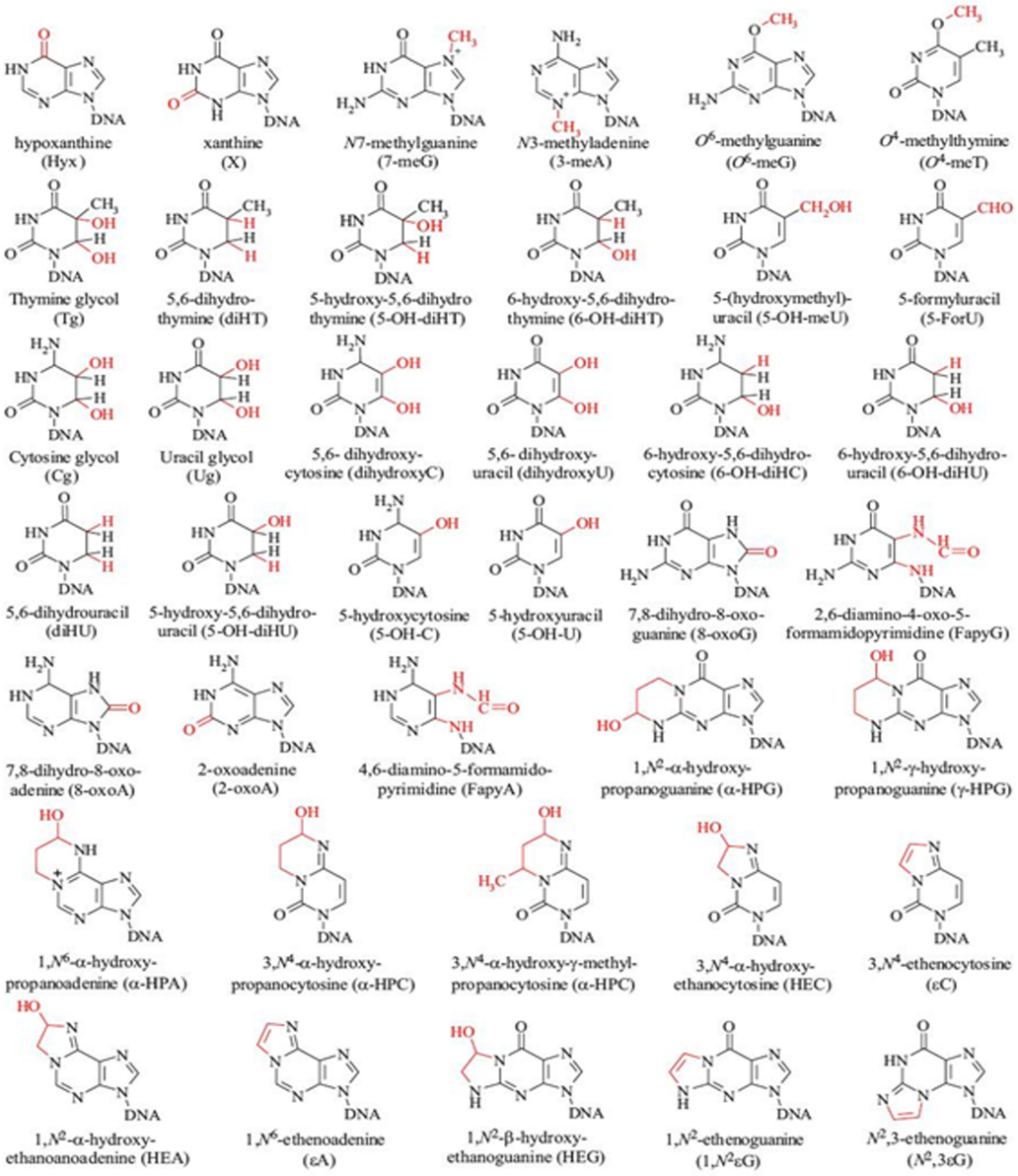

Figure 1.1. Major Endogenous DNA Damages. The major endogenous DNA damages are depicted with alterations highlighted in red. Important examples shown here are, uracil (from cytosine deamination), cyclobutane pyrimidine dimer (produced by UV light exposure), 8-oxoguanine (produced by reactive oxygen species) and O6-methylguanine (produced by alkylating agents). Reproduced from Krwawicz et al., 2007.

1.1.2 DNA Damage Caused by Oxidative Stress

ROS are distinct from other classes of DNA-damaging agents because ROS can elicit a multitude of damage that includes direct strand breaks (via oxygen radical attack of deoxyribose) and chemically diverse base modifications, many of which have harmful biological effects. ROS-induced base damages include some 40-60 damage products with varying frequency and type (**Figure 1.1**) [9, 10]. Purine and pyrimidine ring saturation, ring cleavage and contraction products, as well as exocyclic ring additions are among some of the damage products. Major ROS-induced adducts include 8-oxoguanine (8-oxoG), thymine glycol, and 5-hydroxycytosine (**Figure 1.1**). Certain adducts, such as 8-oxoG, are mutagenic and can be bypassed by both DNA and RNA polymerase, leading to mispairing of 8-oxoG with adenine. The subsequent mispairing results in permanent G >T transversion mutations in DNA and may cause mutant RNA transcripts capable of producing altered protein products, termed transcriptional mutagenesis [11]. In humans, oxidative damage contributes indirectly and directly to pathologies including neurological disorders, heart disease, cancer, and the aging process [12].

As cellular ROS is continuously generated and imposes damage on the genome, oxidative damage is considered a ubiquitous threat to the genome. Cancer seems to be particularly affected by ROS, as evidenced by the recent sequencing of multiple tumor genomes. Approximately 21 cancer-associated signatures were revealed from sequencing 30 cancer types, and some signatures were attributable to lapses in DNA repair and DNA polymerase mutations [13]. Other signatures are attributable to the APOBEC cytosine deaminase protein family, resulting in frequent C >T transitions. These signatures are

context-dependent with respect to the nucleotides immediately 5' and 3' to the mutated base [13]. For many of the signatures, no causative mechanism was identified, however oxidative damage may be responsible for some context-dependent signatures, such as the signatures representing C >A mutations.

A related study investigated context-dependent and mutational signatures in breast cancer. Interestingly, when the number of mutations versus distance between mutations was plotted, distinct clusters of mutations showing short intermutational distances were clustered away from the bulk of sequenced mutations, and were termed "mutational showers". Heavier clusters of damage, often corresponding to C >T mutations were termed as "mutational storms". Together these mutational showers and storms were termed "kataegis", derived from the Greek word for thunder. Further analysis showed that mutational storms and showers corresponded to sites of rearrangements, suggesting that specific clusters of base mutations, some of which are potentially caused by ROS, are hot spots for translocations [14].

1.2 Base Excision Repair (BER)

DNA contained in both the nuclear and mitochondrial cellular compartments is subject to damage from multiple sources [1, 15, 16]. DNA damage is broadly classified as occurring through exogenous and/or endogenous sources. Diverse classes of DNA damage are caused by exogenous sources such as UV, ionizing radiation, alkylating agents, and heavy metals [17-21]. Endogenous sources of damage such as reactive oxygen species (ROS) are generated during normal metabolic functions as well as various cellular

transactions [16, 22]. Alkylation DNA damage can also endogenously arise through the methyl group donor S-adenosylmethionine [23]. Spontaneous deamination of cytosine erroneously places uracil in DNA [24], and the *N*-glycosidic bond of purines and pyrimidines is susceptible to hydrolysis via base protonation by water (**Figure 1.2A**), leading to a base-free site (**Figure 1.2B**) [25]. Cells have therefore evolved numerous DNA repair and tolerance pathways to protect the genome from DNA damage.

1.2.1 BER Enzymatic Steps

Base excision repair (BER) is crucial for maintaining genome integrity through repair of non-bulky base damage in both the nuclear and mitochondrial cellular compartments [1, 15, 22]. The span of damage repaired by BER includes oxidation and alkylation of DNA bases, deamination products, DNA base open ring products, and repair of apurinic/apyrimidinic sites [1, 23, 25]. BER is an evolutionarily conserved DNA repair pathway and is generally composed of five coordinated steps: (1) recognition of and removal of a damaged DNA base by the *N*-glycosylase proteins to generate an AP site, (2) cleavage of the AP site by an AP endonuclease or glycosylase AP lyase activity, (3) DNA end cleaning of polymerase blocking moieties, (4) gap filling by a DNA polymerase, and (5) DNA ligase sealing of the remaining nick [1, 23].

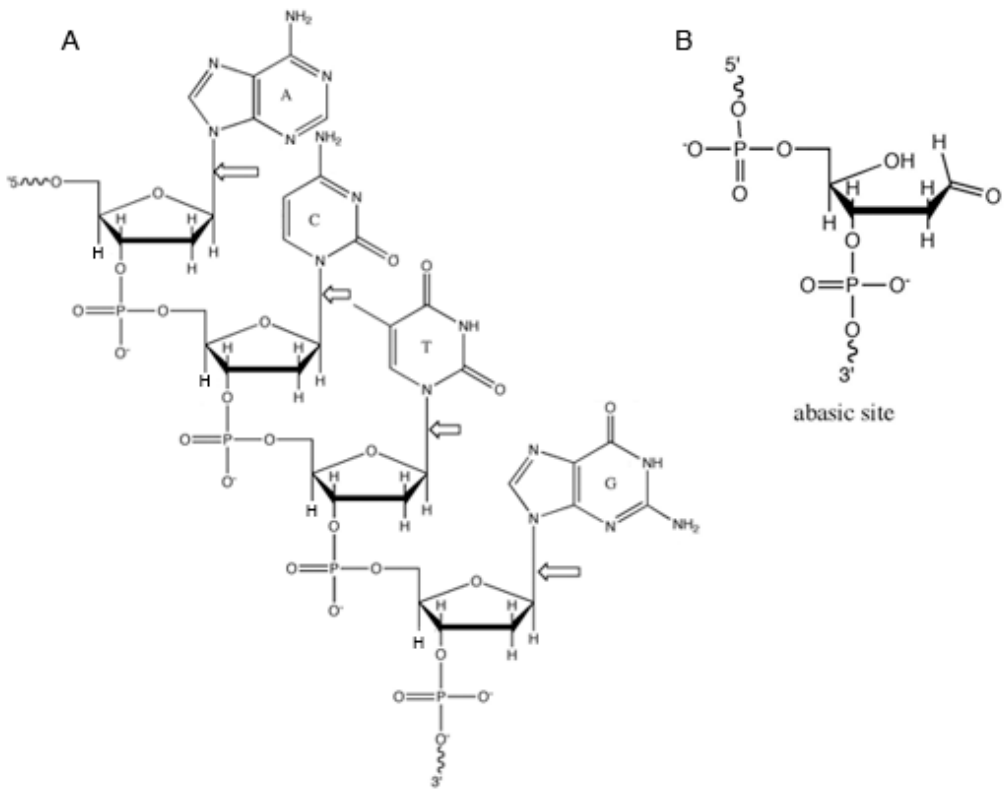

Figure 1.2. **Schematic of DNA bases and AP sites.** A) DNA is comprised of four bases, adenine (A), cytosine (C), thymine (T), and guanine (G). Each base is attached to the C1 carbon of a deoxyribonucleotide to form an *N*-glycosidic bond (arrows) that is targeted by the *N*-glycosylase proteins. B) The *N*-glycosidic bond is cleaved is cleaved to form a base-free AP site.

As depicted in **Figure 1.3**, the *N*-glycosylase proteins initiate BER when they recognize and cleave a specific subset of DNA base damage leaving an AP site. Monofunctional N-glycosylases cleave the damaged base from DNA leaving an AP site that is further processed by AP endonuclease (APEX1, also known as APE1) before gap filling and ligation to repair the initial site of damage [1, 23]. These glycosylases include the uracil DNA glycosylase (UDG) superfamily, which includes the UNG1/2, SMUG, TDG, and MBD4 proteins, and the methyl purine glycosylase, MPG [23]. To cleave the damaged base, monofunctional glycosylases employ a water molecule to attack the aromatic carbon of the base in order to cleave the N-glycosidic bond between the base and deoxyribose sugar [23, 26]. Thus, monofunctional glycosylases generate non-coding AP sites that lack instructional base pairing. AP sites can be mutagenic by single base pair substitutions, as DNA polymerases preferentially insert adenine into AP sites [27]. Furthermore, AP sites are toxic as they block replication and transcriptional machinery [28].

Bifunctional *N*-glycosylases contain an AP lyase activity that further cleaves the DNA phosphodiester backbone 3' of the AP site, resulting in a single strand break [23]. These glycosylases employ an amine moiety in the active site to excise the damaged base and generate a Schiff base intermediate for AP lyase cleavage of the DNA phosphodiester backbone [23, 26]. Glycosylases in this family comprise the MUTYH protein and

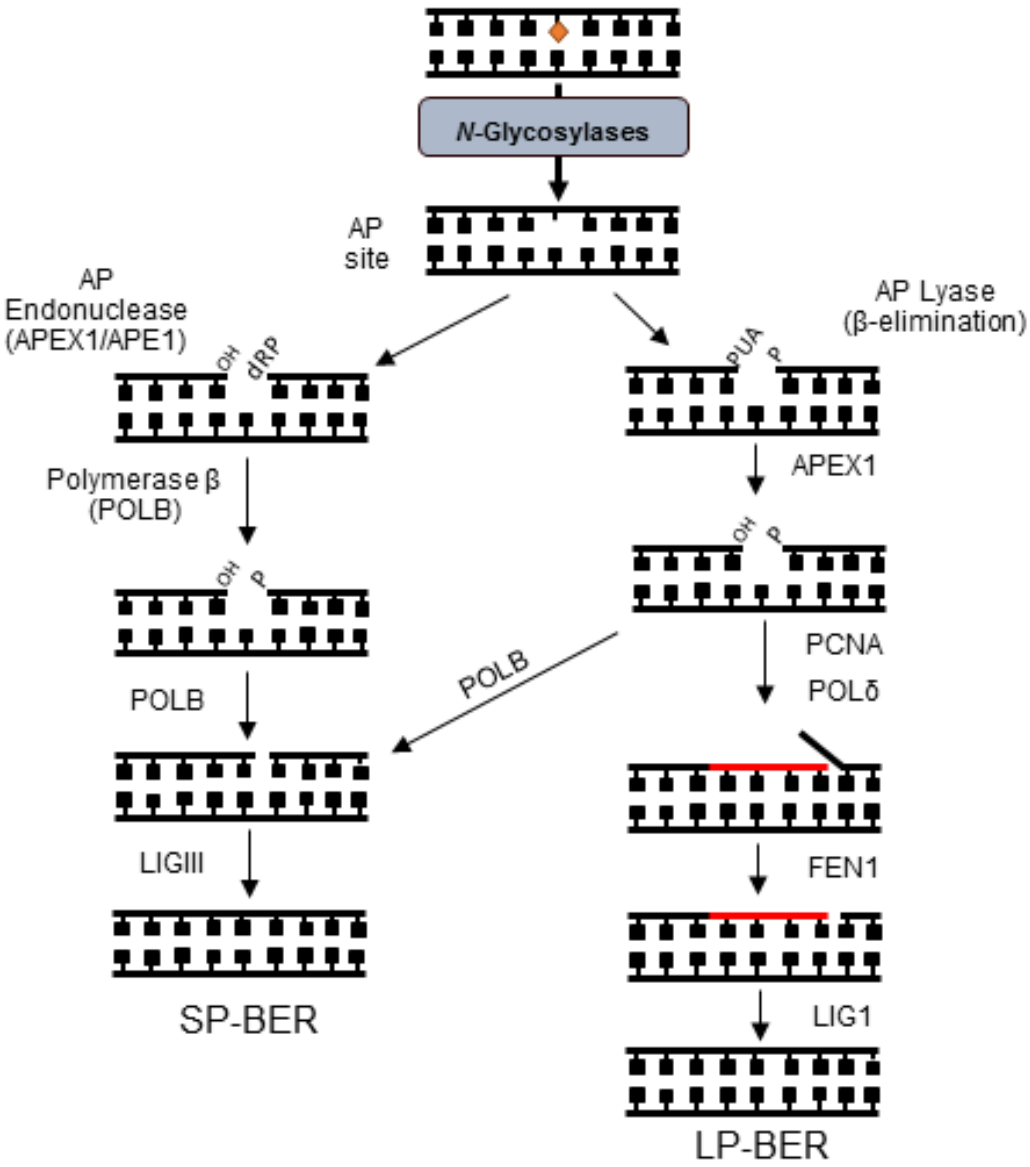

Figure 1.3 the base excision repair pathway. Base excision repair (BER) is initiated when an N-glycosylase recognizes a DNA substrate (diamond) and excises the damaged base to leave an AP site. In short patch BER (SP-BER), the AP lyase activity of a bifunctional glycosylase may nick the DNA phosphodiester backbone using a β-elimination reaction leaving a phosphate (P) and phospho-unsaturated aldehyde (PUA) blocking groups. AP endonuclease (APEX1) will either perform end-cleaning duties following AP lyase activity, or cleave the AP site following a monofunctional glycosylase resulting in a hydroxyl (OH) and a deoxyribose phosphate (dRP) group. DNA polymerase β adds an undamaged nucleotide and ligase III (LIG III) seals the remaining DNA nick. In long patch BER (LP-BER), following APEX1 end processing, PCNA and polymerase δ (POLδ) adds and displaces 2-12 nucleotides (red). FEN1 cleaves the DNA flap and ligase I (LIG I) seals the remaining DNA nick.

glycosylases that repair oxidative DNA damage including NTHL1, OGG1, NEIL1, NEIL2, and NEIL3 [23]. Thus, the result of the bifunctional glycosylases is a single strand break, which must be fully repaired. APEX1 then performs end-cleaning duties following glycosylase AP lyase activity [29]. Further processing by DNA polymerase β and subsequent ligation result in repair of the initial damage site [22].

Upon removal of a damaged base by a monofunctional DNA glycosylase, APEX1 will cleave the DNA phosphodiester backbone 5' to the abasic site to generate a 3'-hydroxyl group and a 5'-deoxyribose phosphate (dRP) [23, 29]. The dRP moiety cannot be directly utilized by DNA polymerases to complete repair; hence dRP is "cleaned up" by the dRP lyase activity of DNA polymerase β (POLB) to generate a 5'-phosphate group before gap filling ensues [23, 29]. Alternatively, APEX1 also follows the AP lyase activity of bifunctional *N*-glycosylases to process the remaining non-conventional 3'-phospho-α,β-unsaturated aldehyde (PUA) group [23]. APEX1 processes PUA to a 3'-hydroxyl group required for gap filling by DNA polymerases [23].

BER can deviate following end processing to engage short patch BER or long patch BER components [23, 30-33]. During short patch BER (SP-BER), a single nucleotide is replaced at the original site of damage [1, 23]. Following POLB processing of the dRP group, POLB will insert a single nucleotide, and ligation by DNA ligase III completes BER repair [1, 23]. Long patch BER (LP-BER) replaces anywhere from 2-12 nucleotides via strand displacement synthesis by DNA polymerase δ and PCNA, resulting in a flap intermediate [30-33]. Flap endonuclease 1 (FEN1) cleaves the displaced DNA so ligation

by DNA Ligase I can complete BER [23, 34]. The exact trigger to initiate LP-BER is unclear, but evidence suggests that the energy state of the cell through ATP levels contributes to the choice between SP-BER and LP-BER [1, 23, 35]. Sufficient ATP levels allow for DNA ligase III to immediately utilize ATP in the final ligation step of SP-BER [35]. During times of metabolic stress where ATP levels are decreased, LP-BER employs poly (ADP-ribose) polymerase (PARP), which is dependent on the NAD^+ cofactor instead of ATP [35, 36]. Poly (ADP-ribose) hydrolysis may be used to generate the ATP necessary for the final ligation step of LP-BER [37]. Thus, BER is a complex, multi-step process that requires proper regulation of each protein component and hand off to subsequent steps.

1.2.2 Consequences of BER dysregulation

BER glycosylases generate apurinic/apyrimidinic (AP) sites and/or strand breaks, which are themselves forms of DNA damage [38, 39]. Proper coordination and regulation of BER glycosylase proteins is important to ensure these intermediates do not accumulate. As APEX1 also generates single strand breaks during BER, the coordinated handoff from glycosylase to APEX1 to downstream BER proteins must be properly regulated to avoid accumulation of BER intermediates. Therefore, a balance of BER components is crucial to ensure rapid and efficient repair of base damage, while simultaneously ensuring that potentially deleterious BER intermediates do not accumulate (**Figure 1.4**).

Because the *N*-glycosylases initiate BER and generate repair intermediates, elucidating how these enzymes are regulated is paramount to understanding how BER maintains

stability of the genome. In fact, *N*-glycosylase activity appears to be affected through protein interactions of BER with various protein components of other DNA repair pathways, a topic extensively covered in section in 1.4. Examples include glycosylase turnover from DNA by interactions with components from nucleotide excision repair (NER) that are independent of NER activity [40-42]. For example, the interaction between NTHL1 and the NER protein, XPG, increases NTHL1 binding to its substrates for excision [40]. Other examples of crosstalk include the interaction of the NER protein, CSB, with the BER N-glycosylases OGG1 [43], NEIL1 [44], and NEIL2 [42]. Therefore, the crosstalk of BER repair proteins with other DNA repair pathways may play an important role in maintaining genome stability.

1.3 NTH endonuclease-like 1 (NTHL1)

Nth endonuclease-like 1 (NTHL1) is a highly conserved BER glycosylase that is annotated in species from bacteria to humans and is responsible for repair of oxidized pyrimidine bases (thymine and cytosine) that result from exposure to ROS [45]. cBioPortal cancer datasets indicate that NTHL1 is frequently amplified across many cancer types. However, the cellular consequences of NTHL1 amplification or protein overexpression remain unknown.

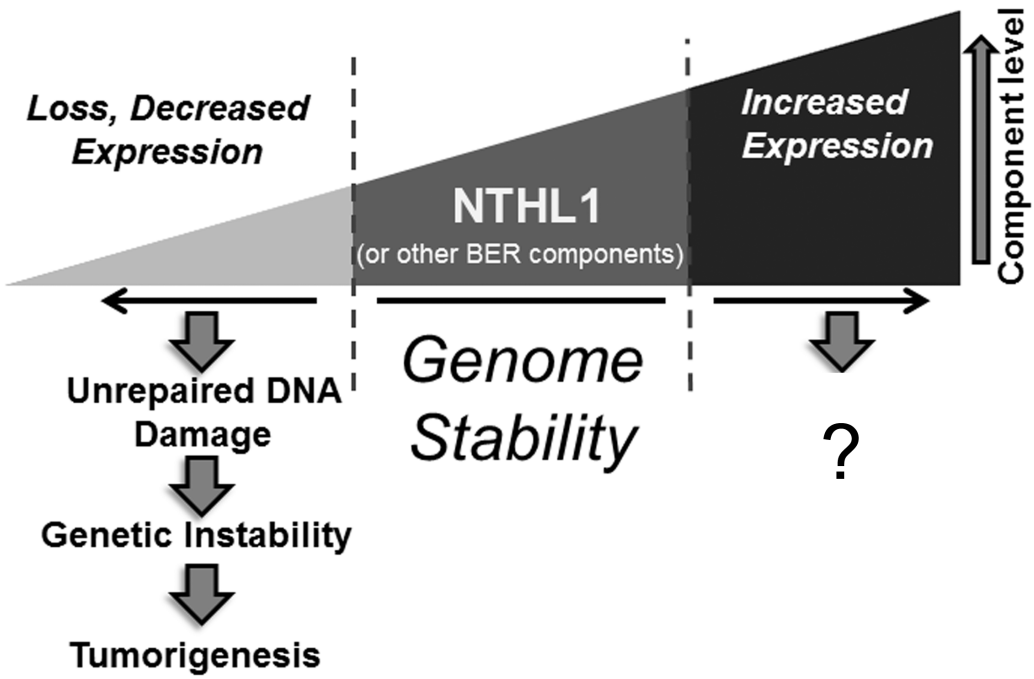

Figure 1.4. **A proper balance of base excision repair protein components is required for genome stability.** Dysregulation of BER through loss or decreased protein levels may contribute to tumorigenesis through the accumulation of unrepaired DNA damage, as depicted in [46, 47]. Unrepaired DNA damage may lead to the mutations necessary for cancer development. The contribution of BER protein overexpression to the tumorigenic process is not well understood for the BER *N*-glycosylases.

1.3.1 *NTHL1* gene organization

The *NTHL1* gene is found on the short arm of chromosome 16 at position 13.3 (16p13.3) between the tuberous sclerosis 2 gene (*TSC2*) and the gene for regulatory factor 2 of the Na^+/H^+ solute carrier family isoform A3 (SLC9A3R2) [45]. Interestingly, the promoters for *NTHL1* and the tumor suppressor, *TSC2*, are located in a 5'-to-5' orientation within a 357 bp span of each other [45]. These promoters putatively overlap or are bidirectional and present another layer of regulatory complexity for studying the *NTHL1* gene [45]. Loss or mutation of *TSC2* results in the tuberous sclerosis syndrome, and the large deletions of *TSC2* responsible for the syndrome frequently encompass the *NTHL1* gene in a subset of patients [48], demonstrating the impact these genes may exert on each other due to close proximity of their promoters.

1.3.2 *NTHL1* transcript isoforms

Three mRNA isoforms for *NTHL1* are annotated in the NCBI database (**Figure 1.5**). Isoform 1 of *NTHL1* encodes for a transcript that spans six exons and 5 introns (NP_002519.1) resulting in a 312 amino acid protein, which is the focus of our study. Two other isoforms of *NTHL1* have been annotated in the NCBI database that varies in their 5' untranslated region (UTR) [49]. Isoform 2 encodes for a transcript that lacks an alternate exon beginning in the 5' coding region of exon 1 and internally lacks exon 3, resulting in a shorter NTHL1 protein (NP_001305122.1) [49]. Isoform 3 encodes a transcript that initiates transcription at an alternative start site encompassing exons 3-6, resulting in a truncated NTHL1 protein (NP_001305123.1) [49]. However, initial characterization of NTHL1 mRNA levels across various tissue types revealed that in liver

tissue, the predominant isoform of NTHL1 is significantly shorter than predicted [45, 50]. This result suggests that in liver tissue, the predominant NTHL1 transcript is isoform 3. A separate study annotated an NTHL1 transcript of intermediate size in heart, spleen, and thymus tissue using Northern blotting [50]. This result suggests the presence of NTHL1 transcript that corresponds to isoform 2. To our knowledge, both isoforms 2 and 3 have yet to be studied in a cellular context, and future studies could focus on various functions of each NTHL1 isoform in a tissue dependent context.

1.3.3 NTHL1 protein organization

NTHL1 protein translated from isoform 1 contains a mitochondrial targeting sequence (MTS) distally on the N terminus, with a closely adjoined bipartite nuclear localization sequence (NLS) at amino acid positions 27-36 (NLS 1) and 45-52 (NLS 2) (**Figure 1.6**) [51]. A third putative NLS (NLS 3) is found at amino acid positions 115-123, but was shown to be dispensable for NTHL1 nuclear sorting of isoform 1 [51]. However, as NTHL1 isoform 3 was recently identified to encompass exons 3-6, whether NLS 3 contributes to nuclear localization of this shorter NTHL1 protein remains unknown. NTHL1 also contains a helix-hairpin-helix (HhH) motif that encompasses the enzyme active site and an iron-sulfur cluster (4Fe-4S) on the C terminus tail (**Figure 1.6**) [50-52]. Excepting the N terminus and the sub-cellular localization signals, NTHL1 is highly conserved across eukaryotes as well as to the bacterial nth endonuclease III ortholog [50]. While no crystal structure for NTHL1 has been reported, the yeast homolog to NTHL1, Ntg1, was homology modeled against the bacterial nth glycosylase, showing similarity in the catalytic core between both proteins with an unstructured N terminal tail [53].

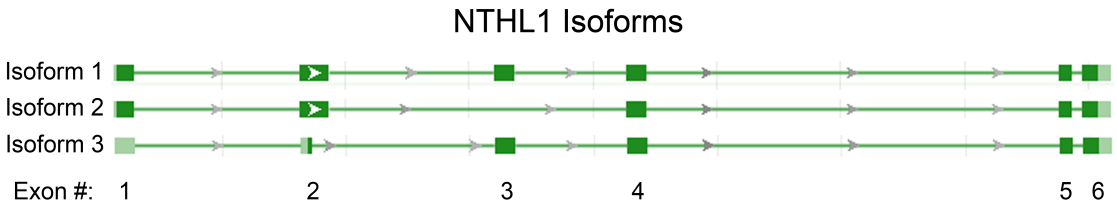

Figure 1.5. NTHL1 isoforms. Three isoforms of *NTHL1* exist in humans. Each dark green box annotates an exon of NTHL1, with exons 1-6 labeled. Green lines connecting exons denote introns, and the arrows indicate the direction of gene transcription.

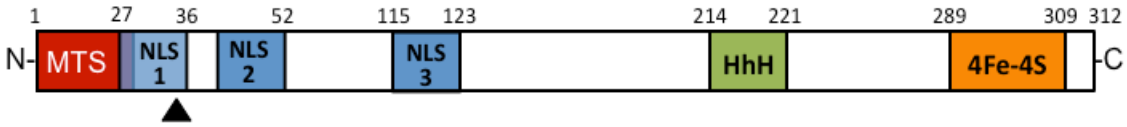

Figure 1.6. **Schematic of NTHL1 protein organization**. NTHL1 is amplified across multiple cancer types. NTHL1 is a 312 amino protein with an N terminus mitochondrial targeting sequence (MTS) at amino acid positions 1-30. A bipartite nuclear localization sequence (NLS) is located at amino acids 27-36, and 45-52. NLS 1 and the MTS sequences overlap with a predicted MTS cleavage site at amino acid 35 (black triangle). A third putative NLS is found at positions 115-123. The catalytic core is contained in the helix-hairpin-helix (HhH) motif at amino acids 214-221. The C terminus tail contains an iron-sulfur core (4Fe-4S) located at positions 289-309.

1.3.4 NTHL1 biochemistry

Biochemically, NTHL1 repairs a range of oxidative DNA damage to primarily pyrimidine bases, including thymine glycol (Tg), 5'-hydroxycytosine, dihyrdouracil, urea, and the ring opened form of the purine base, guanine [54, 55]. Furthermore, NTHL1 is classified as a bifunctional glycosylase and contains both glycosylase and AP lyase activity for repair of DNA substrates [52]. NTHL1 excises these base damages using an amine moiety of a lysine at amino acid position 220 in the catalytic core of the HhH domain [26, 52]. This lysine is highly conserved across species [50], making studies in model systems ideal for understanding the basic biochemistry of NTHL1 and its orthologs. The amine group is used as a nucleophile to cleave the deoxyribose base and form a Schiff base intermediate with DNA [26, 52]. Resolution of this intermediate is completed by the AP lyase activity of NTHL1 where the phosphodiester backbone is cleaved by β-elimination [26, 52]. Mutagenesis of lysine 220 to glutamine abolishes both glycosylase and AP lyase activity of NTHL1 [52]. Furthermore, this catalytically dead NTHL1 protein cannot bind a radiolabeled oligonucleotide containing an NTHL1 substrate; thus, subsequent DNA repair cannot occur with this mutation [52].

1.3.5 NTHL1 regulation

Further biochemical analysis of NTHL1 showed that the N terminus tail is autoinhibitory for NTHL1 release from the final, cleaved DNA product [56]. In fact, upon progressive truncation of the N terminus, NTHL1 activity increased three- to four-fold in kinetic activity assays [56]. This result suggests that release of the nicked DNA is a controlled process and that coordinated handoff of the BER intermediate to APE1 is necessary, as

BER intermediates are types of DNA damage. In support of this idea, NTHL1 release from the cleaved product was determined to be the rate-limiting step of the enzyme [56].

To overcome this inhibitory effect, homodimerization of the NTHL1 N-terminus tail prompts release of the cleaved product [57]. NTHL1 dimerization and turnover occurs in a protein concentration dependent manner *in vitro*, and extension of this observation into HeLa cells confirmed dimerization in a tissue culture model [57]. This study provided insight into how NTHL1 regulates its own enzymatic turnover, and adds to the repertoire of NTHL1 regulation that includes regulation by the XPG, YB-1, and APEX1 proteins [40, 57-59].

NTHL1 regulation by XPG is extensively covered in section 1.4.1. YB-1 is a DNA damage inducible transcription factor that stimulates NTHL1 DNA glycosylase and AP lyase activity, as shown through kinetic cleavage assays *in vitro* [58]. YB-1 does so though protein-protein interaction with NTHL1 and pushing the kinetics of the reaction towards an enzyme-DNA-Schiff base intermediate, thus increasing the rate of β-elimination to produce a DNA nick [58]. APEX1 regulates NTHL1 enzymatic processing in a substrate specific manner depending on the base pairing of the DNA damage (i.e. Tg:A or Tg:G) [59]. The biochemistry, kinetics, and mechanism of action vary depending on the damage substrate, and a detailed explanation for how APEX1 influences NTHL1 activity can be found in [59].

Another level of NTHL1 regulation is found in the context of sub-cellular localization, as

NTHL1 may be localized to the nuclear, mitochondrial, or cytoplasmic compartments [51, 52, 60]. Studies addressing NTHL1 localization in the absence of DNA damaging agents are conflicting and may be tissue dependent. For example, Ikeda et al showed that NTHL1 is localized to the nucleus and cytoplasm in HeLa cells (cervical cancer origin) [52]. In a separate study using HeLa cells, Ikeda et al looked at NTHL1 localization and determined that NTHL1 localizes to the nucleus and mitochondria, but did not find cytoplasmic localization [51]. Yet others examined NTHL1-Flag localization in COS-7 cells (African green monkey kidney cells) and determined that NTHL1 is present in the nucleus and mitochondria [60]. Luna et al showed that NTHL1-GFP is exclusively sorted to the nucleus in HeLa cells [55]. As HeLa is a cancer cell line, whether an acquired mutation in the NTHL1 localization sequence(s) accounts for the differences in localization is undetermined. Additionally, increased oxidative stress burden in certain cells could direct the localization of NTHL1 to either the nucleus or mitochondria, and account for the differences seen between studies.

Interestingly, sub-cellular localization of the NTHL1 orthologs may also be species dependent. For instance, the mouse NTH1 (mNTH1) is highly conserved in the core domain compared to human NTHL1 at 89% identity [55]. However, in the less conserved N-terminus tail, mNTH1 and NTHL1 only share 59% identity [55]. Differences in these amino acids could explain why mNTH1 is predominately localized to mitochondria, in contrast to human NTHL1 [55]. Mutation of the mNTH1 MTS sequence results in a switch to nuclear localization of mNTH1 [55], suggesting sub-cellular sorting preferences between mNTH1 and NHTL1. This is a noteworthy finding because if the mNTH1

enzyme does not behave the same as the human NTHL1 enzyme, knock out of *mNTH1* in mouse models may not recapitulate the NTHL1 biology found in humans. A knockout mouse model for mNTH1 exists, and under conditions where the animals were not exposed to genotoxic stress, little spontaneous tumor formation is observed [61]. In order for spontaneous tumor formation to occur, another oxidative DNA damage glycosylase, *Neil1*, must also be knocked out in conjunction with *mNTH1* [61]. However, tumor formation in both the single and double knockout mice for mNTH1 was not measured in response to DNA damaging agents. Given that human *NTHL1* loss contributes to a colon cancer predisposition syndrome [46, 47], mouse models, at least in the case for NTHL1, may not accurately reflect human tumorigenesis processes in each instance.

Alternatively, localization in response to DNA damaging agents is characterized for the NTHL1 homologs, Ntg1 and Ntg2 in the budding yeast *Saccharomyces cerevisiae.* NTHL1 is closely related to Ntg2, as Ntg2 contains an iron-sulfur cluster [62]. However, Ntg2 is exclusively localized to the nucleus [63]. Ntg1 does not contain an iron-sulfur cluster, but has an N terminus tail that contains both MTS and NLS signals that are analogously found in NTHL1 [62]. Thus, both Ntg1 and NTHL1 could localize to either the nucleus or mitochondria in response to DNA damaging agents.

For these studies, hydrogen peroxide (H_2O_2) was used a general ROS DNA damaging agent, while antimycin A was utilized to preferentially cause an increase of ROS in the mitochondria [63]. Upon exposure to H_2O_2, Ntg1-GFP was found to predominately localize to the nucleus of yeast, and analogously Ntg1-GFP predominately localizes to

the mitochondria following exposure to antimycin A plus H_2O_2 [63]. This process of preferentially responding to DNA damage contained in a sub-cellular compartment was termed dynamic localization. In the absence of dynamic localization in yeast, mutation rates in the nucleus or mitochondria increased in cells that contained mutated MTS or NLS sequences. Thus, transfer of BER glycosylases to the correct organelle upon exposure to DNA damaging agents is a critical component of the DNA damage response. Whether human BER glycosylases respond to DNA damage utilizing dynamic localization remains unknown.

1.3.6 NTHL1 in cancer

Human NTHL1 was found in both the nucleus and mitochondria in HeLa cells in support of the idea that NTHL1 could undergo dynamic localization [52]. Evidence for the potential importance of NTHL1 localization in primary cancer samples was later found in gastric and colon cancer [64, 65]. Primary gastric cancer samples showed that in 24% of examined cases, NTHL1 was exclusively mislocalized from the nucleus to the cytoplasm [64]. At least for this set of gastric cancer, NTHL1 mislocalization did not have an association with tumor stage [64]. As NTHL1 repairs oxidative DNA damage in the nucleus, this result suggests that repair of NTHL1 substrates may not occur in the nucleus, potentially leading to the accumulation of nuclear mutations contributing to tumor development. The mechanism of NTHL1 mislocalization in gastric cancer has not been further studied.

A separate study of NTHL1 localization in colon cancer reported that in 35% of cases

studied, NTHL1 was mislocalized from the nucleus to the cytoplasm [65]. Unlike in gastric cancer, NTHL1 mislocalization in colon cancer does associate with lymph node metastasis (p=0.001), tumor grade (p=0.005), and poor disease free survival (p=0.04) [65]. Again, no mechanism for NTHL1 mislocalization was further studied. These studies underscore the need to understand NTHL1 regulation and dysregulation in cancer to elucidate how BER glycosylases may contribute to the tumorigenic process.

Recent studies on the protein expression levels of NTHL1 demonstrate that a recessive, homozygous loss-of-function germline mutation in *NTHL1* contributes to a novel colon cancer predisposition syndrome [46, 47]. This *NTHL1* variant was identified by Sanger sequencing in three non-related families, and revealed that a germline variant at nucleotide position 268 of the *NTHL1* gene was responsible for the disorder [46, 47]. Mutation at nucleotide c.268C >T (cytosine to thymine) results in a nonsense amino acid substitution encoding for glutamine at position 90 (p.Gln90*). Mechanistically, this amino acid change results in NTHL1 protein decay, thus resulting in the absence of cellular NTHL1 protein [46, 47]. These patients also appear to be prone to multiple cancer types beyond colon cancer. For instance, one male patient presented with colon, pancreatic, duodenal, and biliary tract cancers between 47 to 52 years of age [47]. Females in these cohorts were also diagnosed with either endometrial cancer or endometrial hyperplasia [47]. Fortunately, the prevalence of homozygosity of the p.Gln90* allele is rare, with calculations showing that the occurrence is approximately 1 in every 75,076 individuals after examination of over 63,000 exomes [47]. However, the pGln90* allele was more prevalent in European populations, especially in persons with

Dutch ancestry. This study emphasized that loss of NTHL1 protein in humans predisposes to tumor formation, in contrast to the previously mentioned mouse mNTH1 knockout models. However, patients presented with cancers beginning in midlife, and the longevity discrepancies between humans and mice could also account for why loss of NTHL1 in humans results in cancer. Humans are also exposed to DNA damaging agents throughout life and these additional exposures may contribute to the NTHL1 cancer syndrome. Taken together, these findings further support the notion of NTHL1 as a tumor suppressor.

We analyzed *NTHL1* expression in cBioPortal cancer datasets for loss of *NTHL1* in other cancer types. Contrary to the above findings, our analysis revealed that *NTHL1* is frequently amplified or mRNA upregulated across several cancer types including breast, pancreas, prostate, and lung cancers (**Figure 1.7**). Whether *NTHL1* amplification or mRNA upregulation results in increased NTHL1 protein levels, and whether an increase in NTHL1 protein levels can directly contribute to the tumorigenic process has not been examined. Classic examples of gene amplification in cancer include *MYCN*, *EGFR*, and *ERBB2* across various cancer types [66]. As gene amplification is thought to predict areas of the genome that promote tumor formation [66], the role that *NTHL1* amplification may contribute to tumorigenesis was investigated. However, overexpression of other BER glycosylases has cellular impacts. For example, overexpression of the MPG glycosylase causes temozolomide resistance in gliomas, and patients with these tumors have poorer outcomes [67, 68]. Overexpression of OGG1 results in a greater sensitivity to ionizing radiation in a cell culture model [69], and SMUG1 overexpression is also correlated with

poor disease-free survival in gastric cancer patients [70]. Thus, an increase in NTHL1 protein could also have implications for tumorigenesis.

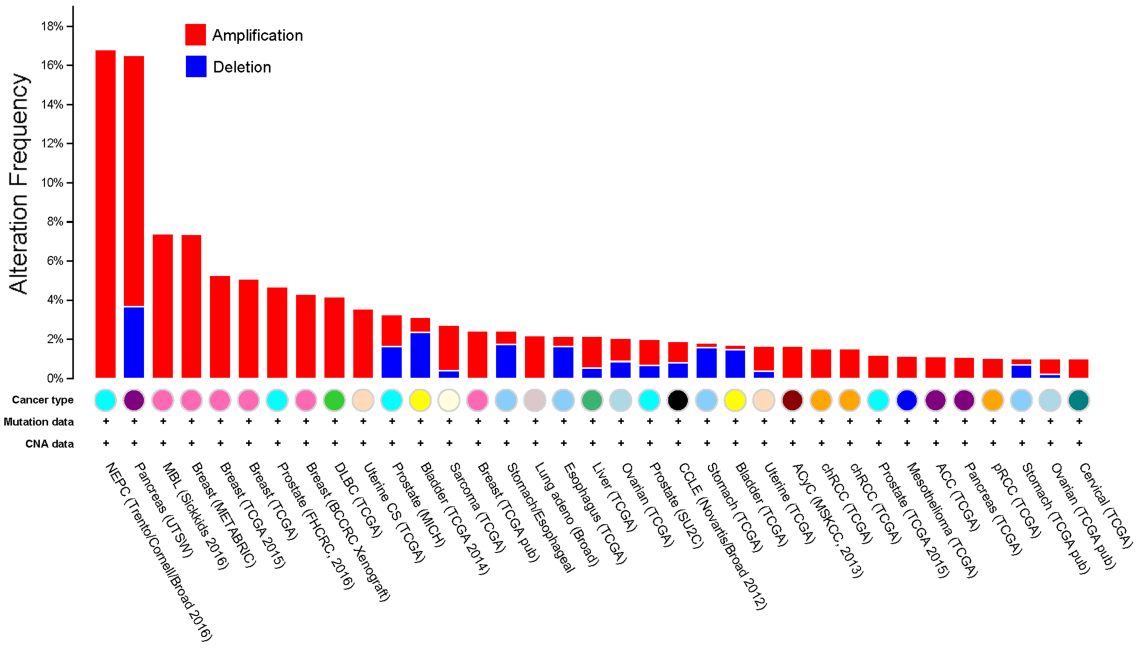

Figure 1.7. **NTHL1 is amplified across multiple cancer types**. Cancer datasets in the cBioPortal database show that *NTHL1* copy number is frequently amplified (red). Deletion of the *NTHL1* gene (blue) is found in certain cancer types such has colon cancer. The alteration frequency is displayed on the y-axis, and the cancer type is shown across the x-axis. Colored circle denote a specific cancer type. For instance, a pink circle is breast cancer, light blue is prostate cancer, purple is pancreatic cancer, and grey is lung cancer.

1.4 DNA Repair Pathway Crosstalk

Section 1.4 is adapted from the review article “BERing the burden of damage: Pathway crosstalk and posttranslational modification of base excision repair proteins regulate DNA damage management," authored by Kristin Limpose, Anita H. Corbett, and Paul W. Doetsch. DNA Repair, Accepted April 2017 for publication in August 2017.

KL and PWD generated the article concept. KL, AHC, and PWD all contributed to writing and editing the article for publication.

Originally, each DNA repair pathway was analyzed in isolation to define repair of a specific subtype of DNA damage. For instance, base excision repair (BER) handles non-bulky DNA base damage, nucleotide excision repair (NER) manages bulky lesions, and homologous recombination (HR) repairs double strand breaks in S phase [71]. As each DNA repair pathway was characterized beyond the individual biochemical steps, it became apparent that coordination between DNA repair pathways is essential for proper cellular responses to DNA damage. We refer to such coordination as pathway crosstalk. In this context, pathway crosstalk occurs when components of one, biochemically distinct DNA repair pathway influence the repair of a substrate that is corrected by a different DNA repair pathway. For example, components of NER are indispensable for efficient repair of BER substrates through interactions with several *N*-glycosylases that initiate BER [40, 42, 44]. We focus on pathway crosstalk events primarily mediated through protein-protein interactions that could affect BER activity and pathway crosstalk in response to DNA damage through multiple mechanisms. We describe several classical, as well as recently reported examples of pathway crosstalk, with an emphasis on how BER components are regulated in human cells.

Of note, a majority of BER crosstalk with other DNA repair pathways takes place at the initiating steps of BER. As BER glycosylases generate apurinic/apyrimidinic (AP) sites and/or strand breaks, which are themselves forms of DNA damage [38, 39], proper coordination and regulation of BER glycosylase proteins is important to ensure these intermediates do not accumulate. APEX1 (also known as APE1) also generates single

strand breaks during BER and the coordinated handoff from APEX1 to downstream BER proteins must be properly regulated to avoid accumulation of BER intermediate.

1.4.1 Nucleotide Excision Repair Crosstalk with BER Components

The canonical function of nucleotide excision repair (NER) is to eliminate bulky DNA damage, which can arise from exposure to UV radiation or certain chemical agents [71]. These lesions include UV-induced cyclobutane pyrimidine dimers, pyrimidine-pyrimidone photoproducts, and bulky chemical adducts [25, 72]. NER also participates as a backup mechanism to base excision repair (BER) for the repair of certain oxidative induced DNA base damage [73]. Thus, the interplay between multiple NER and BER components is critical to ensure efficient BER processing of base damage. Several BER glycosylase-NER protein interactions have been characterized that impact the function of key BER glycosylase enzymes in processing of their respective DNA damage substrates. Several specific examples that illustrate this interplay are described here.

TDG Glycosylase

Well documented evidence that NER components can influence BER glycosylase activity comes from the analysis of the interaction between the BER thymine DNA glycosylase, TDG, and the NER XPC protein (**Figure 1.8**) [41, 74]. In their respective pathways, TDG recognizes G:T mismatches in DNA and excises the mismatched thymine [75]. The XPC protein is involved in global genome NER (GG-NER) [76]. TDG is strongly product inhibited by the AP site that is produced in DNA following thymine cleavage and APEX1 helps displace TDG from these AP sites [77]. Previous work identified a physical

interaction between TDG and XPC [41]. To assess whether XPC is an additional factor that contributes to displacement of TDG from AP sites, APEX1 and XPC were simultaneously added to DNA-bound TDG. Individually, APEX1 or XPC stimulated moderate TDG release from DNA. When present together, APEX1 and XPC resulted in a 6-fold increase in TDG release from DNA compared to control reactions [41]. These findings demonstrated that XPC is an additional component that triggers TDG release from DNA product. However, an exact mechanism for how XPC stimulates TDG release has yet to be defined. Furthermore, as XPC increased the ability of APEX1 to aid in TDG turnover, future studies are required to determine if XPC binds to and/or influences APEX1 activity.

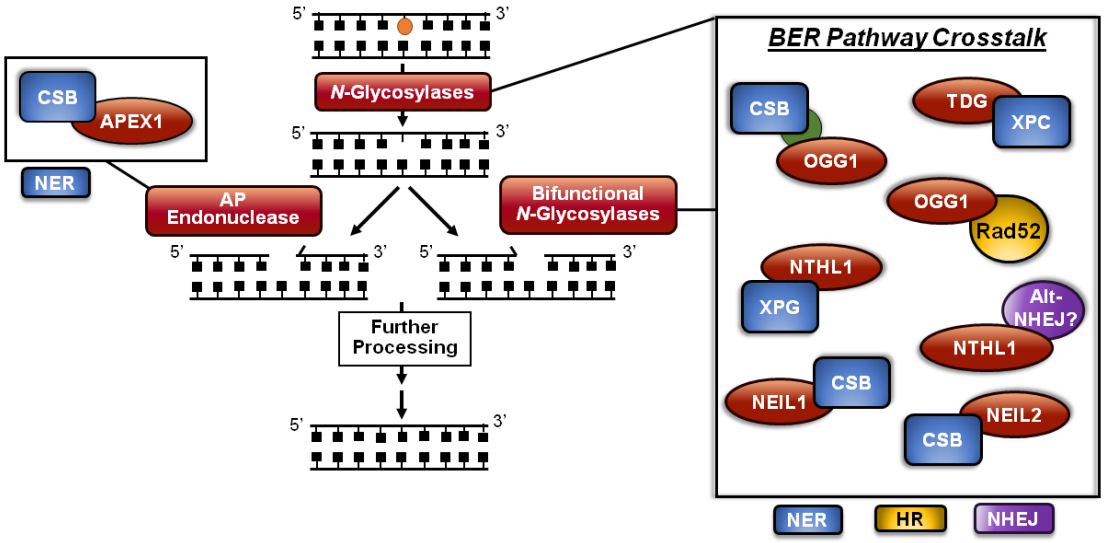

Figure 1.8. **Key human *N*-glycosylases and APEX1 interactions with components of other DNA repair pathways enhance BER activity**. In the initiating steps of BER, a damage substrate is detected, and an N-glycosylase cleaves the damaged base, leaving an apurinic/apyrimidinic (AP) site. Bifunctional glycosylases cleave the phosphodiester backbone and create a single-strand break, while monofunctional glycosylases require APEX1-mediated cleavage of the phosphodiester backbone. Further processing results in repair of the initial damage site. Functional interactions with BER proteins in this review are depicted, with an emphasis on the initiating steps of BER. Pathway crosstalk of BER proteins (red) at the initiating steps of BER includes interactions with components of the NER (blue), HR (yellow), and alt-NHEJ (purple) pathways. We highlight recent advances that provide insight into BER functions, and the interactions displayed are discussed in this review.

NER components interact with BER substrates

Another example of BER and NER interplay includes two NER proteins, XPC and CSB. Both XPC and CSB localize to sites of oxidatively-induced DNA damage generated by laser (405 nm) excitation of a photosensitizer [78]. The primary product of this reaction is the BER substrate 8-oxoguanine (8-oxoG) [78]. Fluorescently tagged XPC and CSB were employed to track the localization and kinetics for both proteins within the nucleus. Upon DNA damage, XPC localized to sites of DNA damage exclusively in the nucleoplasm while CSB localized to sites of damage in both the nucleolus and the nucleoplasm [78]. As the nucleolus is a site of high transcriptional activity due to ribosomal DNA [79], these results are in line with previous data assigning XPC to global genome NER and CSB to transcription coupled repair [76]. In contrast, downstream NER components such as the XPA and XPB proteins were not recruited to these sites of oxidative base damage [78]. This result indicates that the recruitment of XPC and CSB is independent of their respective NER functions and supports a role for XPC and CSB in influencing BER-mediated repair of oxidatively-induced DNA damage.

The NER protein, CSB, also influences binding and excision of the oxidative damage 8-oxoG by the BER glycosylase, OGG1 [41, 80-82]. Despite a functional link to 8-oxoG repair, no direct interaction between the OGG1 and CSB proteins has been detected, suggesting that these proteins could function as part of a protein complex to ensure efficient BER function (**Figure 1.8**) [43]. By analyzing the kinetics of protein recruitment to DNA damage, Menoni et al. concluded that CSB is recruited to DNA damage prior to OGG1 [78]. Consistent with this model, there was no change detected in either XPC or

CSB recruitment to damage in cells deficient for OGG1. Thus, OGG1 is not required for recruitment of CSB to sites of oxidative DNA damage [78]. XPC was recently described as a general DNA damage sensor independent of NER [83]. This role for XPC is supported both by the XPC link to TDG glycosylase and nucleoplasm localization of XPC to sites of DNA damage independent of other NER components [78]. How XPC may generally influence other BER glycosylases as a sensor for other BER substrates is unknown and will require further study.

NEIL Glycosylases

Another class of BER glycosylases that is modulated by CSB are the NEIL1 and NEIL2 glycosylases (**Figure 1.8**) [42, 44, 84]. While the NEIL glycosylases have substrate specificity that overlaps with other BER *N*-glycosylases, they are unique in their ability to excise oxidative DNA damage from single-stranded DNA that mimics a transcription bubble [85]. NEIL1 substrates include the 2,6-diamino-4-hydroxy-5-formamidopyrimidine (FapyG) and 4,6-diamino-5-formamidopyrimidine (FapyA) open ring base damages [44]. As both FapyG and FapyA share an intermediate structure with 8-oxoguanine [86], and as CSB impacts OGG1-mediated repair of 8-oxoguanine [43], this led to the hypothesis that CSB could impact NEIL1 activity [44]. To address this question, levels of FapyG and FapyA damage were analyzed in the brain, liver, and kidneys of $CSB^{-/-}$ mice revealing that FapyA levels are increased in all three tissues relative to control mice. FapyG damage was also elevated in the brain and kidneys compared to control mouse tissues, providing evidence that CSB is required for efficient repair of NEIL1 substrates. Further analysis using NEIL1 *in vitro* incision assays

revealed that CSB increases NEIL1-mediated incision activity up to 4-fold for FapyG, up to 2.5-fold for FapyA, and also stimulates NEIL1 AP lyase activity. The stimulation of NEIL1 by CSB was mapped to a region within the N-terminal domain of CSB (amino acids 2-341) and was independent of CSB ATPase activity. These results suggest that CSB does not mediate chromatin remodeling during repair of NEIL1 substrates. Immunoprecipitation experiments from HeLa lysates demonstrated that NEIL1 and CSB are present in the same protein complex although whether a direct interaction occurs between NEIL1 and CSB remains to be determined [44]. An interesting question is whether NEIL1 has a role in transcription-coupled repair of oxidatively-induced DNA base damage in light of the interaction with CSB and the ability of NEIL1 to initiate repair of single-stranded DNA.

Further work revealed that CSB also stimulates NEIL2 activity [42, 84]. Immunoprecipitation experiments from HeLa cells revealed that CSB and NEIL2 are present in the same protein complex and that the protein-protein interaction is increased following exposure to the oxidizing agent menadione. Further analysis revealed a direct protein-protein interaction between CSB and NEIL2 [42]. Incision assays of a NEIL2 substrate revealed that CSB stimulates NEIL2-mediated incision of FapyA up to 4-fold in duplex DNA and up to 3-fold for 5-hydroxyuracil present in a bubble DNA structure. However, CSB did not affect NEIL2 binding to DNA damage, suggesting that CSB may play a role in NEIL2 release from the final DNA product. Collectively, these results demonstrate that repair of BER substrate from single-stranded DNA can be coordinated through interactions between BER and NER components.

NTHL1 Glycosylase

A crucial BER glycosylase, NTHL1, repairs a large subset of oxidized DNA bases including dihydrouracil and the replication and transcription blocking base damage, thymine glycol (Tg) [40]. However, *in vitro* reconstitution experiments showed that purified NTHL1 has poor incision activity on an oligonucleotide containing an NTHL1 damage substrate [40]. Based on findings that NER-deficient patient cells, which lack the XPG protein, also showed poor excision of Tg [87], purified XPG was added into the reconstituted NTHL1 BER *in vitro* system. Surprisingly, the addition of XPG stimulated NTHL1 incision and release of DNA product for both Tg and dihydrouracil substrates. The addition of other NER components, XPA or XPC, had no such stimulatory effect on NTHL1 activity. To address whether XPG endonuclease activity is required to achieve this stimulation of NTHL1, two XPG protein variants (E791A and A792V) that have no XPG nuclease activity were employed [40]. Both variants could stimulate NTHL1-mediated base excision as effectively as wild type XPG, indicating that the XPG-dependent stimulatory effect on NTHL1 is independent of XPG nuclease function. Overall, this study demonstrates that XPG plays a critical role in NTHL1-mediated base excision of damage and release from DNA (**Figure 1.8**). Whether NTHL1 protein has an impact on XPG-mediated NER functions is not known.

AP Endonuclease (APEX1)

Platinum-based chemotherapeutics are used in the clinic as an effective treatment for multiple cancer types [88-92]. A common platinum therapeutic is cisplatin, which causes intra- and interstrand crosslinks primarily between guanine bases [92, 93]. However,

cisplatin has multiple negative side effects including peripheral neuropathy [94, 95], nephrotoxicity [92], and an increase in cellular reactive oxygen species (ROS) [16, 96]. NER is the main repair pathway to handle cisplatin adducts [97, 98] while BER initiates repair of the ROS-induced base damage [98].

Previous studies showed that increased protein levels of the BER protein, APEX1 (also known as APE1), protect cells against cisplatin toxicity [96, 99]. To assess the role of APEX1 in repair of cisplatin adducts, APEX1 levels were modulated in a neuronal tissue culture model exposed to cisplatin [95]. In this study, knockdown of APEX1 caused an increase in the level of unrepaired cisplatin adducts. Furthermore, cisplatin adduct repair was dependent on APEX1 endonuclease activity, and this repair activity was separate from APEX1 redox functions that are critical for transcriptional regulation [96]. Interestingly, when APEX1 was lost, an increase in the level of the NER protein, XPA, was detected. Whether this increase in XPA protein resulted from regulation at the protein or RNA level is unknown, and the exact mechanism of how APEX1 is involved in the removal of cisplatin adducts remains unclear. Thus, the level of cisplatin-DNA adducts, which are repaired by NER, increased in the absence of the BER protein, APEX1. Whether APEX1 involvement in cisplatin adduct repair is dependent on global genome NER (GG-NER) or transcription-coupled (TC-NER) NER is not known.

In support of a model where APEX1 plays a role in TC-NER of cisplatin adducts, the APEX1 and CSB proteins directly interact (**Figure 1.8**) [100]. CSB protein has been implicated in altering DNA conformation as well as chromatin remodeling during NER

[100, 101]. $CSB^{-/-}$ cells display hypersensitivity to reactive oxygen species (ROS)-generating agents, supporting a role for CSB in BER-mediated processing of oxidative DNA damage [102-104]. In fact, CSB also interacts with other BER proteins, including PARP1 and FEN1 [100], in addition to APEX1. *In vitro*, APEX1 endonuclease activity is increased up to 4-fold on duplex DNA and up to 6-fold on single-stranded DNA by the interaction with CSB [100]. The larger stimulation for single-stranded DNA indicates that the CSB/APEX1 interaction might have a greater impact on damage present in transcriptionally active DNA compared to double-stranded DNA. Addition of ATP was not needed for the stimulation of APEX1 endonuclease activity by CSB, suggesting that increased APEX1 endonuclease activity is not due to CSB-mediated chromatin remodeling. While these results show that APEX1 has a role in repair of cisplatin adducts [95], whether APEX1 has a reciprocal impact on CSB-mediated TC-NER activity is currently unknown.

1.4.2 Non-Homologous End Joining and Homologous Recombination Crosstalk with BER Components

Double strand breaks (DSBs) are the most deleterious class of DNA damage [105], and cells have evolved multiple repair pathways to repair this damage. Homologous recombination (HR) repairs DSBs in the S and G_2 phases of the cell cycle when a homologous sister chromatid is present [106]. Non-homologous end joining (NHEJ), while active in all phases of the cell cycle, primarily repairs DSBs in the G_0 and G_1 phases of the cell cycle. Importantly, BER can generate DSBs as a result of excision of closely opposed base damages or from single strand break intermediates if these

intermediates are encountered by the replication or transcription machinery [1]. To avoid the accumulation of deleterious damage intermediates, BER must therefore be efficiently and precisely regulated to initiate and complete repair. In fact, the BER intermediates generated by the alkyladenine glycosylase induce more robust HR than the initial alkylation damage *in vivo* [107]. The impact of BER and DSB repair pathway cross regulation and how the cell cycle phase contributes to pathway crosstalk to maintain genome stability are areas that require further study. Recent examples that illustrate the interplay between BER and DSB repair pathways are described here.

Alternative Non-Homologous End Joining

End joining repair of DSBs can be prone to loss of genetic material as there is no template for extensive homology searching [108, 109]. The NHEJ pathway is subdivided into two major sub-pathways that include classical NHEJ (c-NHEJ), which is dependent on the KU70/80 proteins, and alternative NHEJ (alt-NHEJ), which uses short stretches of end resection that result in microhomology often found at chromosomal translocations [108]. To elucidate which protein factors mediate alt-NHEJ, an RNAi library directed against DNA repair factors was screened using a fluorescent reporter assay for alt-NHEJ [109]. Interestingly, proteins from diverse DNA repair pathways were identified as top candidates required to perform alt-NHEJ. These proteins include the BER glycosylases, NTHL1 and UNG, the mismatch repair protein, MSH6, and the crosslink repair protein, FANCA. Subsequent analyses demonstrated that knockdown of the NTHL1 and UNG glycosylases significantly decreased alt-NHEJ. Furthermore, this result was specific for alt-NHEJ, as knockdown of either glycosylase did not have a significant impact on HR or

single strand annealing events as determined by reporter assays. HR and alt-NHEJ are most active in the S and G_2 phases of the cell cycle. Thus, accumulation of cells in G_1 upon depletion of either NTHL1 or UNG could account for the apparent decrease in alt-NHEJ. However, NTHL1 knockdown resulted in an increase in the percent of G_2 cells while UNG knockdown resulted in an increase of G_1 phase cells. These distinct cell cycle changes suggest that, at least in the case of NTHL1, the cell cycle status does not account for the decrease observed in alt-NHEJ activity. This result implies that the NTHL1 glycosylase plays a role in promoting alt-NHEJ, perhaps while suppressing HR. Future experiments will need to pinpoint how BER and end joining repair mechanisms coordinate their activities to ensure efficient repair of DSBs.

Homologous Recombination

Another well-known NTHL1 partner is the NER protein, XPG (**Figure 1.8**). Recent work reveals that XPG is indispensable for HR recovery from collapsed replication forks [110]. Genomic instability can result from an inability to repair DSBs that result from these collapsed replication forks, which are normally repaired by HR [111]. In this study, loss of XPG led to DNA damage that resulted in genomic instability [110]. XPG is required for efficient loading of the Rad51 presynaptic filament by the BRCA2/PALB2 complex for HR following end resection [110]. NTHL1 is also upregulated in S phase [55], presumably for BER glycosylase function. However, because NTHL1 appears to promote alt-NHEJ and is a binding partner for XPG, NTHL1 could modulate DSB repair pathway choice during S phase. Future experiments will be needed to assess the functional consequences of NTHL1 protein regulation for DSB repair.

In addition, recent results reveal a reciprocal effect between the BER protein, OGG1, and the HR protein, Rad52 [112]. Rad52 is part of the Rad51 epistasis group that functions in presynaptic filament formation during HR [112]. Previous studies in budding yeast demonstrated that Rad52 aids in strand exchange by forming a bridge between RPA-coated single-stranded DNA and Rad51 [113, 114]. Curiously, yeast deficient in BER are further sensitized to oxidative damage when the *RAD52* gene is disrupted, underscoring the importance of repair pathway crosstalk [115]. Studies determined that mammalian OGG1 and Rad52 proteins directly interact, and this interaction is increased in response to oxidative stress [112]. Furthermore, this interaction had reciprocal effects on the function of both the BER and HR pathways. Rad52 stimulates OGG1-mediated incision of 8-oxoG by up to 3-fold, and Rad52 promotes OGG1 release from DNA [112]. Conversely, OGG1 inhibits Rad52 single-strand annealing and strand exchange activity, while another glycosylase, UNG, has no such effect [112]. RNAi-mediated knockdown of Rad52 caused an increase in 8-oxoG and FapyG accumulation in genomic DNA [112]. Taken together, these results demonstrate that HR proteins can impact BER activity, and, conversely, a BER protein can influence HR function. These findings raise the question of whether BER glycosylases have reciprocal effects on the efficient function of the DNA repair pathways that interact with BER such as NER. Future research will be required to investigate BER protein regulation and understand how BER proteins influence the activities of other repair pathways.

Chapter 2: Overexpression of the base excision repair NTHL1 glycosylase results in genomic instability and the acquisition of cellular transformation

This chapter is modified from K Limpose, K Trego, SW Leung, J Shah, EM Werner, PK Cooper, AH Corbett, PW Doetsch.

Under revision, June 2017, Nuclei Acids Research.

2.1 Author's Contribution and Acknowledgement of Reproduction

KL, KT, SWL, JS, EMW performed the experiments. KL, AHC, PWD conceived the idea for the project. PKC, PWD, AHC contributed to experimental design and interpretation of results. KL, AHC, PWD wrote the manuscript, with feedback and editing by all authors.

2.2 Abstract

Base excision repair (BER), which is initiated by the DNA *N*-glycosylase proteins, is the frontline for repairing potentially mutagenic DNA base damage. The NTHL1 glycosylase, which excises DNA damage caused by reactive oxygen species, is thought to be a tumor suppressor. In addition to *NTHL1* loss of function mutations, our analysis and data mining of cancer genomic datasets in cBioPortal reveal that *NTHL1* frequently undergoes amplification or mRNA upregulation in certain cancer types. The contribution that NTHL1 overexpression could have to cancer has not previously been explored. We report that NTHL1 protein levels are elevated in a panel of lung cancer cell lines. To address the functional consequences of NTHL1 overexpression, transient NTHL1 overexpression was employed. Both NTHL1 and a catalytically-dead NTHL1 (CATmut) induce DNA damage and genomic instability in non-transformed human bronchial epithelial cells (HBEC) when overexpressed. Strikingly, overexpression of either NTHL1 or CATmut causes replication stress signaling and a decrease in homologous recombination (HR) activity. HBEC cells that overexpress NTHL1 and CATmut acquire the ability to grow in soft agar and exhibit loss of contact inhibition, suggesting that a catalytic-independent mechanism of NTHL1 contributes to the acquisition of cancer phenotypes. Our results demonstrate that dysregulation of a base excision repair protein, NTHL1, can induce genomic instability, interfere with HR repair activity, and result in the acquisition of cellular transformation markers.

2.3 Introduction

DNA damage can be caused by reactive oxygen species (ROS) produced by both endogenous and exogenous sources [1, 16, 21, 22]. Faced with ROS-induced DNA damage, cells have evolved the base excision repair (BER) pathway to maintain genome stability. BER is initiated by *N*-glycosylase proteins, which recognize specific subsets of base damage [1, 22]. In this initial step, cleavage of the damaged base from the phosphodiester backbone forms an apurinic/apyrimidinic (AP) site [1, 23]. The bifunctional *N*-glycosylases also possess associated AP lyase activity and cleave the DNA phosphate backbone yielding single-strand breaks with ends that require further processing by either apurinic/apyrimidinic endonuclease 1 (APE1) or polynucleotide kinase phosphatase (PNK) [1, 23]. As BER intermediates are types of DNA damage [116], a balance of BER components is crucial to ensure rapid and efficient repair of base damage, while simultaneously ensuring that potentially deleterious BER intermediates do not accumulate.

The bifunctional DNA glycosylase, NTHL1, removes a large subset of oxidatively-induced DNA base damages in mammalian cells [40]. Consistent with a critical role for NTHL1 in repairing oxidative damage and protecting the integrity of the genome, a double knockout mouse for the *NTHL1* and the *NEIL1 N*-glycosylases displays increased spontaneous tumor formation in the lung and liver [61]. More recently, studies linked biallelic inactivation of *NTHL1* to a novel human colon cancer predisposition syndrome [46, 47], presumably through impaired DNA repair. These data support a tumor

suppressive function for NTHL1 consistent with an essential role in repair of oxidative base damage.

In striking contrast to loss of NTHL1 function, cBioPortal [117, 118] cancer genomic datasets (www.cBioprtal.org) show *NTHL1* amplification or mRNA upregulation in a number of cancer types including lung, breast, prostate, and pancreas. Classic examples of gene amplification in cancer include *MYCN*, *EGFR*, and *ERBB2* in various cancer types [66]. Typically, gene amplification occurs in regions of the genome encoding genes that promote tumor formation [66]. Indeed, overexpression of other BER glycosylases has been linked to cancer outcomes. For example, overexpression of the MPG glycosylase causes temozolomide resistance in gliomas and poorer patient outcomes [67, 68], and SMUG1 glycosylase overexpression is also correlated with poor disease-free survival in gastric cancer patients [70]. Thus, an increase in NTHL1 protein could also have implications for tumorigenesis.

An increase in NTHL1 levels could be a protective response in cancer to manage an increased DNA damage burden. Alternatively, an increase in NTHL1 protein levels could directly contribute to tumorigenesis. Given that in non-small cell lung cancer (NSCLC) cBioPortal datasets *NTHL1* is amplified instead of deleted, we focused our efforts on assessing the functional consequences of NTHL1 overexpression. We show that NTHL1 overexpression causes genomic instability and cellular transformation.

2.4 Materials and Methods

Cell lines, Cell Culture and Reagents

All reagents unless otherwise stated, were obtained from Invitrogen (Carlsbad, CA, USA). The human bronchial epithelial cells (HBEC), clone 3T, [119] were a gift from Michael Story (UT Southwestern) and were cultured as previously described [120]. Karyotyping of HBEC cells in 2014 verifies the previously published HBEC karyotype [119] and validates cell line identity. The U2OS (osteosarcoma) and A549 (NSCLC adenocarcinoma) cell lines were cultured in DMEM containing 10% fetal bovine serum. Cell lines derived from HBEC soft agar clones were grown in Keratinocyte Serum-Free Media (SFM) supplemented with bovine pituitary extract and EGF. RPMI-1640 containing 5% fetal bovine serum in a 1:1 ratio, once isolated from soft agar. Beas2B cells were cultured in MEM containing 5% fetal bovine serum. H460 and H1299 cell lines were grown in RPMI containing 10% fetal bovine serum, and were provided by Dr. Wei Zhou (Emory University). The Calu-1 cell line was maintained in McCoy's 5a Modified Media and 10% FBS. H1975, H522, H226, H1792, and HCC827 cell lines were maintained in RPMI-1640 media supplemented with 10% FBS, and were obtained from the Winship Cancer Institute (Atlanta, Georgia). All cell lines were supplemented with penicillin and streptomycin and maintained at 37°C in 5% CO_2. Cells were routinely screened for mycoplasma contamination using the MycoAlert Mycoplasma Detection Kit (Lonza, Walkersville, MD, USA). Cell lines were verified by STR analysis.

DR-U2OS cells [121] were provided by Dr. M. Jasin (Memorial Sloan-Kettering Cancer Center, USA). Cells were cultured under ambient oxygen levels and 10% CO_2 in DMEM

supplemented with 10% fetal calf serum and 1% antibiotic/antimycotic.

cBioPortal Data Mining

Data was accessed from cBioPortal (www.cBioPortal.org). All cancer datasets, or specific non-small cell lung cancer (NSCLC) cancer datasets were searched for NTHL1 gene amplification and/or mRNA data. Percentages calculated in Figure 1A are the sum of values from each NSCLC dataset.

Plasmids and Site Directed Mutagenesis

The *NTHL1* gene was sub-cloned from the RG214598 plasmid (Origene, Rockville, MD, USA) using the restriction sites Sgf1 and Mlu1, and was cloned into the pCMV6-AC-GFP plasmid to create a C-terminally tagged *NTHL1* gene used in FACS sorting experiments, micronucleus, and localization studies (Origene). For NTHL1-Flag, *NTHL1* was cloned from the pCMV6-AC-NTHL1-GFP plasmid into the pcDNA3.1(+) vector using the HindIII and BamHI restriction sites (see **Table 2.1** for plasmids for Flag tag addition). The pDsRED-Express-N1 plasmid was obtained from Clontech (Mountain View, CA, USA) and used as negative control in the micronucleus experiments.

Site directed mutagenesis of *NTHL1* to create the catalytically dead NTHL1 K220Q mutant was performed on the pcDNA3.1(+) NTHL1-Flag construct (see **Table 2.1** for primers), and the Q5 Site-Directed Mutagenesis Kit (New England BioLabs, Ipswich, MA, USA). All plasmids were sequenced to ensure no mutations were inadvertently introduced, and to verify the presence of the NTHL1 K220Q mutant.

Transfection and Drug Treatments

HBEC cells were plated at a density of $2.3x10^5$ cells per well in a six well dish, trypsinized until rounded, then transfected using Fugene HD Transfection Reagent (Promega, Madison, WI) in a 3:1 (Fugene: 1 μg DNA) ratio in OPTIMEM. Cells were incubated for three hours, and transfection media was replaced with fresh HBEC media. U2OS cells were seeded at a density of $1.5x10^5$ cells per well of a six well dish and transfected with Lipofectamine 2000 (Invitrogen) for 6 hours before fresh media was added. Plasmid concentration was 1 μg per well for all experiments described in six well plate format, and scaled down for 24 well plates based on well area. Replication stress was induced in HBEC and U2OS cells with 2 mM hydroxurea (HU) (Sigma, St. Louis, MO, USA) for 24 hours in media. Camptothecin (CPT) (Sigma, St. Louis, MO, USA) treatments were performed in SFM media with 1 μM CPT for 24 hours.

RNA Isolation and Real Time PCR

HBEC, Beas2B, A549, H460, and H1299 cell lines were plated the day before at a density of $1.5x10^6$ cells per 100 mm dish. Cells were pelleted, resuspended, and divided in half for Immunoblotting and RNA preparation. Trizol RNA isolation was performed as previously described [122]. Briefly, 1 μg of total RNA isolated from cell pellet was reversed transcribed using M-MLV (Invitrogen). All real-time qPCR reactions were performed with 5 ng of cDNA, 0.5 μM of primers (see **Table 2.1** for primer sequences), and 10 μL of Quantitect Sybr Green PCR mix (Qiagen) using StepOnePlus system and software (Applied Biosystem). Post real-time analyses were performed using Microsoft Excel.

Immunohistochemistry

Immunohistochemistry (IHC) was performed by the Emory University Pathology Core using a standard IHC protocol on paraffin embedded NSCLC tissue. Novus Biological (Littleton, CO, USA) NTHL1 antibody (NB100-108) was used for IHC staining. NTHL1 antibody was validated, per manufacturers recommendation, on kidney cancer. Antibody optimization with titrations of 1:100, 1:200, and 1:400 were performed, and an antibody concentration of 1:200 was chosen for NSCLC tissue staining.

Immunoblotting

All cell lines were lysed using RIPA buffer (50 mM Tris HCL, pH 8.0, 150 mM NaCl, 5 mM EDTA, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS), except where indicated and sonicated three times at 3W and 0.5 amplitude on a Misonix Sonicator 3000 (Newtown, CT, USA). Lysates were spun in a cold room at 13,000 rpm for 15 minutes to obtain the final cleared lysate. Whole cell lysate protein concentrations were determined using the Bradford assay. The Invitrogen NuPAGE system was used with MOPS or MES buffer based on the molecular weight of the target protein (10% acrylamide gels). Transfer was performed onto nitrocellulose membranes (0.2 μm pore) in the cold room at 100V for 70 minutes. Blocking was performed in 5% Amersham ECL Prime Blocking Reagent (Pittsburgh, PA, USA) for 1 hour at room temperature. Primary antibodies used include NTHL1 1:500 (ab70726, Abcam, Cambridge, MA, USA); NEIL2 1:10,000 (180576, Abcam); Actin 1:1,000 (ab8224, Abcam); RPA 1:750 (abcam, ab2175, clone 9H8); Tubulin 1:1,000 (F2168, Sigma; Clone DM1A); ATR phospho Thr1989 1:500 (GTX128145, GeneTex, Irvine, CA, USA); ATR 1:500 (sc-1887, Santa Cruz, Santa

Cruz, CA USA); KRas 1:1,000 (F234, Santa Cruz); TSC2 1:1,000 (3612, Cell Signaling, Danvers, MA, USA); Flag 1:1,000 (2368, Cell Signaling), GFP antibody 1:2000 (#2555, Cell Signaling), and RPA phosphor Ser4/Ser8 1:500 (A300-245A, Bethyl Laboratories, Montgomery, TX, USA). Secondary antibodies used were HRP-conjugated anti-mouse 1:3,000 (W4028, Promega) and anti-rabbit 1:3,000 (W4018, Promega, Madison, WI, USA), and anti-goat 1:5,000 (sc-2020, Santa Cruz). Horseradish peroxidase was detected using the Thermo Fisher Scientific pico SuperSignal Chemiluminescent kit (#32106). Immunoblot figures were cropped and made in Photoshop. Blots were quantified using the GE Healthcare ImageQuant program.

In order to perform semi-quantitative analysis of immunoblots for fold NTHL1 overexpression change, NTHL1 and CATmut lysates were serially diluted three-fold starting at a total protein concentration of 2 μg, for final protein concentrations of 0.5 μg, 1 μg, and 2 μg for NTHL1 and CATmut protein (wedges). Empty vector control (Vector) was serially diluted, starting at 10 μg for final protein concentrations of 2.5 μg, 5 μg, and 10 μg (wedges). Ten 10 μg of non-transfected control (NT) was loaded.

DR-U2OS cells for the gene conversion assay and cells to evaluate replication stress markers were lysed in SDS sample buffer (3% SDS, 10% glycerol, 100 mM Tris-HCl, pH 6.8) and heated at 95° C for 5 minutes. Protein concentrations were determined by the BCA assay (Pierce). Samples were mixed with SDS sample buffer + 200 mM DTT + 0.5 mM bromophenol blue, prior to resolution on 4-12% Tris-Glycine gels (Invitrogen), and were transferred to nitrocellulose membranes and probed with the corresponding

antibodies. Antibodies used were mouse anti-HA (clone 16B12, MMS-101P, Covance), mouse anti-tubulin (CP06, Calbiochem), mouse anti-FLAG (F3165, Sigma), and sheep anti-mouse (NA931V, GE Healthcare) conjugated to horseradish peroxidase (HRP).

ImageQuant was used to quantitate immunoblots for the replication stress signaling. All protein values were normalized to actin to control for protein loading. The ratio of phosphorylated protein:total protein in each lane was then determined. Values were normalized so that HU control signal was set as 1.

Cytokinesis-block Micronucleus Assay

HBEC cells were plated at a density of 30,000 cells per well and U2OS cells were plated at a density of 20,000 cells per well in a 24 well plate on glass coverslips. Media in the HBEC wells was supplemented with 5 µl of fetal bovine serum per ml. Cells were treated with 3 µg/ml Cytochalasin B in media and stained as previously described [120]. Cells were imaged and scored as described [120] using an Olympus IX81 inverted microscope with SlideBook 5.0 (3i Intelligent Imaging Innovations, Denver, CO, USA).

Immunofluorescence and Microscopy

HBEC cells were seeded at a density of 30,000 cells per well on glass coverslips. Cells were transfected as described, and 24 hours after transfection, were washed in cytoskeletal (CSK) buffer (0.15% Triton-X 100, 360 mM piperazine diethanesulfonic acid, 25 mM HEPES, 10 mM EGTA, 2 mM magnesium chloride, pH to 6.9) supplemented with protease and phosphatase inhibitors for 5 minutes on ice, followed by

two washes with PBS on ice, in order to retain chromatin bound protein. Cells were fixed in 4% paraformaldehyde. Primary antibodies used were γH2AX 1:500 (Millipore, Billerica, MA, USA) and 53BP1 1:1000 (Novus Biologicals, Littleton, CO, USA). Imaging was performed on a Zeiss LSM510META confocal microscope (Thornwood, NY, USA) using a 20×Plan-Apo objective (NA=0.75). Three independent experiments were performed, and 50 cells per replicate were counted for quantification.

Comet Assay

Comet assays were performed as previously described [123, 124]. In brief, cells were transfected with each NTHL1 construct and cells were suspended in low melting point agarose 24 hours following transfection. Cells were then alkaline lysed, electrophoresed, and neutralized, followed by staining with SYBER green. Cells were imaged on an Olympus IX81 inverted microscope with SlideBook 5.0. Fifty cells from each replicate experiment were analyzed using the Comet Score software (TriTek Corporation).

Gene Conversion and Single-Strand Annealing Assays

DR-U2OS cells were seeded, and transfected 24 hours later with Lipofectamine2000 (Invitrogen) and 3 μg total DNA plasmid consisting of I-SceI-encoding plasmid (1.5 μg) and pcDNA3.1 plasmid (1.5 μg) with empty vector control, NTHL1-FLAG, or NTHL1-FLAG cat dead. Flow cytometry was carried out 72 hours later to measure GFP positive cells. GFP conversion was normalized to GFP (%) in the empty vector control for each biological experiment. Statistical significances were determined using the Student's T test.

Soft Agar Colony Formation

HBEC cellular transformation was measured by using soft agar colony formation was performed as described except that NTHL1 overexpressing cells were plated 72 hours following transfection [125]. In brief, low melting temperature agarose containing a 1:1 ratio of HBEC media and RPMI-1640 media was diluted to a final concentration of 0.7% and solidified at 4 °C in each well of a 6 well plate. After fluorescence activated cell sorting (FACS) of GFP positive cells or following transient transfection, 55,000 cells per well were immediately plated in a top layer of low melting temperature agarose at a final concentration of 0.37%. Cells were cultured at 37 °C with 5% CO_2 for three weeks. Upon colony appearance, the media was changed from HBEC media to a 1:1 ratio of HBEC and RPMI-1640 media. Colonies were visually counted through an Olympus IX81 inverted microscope. Colonies were isolated from soft agar, trypsinized into single cell suspension, and grown into clonal cell lines for loss of contact inhibition assays and persistent genomic instability.

Loss of Contact Inhibition

Loss of contact inhibition was measured as described [126]. Loss of contact inhibition was visualized macroscopically and on an Olympus IX81 inverted microscope with QCapture (QImage, Surrey, CAN) using a10x objective

FACS Sorting

Cells were plated at a density of $1.7x10^6$ cells per 100 cm^2 dish and transfected with NTHL1-GFP plasmid or GFP alone. Twenty four hours following transfection, cells were

trypsinized, spun down, and resuspended in filter-sterilized 1x phosphate buffered saline containing 5 mM EDTA, 25 mM HEPES pH 7.0, 1% BSA, and 10 U/ml of DNAse. Cells were sorted on a BD FACSAria machine using a 100 μM diameter nozzle. Cells were binned into three populations based on NTHL1-GFP intensity deemed as high, intermediate, and low based on the observed distribution of GFP positive cells. NTHL1-GFP positive cells from each population were immediately diluted and plated in fresh Keratinocyte serum-free media (SFM) and allowed to recover before each assay or the addition of cytochalasin B.

Statistical Analysis and Data Representation

Error bars in figures represent the standard error from the average of at least three independent experiments. Each measurement was performed in duplicate. Two-way ANOVA was performed in Graph Pad for the analysis of HBEC micronucleus formation at 48 hours and 96 hours, with a=0.05. One-way ANOVA (a=0.05) was calculated in VasserStats for all other experiments, unless otherwise indicated. Micronuclei frequency distribution and DSB frequency distribution was analyzed using the chi-squared test. P values are as follows: $p \leq 0.05$ * ; $p \leq 0.01$ ** ; $p \leq 0.001$ *** ; $p \leq 0.0001$ ****

Project	Primer Direction	Primer Sequence	Original Plasmid	Plasmid Cloned Into	Reference
NTHL1-Flag Subcloning	Forward	5'- ACACTGGCGGCCGTTACTAGTGGATCCT-3'	pCMV6-AC-NTHL1-GFP	pcDNA3.1(+)	Origene (PS100010) Invitrogen (V79020)
	Reverse	5'- ACGACTCACTATAGGGAGACCCAAGCTT-3'			
CATmut-Flag Site Directed Mutagenesis	Forward	5'-TGTTGGGCCCCAGATGGCACACCTGG-3'	pcDNA3.1(+)	——	Invitrogen (V79020)
	Reverse	5'-CCCGGCAGCGCCACCAGC-3'			
qRT-PCR NTHL1	Forward	5'-CAGCATCCTGCAGACAGATGA-3'	——	——	Sigma
	Reverse	5'-GTCCACTGCAATGCCTGACAC-3'			

Table 2.1: Cloning and oligonucleotide constructs used in experiments. Plasmids used for NTHL1 cloning and the corresponding oligonucleotides used are displayed with their sources. Oligonucleotides for qRT-PCR of NTHL1 in control and non-small cell lung cancer are displayed.

2.5 Results

NTHL1 protein levels are elevated in non-small cell lung cancer (NSCLC) cell lines

Recent genomic analysis of multiple families links homozygous loss-of-function germline mutations in *NTHL1* to a novel colon cancer predisposition syndrome [46, 47]. In light of these observations, we mined cBioPortal (www.cBioprortal.org) [117, 118] datasets to assess whether *NTHL1* deletion is linked to other cancer types. Surprisingly, *NTHL1* was more frequently amplified than deleted in multiple cancer types, and *NTHL1* mRNA was also upregulated in many instances. Non-small cell lung cancer (NSCLC) displayed a pattern where the *NTHL1* gene was amplified or mRNA upregulated, with no loss of *NTHL1* reported (**Figure 2.1A**). Collectively, *NTHL1* amplification and mRNA upregulation occurs in ~21% of NSCLC cases (**Figure 2.1A**). Expansion of cBioPortal analysis to other BER genes reveals a pattern where mRNA is frequently upregulated in NSCLC (**Figure 2.1A**). Therefore, a subpopulation of NSCLC tumors may have increased NTHL1 protein.

To determine whether *NTHL1* amplification or mRNA upregulation corresponds to an increase in protein levels, we compared the level of NTHL1 protein in a panel of transformed and non-transformed lung epithelial cell lines. Steady-state NTHL1 protein levels are low in non-tumorigenic HBEC and Beas2B cells, but higher in several NSCLC cell lines examined (**Figure 2.1B and Figure 2.2**). In contrast, protein levels for another BER glycosylase, NEIL2, or protein levels for a gene adjacent to *NTHL1*, TSC2, show relatively consistent expression across the cell lines examined (**Figure 2.1B**). Quantification of immunoblots reveals that the increase in NTHL1 protein

A

Non-Small Cell Lung Cancer (NSCLC)					
BER gene	% Amplified	% Loss	% Mutated	% mRNA upregulation	% mRNA downregulation
NTHL1	4.3	0	2.7	17	0
OGG1	1	1.5	2.7	11.6	1.2
NEIL1	0.9	0.5	3.3	10.2	0
NEIL2	1.6	22.3	2	20.9	9.1
NEIL3	0.4	11.3	13.7	26	0
APE1	7.8	0	1.3	33	0

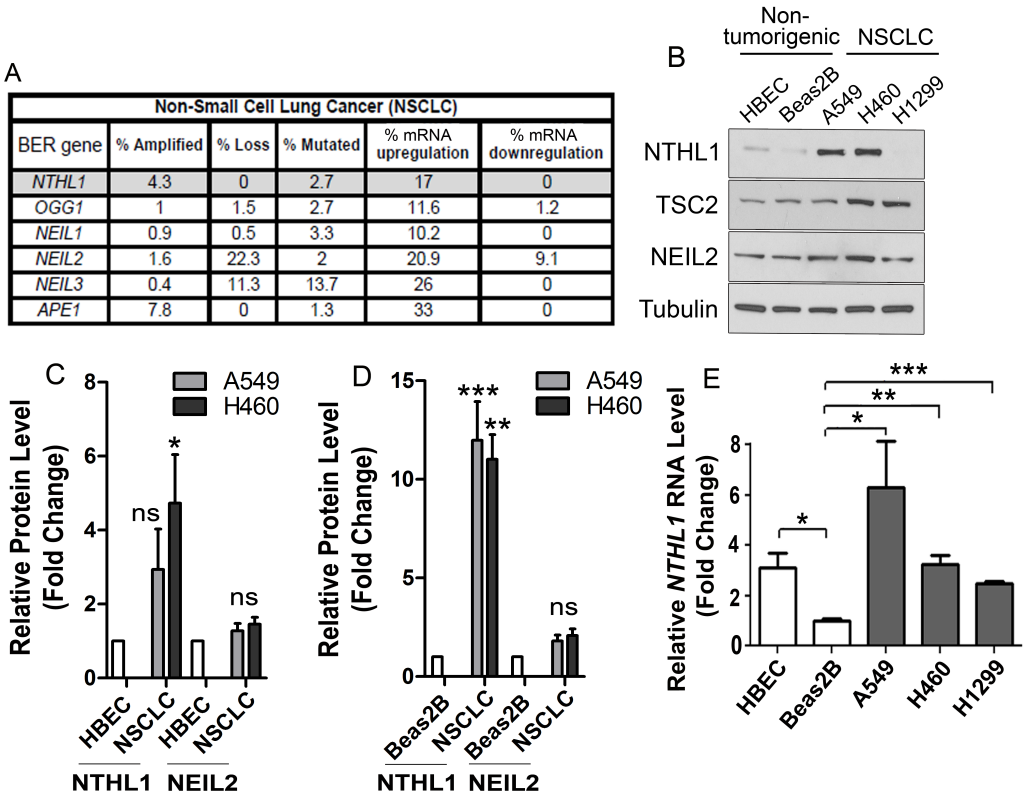

Figure 2.1: **NTHL1 mRNA and protein levels are increased in non-small cell lung cancer (NSCLC) cell lines.** A) cBioPortal primary tumor data sets across all subtypes of NSCLC demonstrate that multiple base excision repair (BER) genes are amplified in addition to loss or mutation in NSCLC. Messenger RNA (mRNA) dysregulation for each BER gene is also observed. B) Immunoblotting for NTHL1 in a panel of non-transformed, immortalized lung epithelial cell lines (HBEC and Beas2B) and NSCLC cancer cell lines (A549, H460, H1299) demonstrate variable NTHL1 protein expression levels. Levels of another BER glycosylase, NEIL2, are constant across all cell lines examined. TSC2 protein levels are also constant in cell lines examined. C) Quantification of relative NTHL1 protein levels from A549 and H460 were normalized to the HBEC cell line. D) Quantification of relative NTHL1 protein levels from A549 and H460 were normalized to the Beas2B cell line. E) qRT-PCR analysis of *NTHL1* transcript levels reveals that *NTHL1* mRNA levels vary between non-transformed and transformed cell lines, and that mRNA levels do not necessarily correspond to NTHL1 protein levels. NS= not significant; $p \leq 0.05 = *$; $p \leq 0.01 = **$; $p \leq 0.001 = ***$

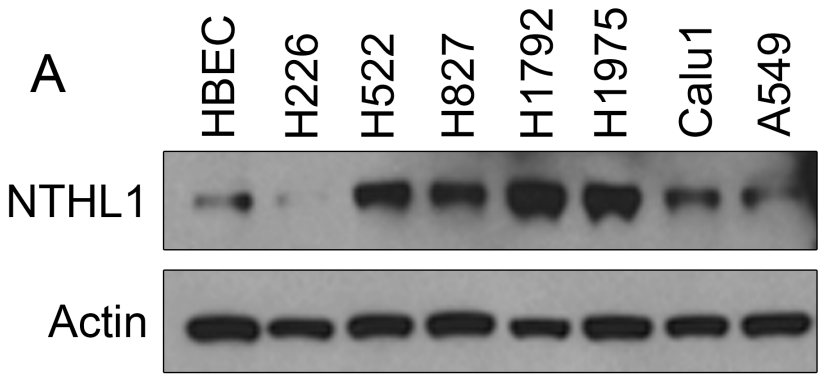

Figure 2.2: NTHL1 protein levels vary between non-small cell lung cancer cell lines (NSCLC). Immunoblot of a panel of NSCLC cell lines reveals that NTHL1 is elevated in NSCLC compared to non-transformed HBEC cells. Gels for lung cancer screens were run with a total protein content of 25 μg of protein

levels in NSCLC cells lines (A549 and H460) is significant when compared to HBEC and Beas2B cell lines (**Figure 2.1C and 2.1D**). To determine whether *NTHL1* transcript levels correlate with protein levels in these cell lines, qRT-PCR was performed. This analysis reveals that *NTHL1* mRNA levels may not necessarily correspond to *NTHL1* protein levels in the cell lines examined (**Figure 2.1E**) indicating the complex regulation of NTHL1 at multiple levels. Nonetheless, these results demonstrate that, in general, steady-state NTHL1 protein is increased in NSCLC cell lines compared to control cells.

NTHL1 is primarily localized to the nucleus in both NSCLC and overexpressing HBEC cells

NTHL1 localization was assessed in matched normal tissue and primary NSCLC tumor tissue by immunohistochemistry (**Figure 2.3A**). NTHL1 is predominately localized to the nucleus in both normal lung and NSCLC primary tissue (**Figure 2.3A**). To assess whether overexpressed NTHL1 is also localized to the nucleus, HBEC cells were transfected with a plasmid that transiently overexpresses NTHL1-GFP. As shown in Figure 2B, NTHL1-GFP is localized predominately to the nucleus in these cells. Furthermore, NTHL1 is chromatin-bound, as NTHL1-GFP signal survives pre-extraction of the soluble fraction with cytoskeletal (CSK) buffer (**Figure 2.3B**). GFP protein alone is not targeted to the nucleus and is soluble, as GFP is removed by CSK pre-extraction (**Figure 2.3B**). Thus, transiently overexpressed NTHL1-GFP localizes to the nucleus and associates with chromatin.

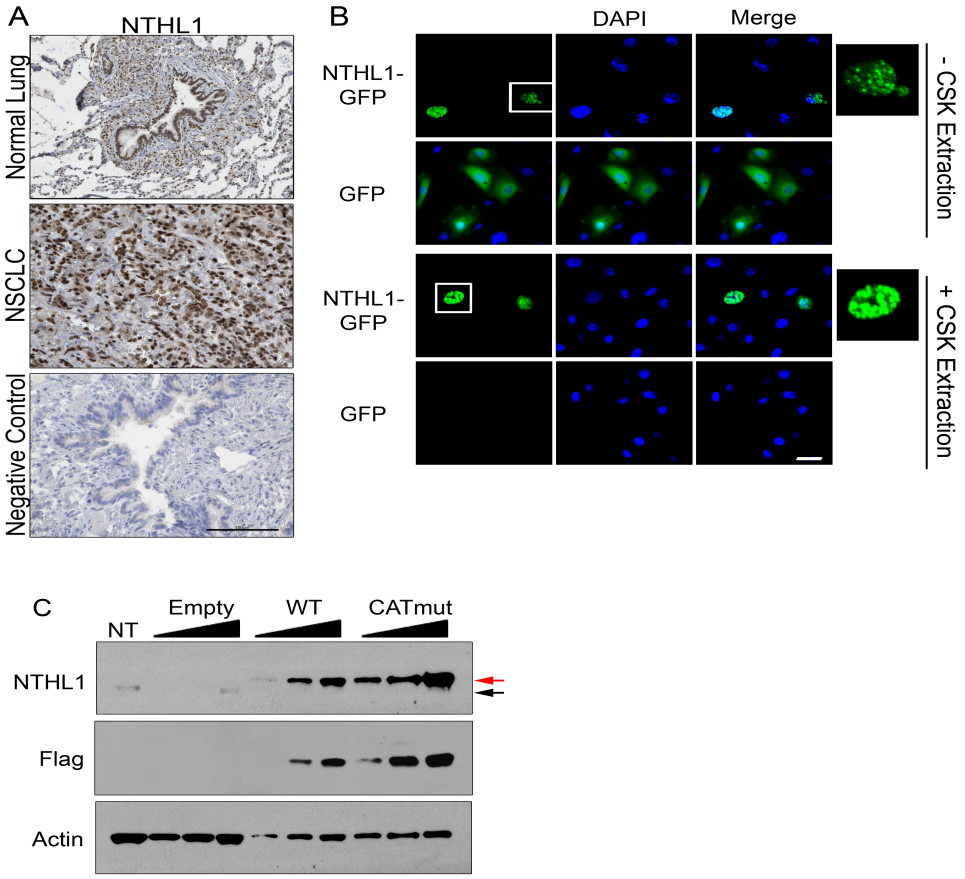

Figure 2.3: NTHL1 is localized to the nucleus in both NSCLC tissue and in HBEC cells. A) Immunohistochemistry of NTHL1 protein (brown) in normal lung and NSCLC tissue. Negative control represents a normal serum control from mice in the absence of primary NTHL1 antibody. The black scale bar corresponds to 100 μm. B) HBEC cells expressing NTHL1-GFP reveal that NTHL1 is localized to the nucleus and is chromatin-associated as demonstrated by pre-extraction with cytoskeletal (CSK) buffer. GFP alone is not targeted to the nucleus or chromatin-associated as demonstrated by CSK pre-extraction. Cells inside the white box have been enlarged and inset on the right. The white bar corresponds to 100 μm. C) Immunoblot of transiently expressed NTHL1-Flag in HBEC cells. The same protein lysate was serially diluted by half to include final concentrations of 2 μg, 1 μg, and 0.5 μg of protein for wildtype NTHL1 and CATmut. Total protein concentration loaded for assessing endogenous NTHL1 in HBEC cells (non-transfected; NT) was 10 μg. The empty vector control (Empty) was diluted by half to final protein concentrations of 10 μg, 5 μg, and 2.5 μg. The lower limit of detection of endogenous NTHL1 from HBEC cells is 10 μg of total protein lysate. The black arrow indicates endogenous NTHL1, while the red arrow denotes exogenously expressed NTHL1-Flag.

To assess the relative fold overexpression of wildtype NTHL1 (NTHL1) compared to endogenous NTHL1, NTHL1 was Flag-tagged. We also created a Flag-tagged, catalytically dead NTHL1 (CATmut) by replacing lysine 220 with glutamine [52]. Because attempts to generate HBEC cell lines stably overexpressing NTHL1 and CATmut proteins were unsuccessful, we used a transient overexpression model for NTHL1 and CATmut. Both NTHL1 and CATmut proteins were highly overexpressed (11-54 fold) compared to endogenous NTHL1 (**Figure 2.3C**).

DNA damage is induced by overexpression of NTHL1 and CATmut

An increase in NTHL1 protein as seen in NSCLC lines could be a protective mechanism in response to increases in cellular reactive oxygen species. However, as BER intermediates are themselves types of DNA damage, increases in NTHL1 protein could overwhelm downstream BER processes, resulting in accumulation of BER intermediates. To determine whether NTHL1 overexpression causes an increase in DNA damage, we used the alkaline comet assay to analyze the sum of DNA single-strand breaks and double-strand breaks [127, 128]. Transient overexpression of NTHL1 or CATmut protein in HBEC cells results in a significant increase in DNA damage (**Figure 2.4A and Figure 2.5A**). This result suggests that elevated NTHL1 levels increase the cellular load of DNA damage in a manner that does not depend on NTHL1 catalytic activity.

As both NTHL1 and CATmut overexpression cause DNA damage, we examined the nature of the DNA damage. NTHL1 or CATmut was overexpressed and co-localization of the DNA double-strand break (DSB) markers γH2Ax and 53BP1 was assessed by

immunofluorescence (**Figure 2.4B**). As a positive control, camptothecin, a topoisomerase I inhibitor that causes DSBs associated with replication [129], was used. In the absence of DNA damaging agents, transient NTHL1 and CATmut overexpression induce DSBs (**Figure 2.4B**) as compared to non-transfected (NT) or empty vector (Vector) controls (**Figure 2.4B**). Quantification of the cells with DSB foci and the number of DSB foci per nucleus (**Figure 2.4C**) reveals that NTHL1 and CATmut overexpression induce a significant increase in the number of cells with DSBs (**Figure 2.4C**). Overexpression of NTHL1 induces a significantly greater percentage of cells (40%) with DSBs compared to CATmut (25%) (**Figure 2.4C**). However, the proportion of cells having large numbers of DSB foci per nucleus (≥3) is equally increased upon NTHL1 or CATmut overexpression (**Figure 2.4C**). Our results support the idea that a novel, catalytically-independent function of NTHL1 contributes to the generation of DSBs.

As DSBs are prevalent in HBEC cells overexpressing both NTHL1 and CATmut, we investigated whether increased NTHL1 protein levels impact DSB repair by homologous recombination (HR). To measure DSB repair by HR, we assayed gene conversion in the direct repeat GFP (DR-GFP) reporter construct integrated into a U2OS cell line [110, 121] (**Figure 2.5B**). This assay directly interrogates HR activity following an I-Sce1-induced DSB in the GFP gene. If HR is active, the cleaved GFP gene can be repaired with a second, transcriptionally inactive GFP gene resulting in recovery of GFP fluorescence. GFP positive cells are measured using flow cytometry. NTHL1-Flag or CATmut-Flag protein was overexpressed in DR-U2OS cells containing this reporter

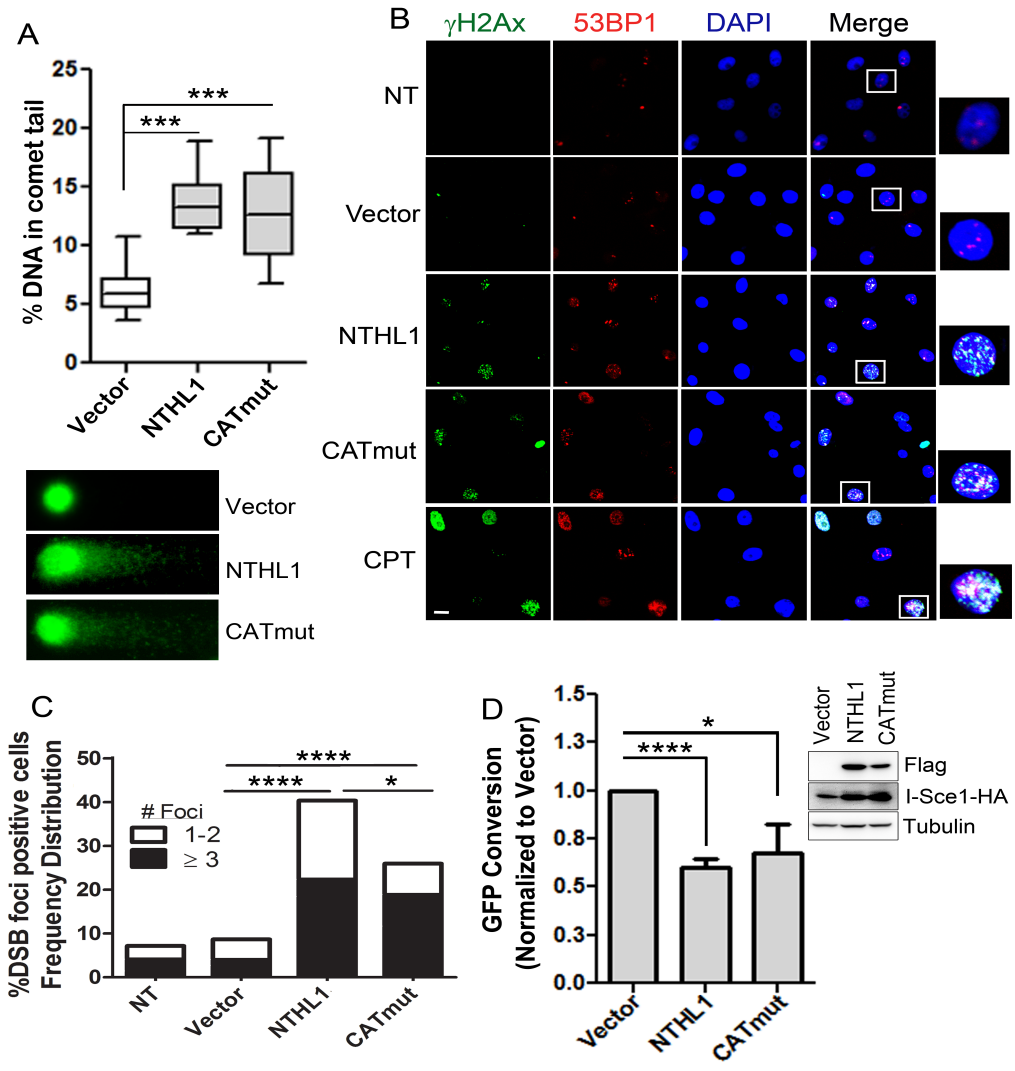

Figure 2.4: DNA damage is induced by overexpression of wildtype NTHL1 and CATmut A) Alkaline comet assay of HBEC cells transiently overexpressing NTHL1 or CATmut NTHL1 24 hours following transfection. DNA damage was measured by %DNA in comet tail. DNA comets were labeled using SYBER green, and representative images of comet tails are shown. B) Immunofluorescence of double-strand break (DSB) markers following NTHL1 or CATmut transient overexpression in HBEC cells. Twenty-four hours following transfection, cells were stained to assess colocalization of γH2Ax (green) and 53BP1 (red) markers for DSBs. Camptothecin (CPT) serves as a positive control for DSBs. The white box indicates an enlarged image of a cell with DSBs, and the white bar corresponds to 50 μm. C) Quantification of the total number of DSB positive nuclei, and DSB foci number per nucleus from experiments in panel B. D) NTHL1 and CATmut overexpression impairs homologous recombination assayed by a reporter for gene conversion in DR-GFP U2OS cells. The fraction of GFP positive cells was determined and normalized to the empty vector control (Vector). Immunoblot insert showing NTHL1 and CATmut expression, along with I-Sce1-HA tagged expression. P values are as follows $p \leq 0.05 = *$; $p \leq 0.001 = ***$; $p \leq 0.0001 = ****$

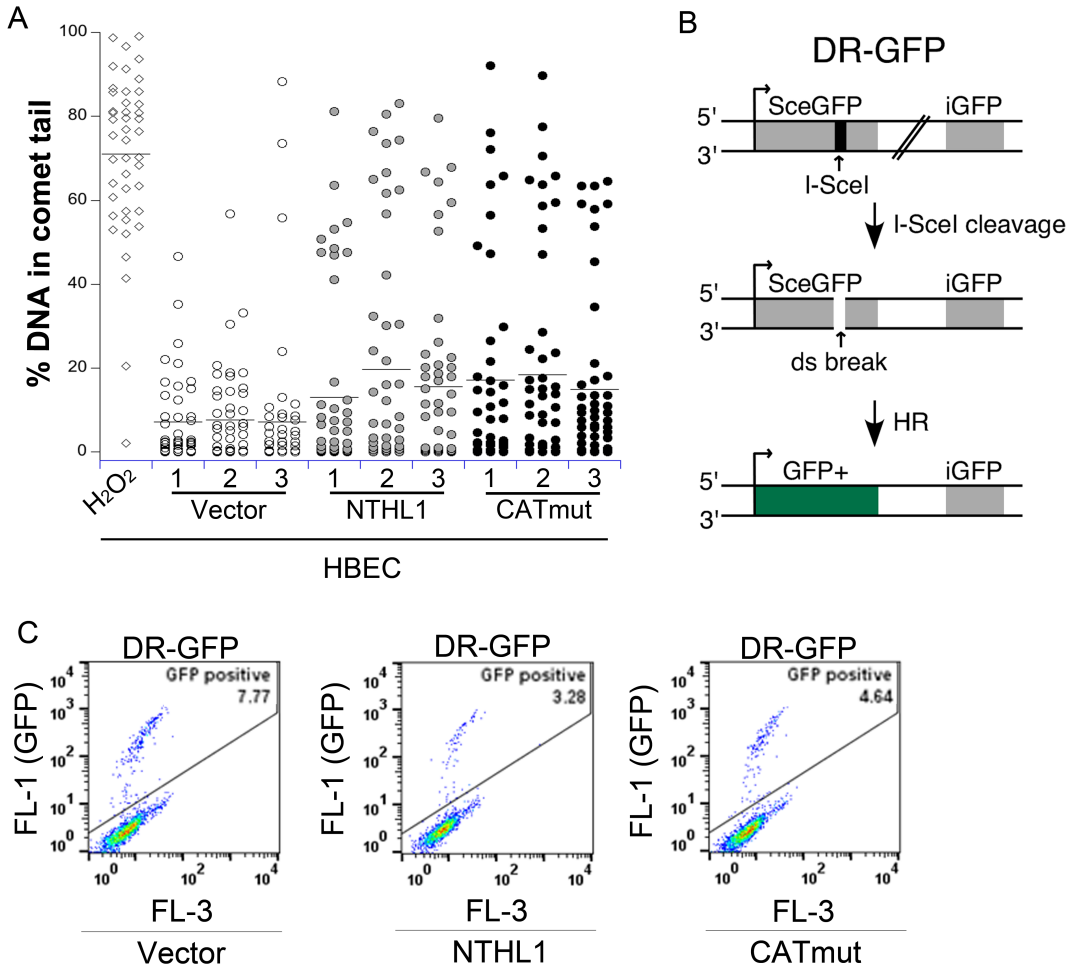

Figure 2.5: DNA damage is induced by NTHL1 and CATmut overexpression. A) Dot plot of DNA damage (% DNA) in comet tails for 50 individual HBEC cells from three independent experimental replicates. Hydrogen peroxide (H_2O_2) treatment was used as a positive control for DNA damage. NTHL1 and CATmut expression results in a cell population that contains an increase in DNA damage compared to empty vector (Vector) control cells. B) Functional schematic of the DR-GFP assay used to measure homologous recombination (HR) activity in U2OS cells. The DR-GFP cassette contains two inactive GFP genes separated by a DNA linker. One is disrupted by the DNA sequence recognized by the exogenously introduced I-Sce1 restriction enzyme (SceGFP). The other is an inactive GFP that serves as the homologous recombination template for repair (iGFP). Upon expression of the I-Sce1 enzyme, a DSB is generated at the I-Sce1 cleavage site. If HR is active, the DSB will be repaired using the iGFP cassette to generate a repaired SceGFP gene that results in fluorescently active GFP protein (GFP+). C) Flow cytometry on U2OS cells for GFP signal (FL-1) after an I-Sce1 induced DSB. GFP positive cells represent cells that are able to repair the I-Sce1 break via HR. NTHL1 and CATmut overexpressing cells display a decrease in the percentage of cells that repair the DSB by HR compared to empty vector (Vector) control cells.

system (**Figure 2.4D**). Overexpression of NTHL1 or CATmut causes an apparent ~50% decrease in HR capacity compared to empty vector control (Vector) (**Figure 2.4D and Figure 2.5C**). Our results suggest that increased NTHL1 or CATmut protein inhibits HR repair of DSBs, providing an explanation for the DSB accumulation in both NTHL1 and CATmut overexpressing cells.

Genomic instability is induced by NTHL1 and CATmut overexpression

To investigate the cellular consequences of NTHL1 and CATmut-induced DSBs, we performed the cytokinesis-blocked micronucleus assay to assesses genomic instability [130]. To verify that cells displaying genomic instability specifically overexpress NTHL1, we isolated NTHL1-GFP expressing cells by fluorescence activated cell sorting (FACS). We further divided the GFP positive cell population into cells expressing low, intermediate or high levels of NTHL1 based on the intensity of the GFP signal (**Figure 2.6A and Figure 2.7A**). Forty-eight hours after sorting, binucleated cells were counted and scored for micronucleus formation. A significant increase in micronucleus formation was observed in cells overexpressing NTHL1-GFP but not GFP alone (**Figure 2.6A and Figure 2.7B**). Furthermore, micronucleus formation correlated with the level of NTHL1-GFP overexpression and was observed even in the population with low levels of overexpression (**Figure 2.6A**). NTHL1 protein levels are displayed (**Figure 2.6B**) from each sorted population, verifying that GFP intensity detected by FACS reflects the amount of protein in the cell. U2OS cells overexpressing either NTHL1-GFP or CATmut-Flag also display micronuclei formation (**Figure 2.6C)**.

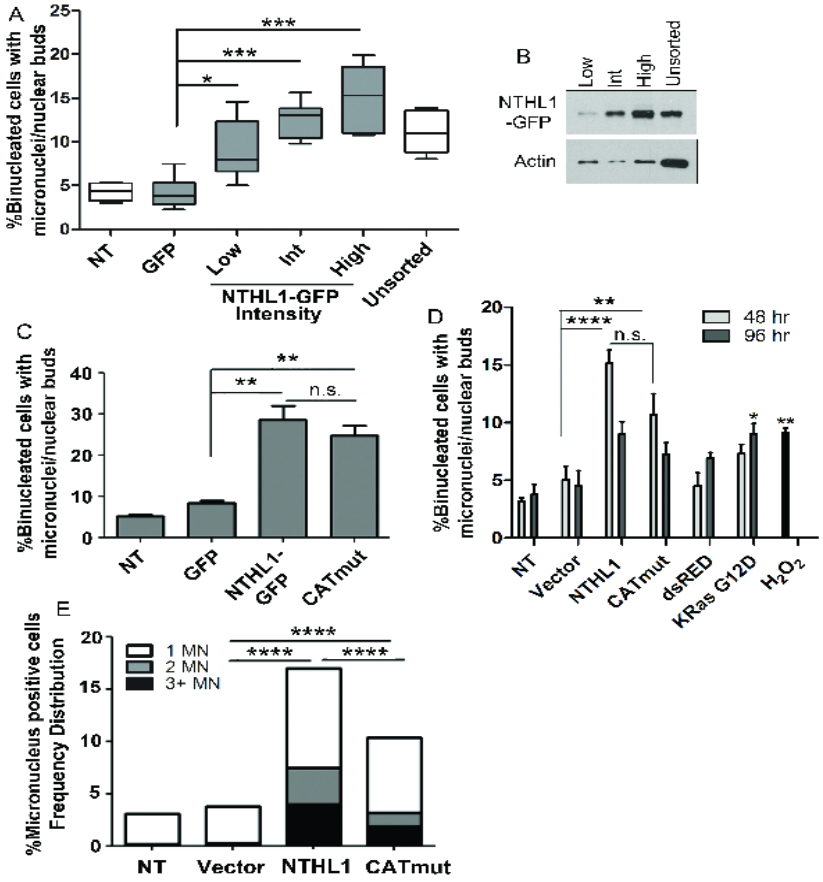

Figure 2.6: Genomic instability is induced by wildtype NTHL1 and CATmut overexpression. A) HBEC cells expressing NTHL1-GFP were FACS sorted based on GFP signal intensity. GFP positive cells were further sorted into three populations based on the intensity of the GFP signal into low, intermediate (Int), and high groups, and micronucleus formation was measured in each population and compared to unsorted NTHL1-GFP cells. Non-transfected (NT) cells and cells expressing GFP alone (GFP), were used as controls. B) Immunoblot of NTHL1 from the FACS sorted cell population of low, Int, and high GFP intensity signals. C) U2OS cells transiently overexpressing either NTHL1-GFP or CATmut-Flag were analyzed for micronucleus formation. D) Overexpression of NTHL1 and CATmut in HBEC cells results in micronuclei induction at 48 following transfection compared to empty vector (Vector) control cells. Expression of red fluorescent protein (dsRED) or of G12D mutant K-Ras and treatment with hydrogen peroxide (H_2O_2) were all used as controls for micronucleus formation. KRas G12D produced statistically significant micronuclei induction at 96 hours compared to empty vector control. dsRED is not statistically significant compared to vector control at 48 and 96 hours. E) The number of micronuclei (MN) per binucleated cell increases in NTHL1 and CATmut overexpressing HBEC cells. Cells were scored as having 1, 2, or 3 or more micronuclei (MN) per binucleated cell. NS= non-statistically significant; $p \leq 0.05 = *$; $p \leq 0.01 = **$; $p \leq 0.001 = ***$; $p \leq 0.0001 = ****$

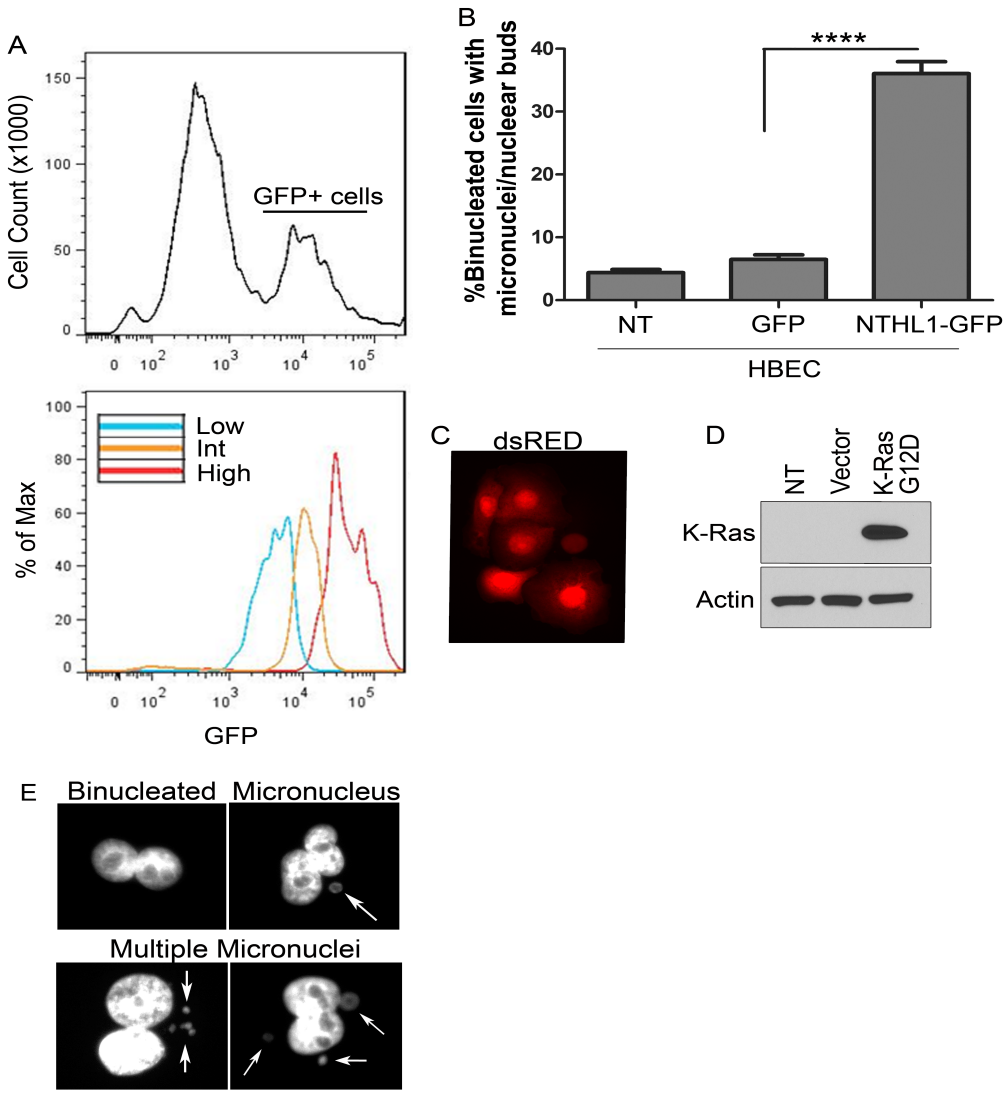

Figure 2.7: Genomic instability is induced by NTHL1 and CATmut overexpression. A) HBEC cells expressing NTHL1-GFP were FACS sorted for GFP. Cells were sorted in three different populations based on GFP intensity termed low, intermediate (Int), and High. B) HBEC cells overexpressing NTHL1-GFP were scored for micronucleus formation. Genomic instability is specific to cells overexpressing NTHL1-GFP and not seen from GFP expression alone. C) Ectopic expression of red fluorescent protein (dsRED) was employed as a control for general protein overexpression in HBEC. D) Expression of the G12D mutant form of K-Ras (a well-known oncogene variant in NSCLC) was employed as a positive control for micronucleus assays. E) Representative examples of a binucleated cell without a micronucleus, typical binucleated cell with one micronucleus, and binucleated cells with multiple micronuclei, as scored for micronucleus assays with NTHL1 and CATmut overexpression.

To determine the time course of genomic instability, NTHL1 and CATmut were overexpressed in HBEC and micronucleus formation monitored at 48 and 96 hours following transfection. NTHL1 or CATmut overexpression induces significant micronucleus formation compared to empty vector control at 48 hours following transfection, with no statistically significant difference between them (**Figure 2.6D**). Increased micronuclei formation is still observed at 96 hours but is not statistically significant compared to empty vector control. Overexpression of an unrelated protein such as dsRED (**Figure 2.6D and Figure 2.7C**) did not induce micronucleus formation, whereas positive controls of treatment with 100 μM H_2O_2 or overexpression of K-Ras G12D protein do induce micronuclei formation (**Figure 2.6D** and **Figure 2.7D**) [131].

In addition to causing elevated levels of micronuclei, NTHL1 and CATmut overexpression also induce the appearance of multiple micronuclei per binucleated cell (**Figure 2.6E** and **Figure 2.7E**). Control non-treated cells (NT) and cells transfected with empty vector alone (Vector) had <4% of binucleated cells with micronuclei. Cells overexpressing NTHL1 or CATmut display >15% or >10% of binucleated cells with micronuclei formation, respectively (**Figure 2.6D and 2.6E**). Remarkably, binucleated cells displaying two (~3% or ~1%, respectively) or even three or more (~4% or ~2%, respectively) micronuclei were observed (**Figure 2.6E**). We conclude that both NTHL1 and CATmut overexpression are potent inducers of genomic instability.

NTHL1 and CATmut overexpression induce replication stress signaling

We sought to determine how NTHL1 overexpression results in DSB formation. As unrepaired DSBs during S phase can result in genomic instability and micronuclei formation [132-134], we investigated whether replication stress contributes to the observed cellular phenotypes. Immunoblotting for known replication stress signaling proteins was performed. Phosphorylation of the ATR (ataxia telangiectasia and Rad3 related) kinase takes place during the replication stress response [135]. As part of this signaling cascade, ATR undergoes autophosphorylation at threonine 1989 (T1989) [135], and then phosphorylates downstream targets such as Chk1 to trigger the intra-S phase checkpoint [136, 137]. Hydroxyurea (HU) treatment, which induces replication stress and ATR activation [135], was used as a positive control (**Figure 2.8A**). Cells overexpressing NTHL1 and CATmut both activated ATR (pATR T1989) (**Figure 2.8A**) and Chk1 (S317) (**Figure 2.8B**), consistent with the presence of replication stress [136]. DSBs resulting from replication fork collapse activate ATM [137], which phosphorylates downstream effectors such as γH2Ax and Chk2 [137]. Phosphorylation of Chk2 on threonine 68 (T68) occurs in response to DNA damage and is essential to mount a proper response to DSBs [137]. Cells overexpressing NTHL1 or CATmut both display an increase in pChk2 T68 compared to negative controls (**Figure 2.8B**). The presence of γH2Ax and 53BP1 foci (**Figure 2.4**), together with pChk2, indicate that overexpression of NTHL1 or CATmut induces DSBs during replication stress through a mechanism that does not require NTHL1 catalytic activity and is consistent with the demonstrated reduction in HR.

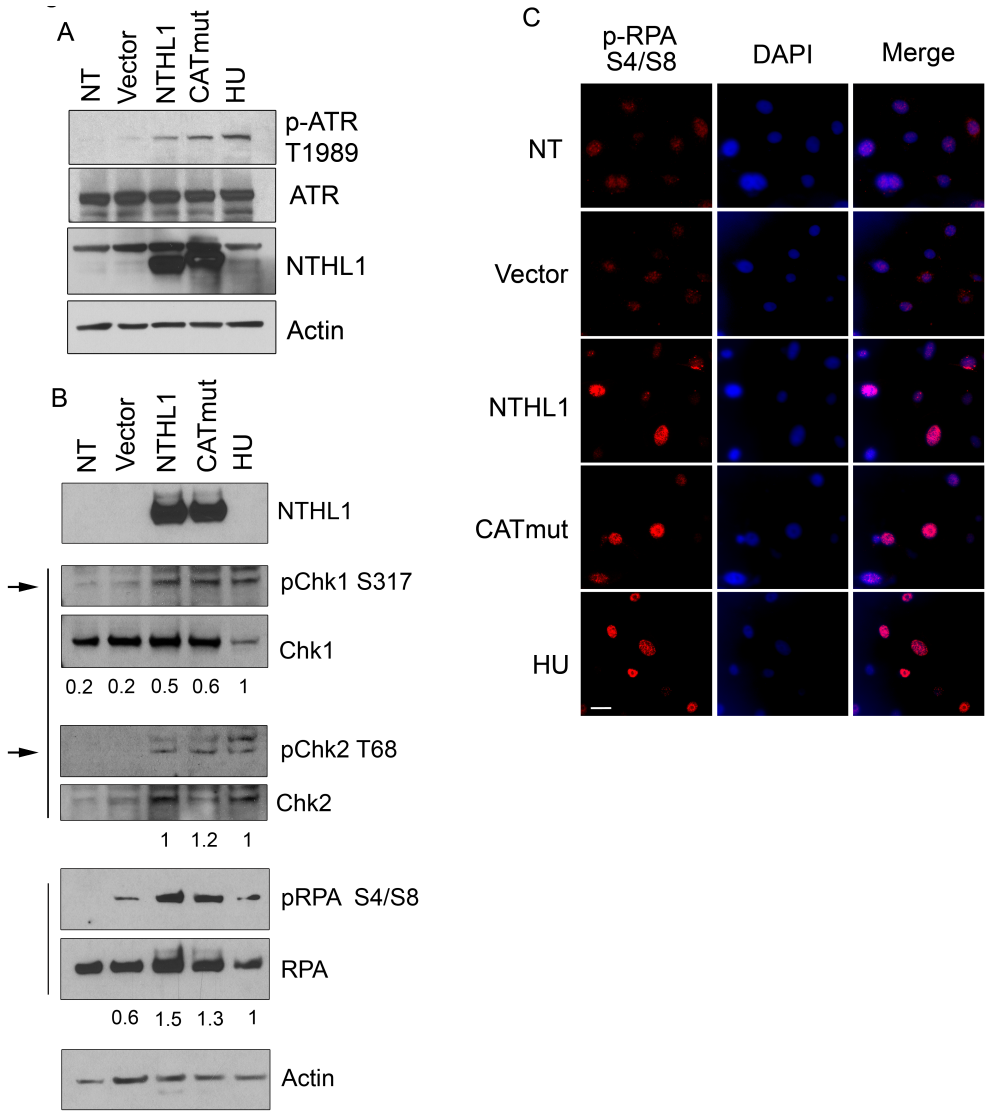

Figure 2.8: Wildtype NTHL1 and CATmut overexpression results in replication stress signaling. A) Immunoblot of HBEC cells transiently overexpressing NTHL1 and catalytically dead NTHL1 (CATmut) 24 hours following transfection. Hydroxyurea (HU) is a positive control for replication stress. Autophosphorylation of ATR at threonine 1989 (p-ATR T1898) is a replication stress marker, and is induced by overexpression of NTHL1 and CATmut. B) Immunoblots showing that cells overexpressing NTHL1 or CATmut display replication stress response signaling by phosphorylation of Chk1 (pChk1 S317) and phosphorylation of Chk2 (pChk2 T68). The arrows indicate the bands at the correct size for each pChk1 S317 and pChk2 T68, while the upper bands are non-specific binding of the antibody. Total RPA and RPA phosphorylation at serine 4 and 8 (pRPA S4/S8) are shown as a marker of HR. Numbers below each lane correspond to the ratio of phosphorylated protein to total protein content. C) Immunofluorescence of the HR marker, pRPA at serine 4 and serine 8 (pRPA S4/S8), in HBEC cells transiently overexpressing NTHL1 and CATmut. The white bar corresponds to 50 μm.

DSBs generated from replication stress are primarily repaired by HR [138], which is initiated by DNA end resection to create single-stranded DNA (ssDNA) [138, 139]. Replication protein A (RPA) coats the ssDNA and is phosphorylated at serine 4 and 8 (pRPA S4/S8) [140]. NTHL1 and CATmut overexpression both induce increased pRPA S4/S8 signal compared to non-transfected (NT) and empty vector controls (**Figure 2.8B and 2.8C**). Each induces at least a two-fold change in the amount of phosphorylated protein for pChk1 and pRPA compared to empty vector control. We conclude that NTHL1 and CATmut overexpression results in increased replication stress signaling and DSB formation that likely account for the observed genomic instability.

Early cancer hallmarks are conferred by NTHL1 and CATmut overexpression

As replication stress and genomic instability are inducers of early cellular markers of transformation [138, 141, 142], we investigated whether NTHL1 overexpression could induce these phenotypes. The ability to form colonies in soft agar is an established criterion for assessing one of the earliest cancer hallmarks, anchorage independent growth [143]. HBEC cells expressing NTHL1-GFP were FACS sorted and GFP-positive cells were plated for the soft agar assay. NTHL1-GFP expressing cells form colonies in soft agar, while cells expressing GFP alone display minimal colony formation (**Figure 2.9A**). HBEC cells overexpressing CATmut-Flag also form colonies in soft agar (**Figure 2.9B**).

Colonies were isolated from soft agar, and two surviving clones for each NTHL1 variant were expanded into cell lines to test for loss of contact inhibition. Non-transformed

HBEC cells form an organized, epithelial monolayer (**Figure 2.9C**). In contrast, clonal cell lines expressing NTHL1 or CATmut show loss of contact inhibition with foci formation and morphological disorganization comparable to the NSCLC cell line, A549 (**Figure 2.9C**). Thus, loss of contact inhibition develops in overexpressing cells that acquired the capability to grow in soft agar.

We assessed NTHL1 protein expression in each clonal cell line and determined that NTHL1 levels were moderately elevated over control HBEC cells, in contrast to the high levels seen in transiently transfected HBEC (**Figure 2.9D**). However, quantification showed that the differences in NTHL1 expression were not statistically significant compared to non-transformed HBEC (**Figure 2.10A**). To investigate whether pre-existing high endogenous levels of NTHL1 selected for soft agar colonies, cells from the parental HBEC cell line were clonally expanded to analyze the variation in expression of endogenous NTHL1. Levels of endogenous NTHL1 in individual parental cells do not reach comparable levels to NTHL1 detected in the soft agar clones (**Figure 2.10B**). Thus, selection in soft agar is most likely not due to high endogenous levels of NTHL1 found in a subset of parental HBEC cells (**Figure 2.10B**). Importantly, these results demonstrate that transient overexpression is sufficient to induce transformation.

Because we show that transient NTHL1 and CATmut overexpression leads to genomic instability, we evaluated whether the clonal cell lines derived from soft agar colonies demonstrated persistent genomic instability as another cancer hallmark. These clones display an elevated level of micronucleus formation that is comparable to the A549

cancer cell line (**Figure 2.9E**). NTHL1 overexpression therefore contributes to the generation of genomic instability that results in transformation of HBEC cells, and through soft agar selection, persistent genomic instability is a permanent characteristic of these transformed cells.

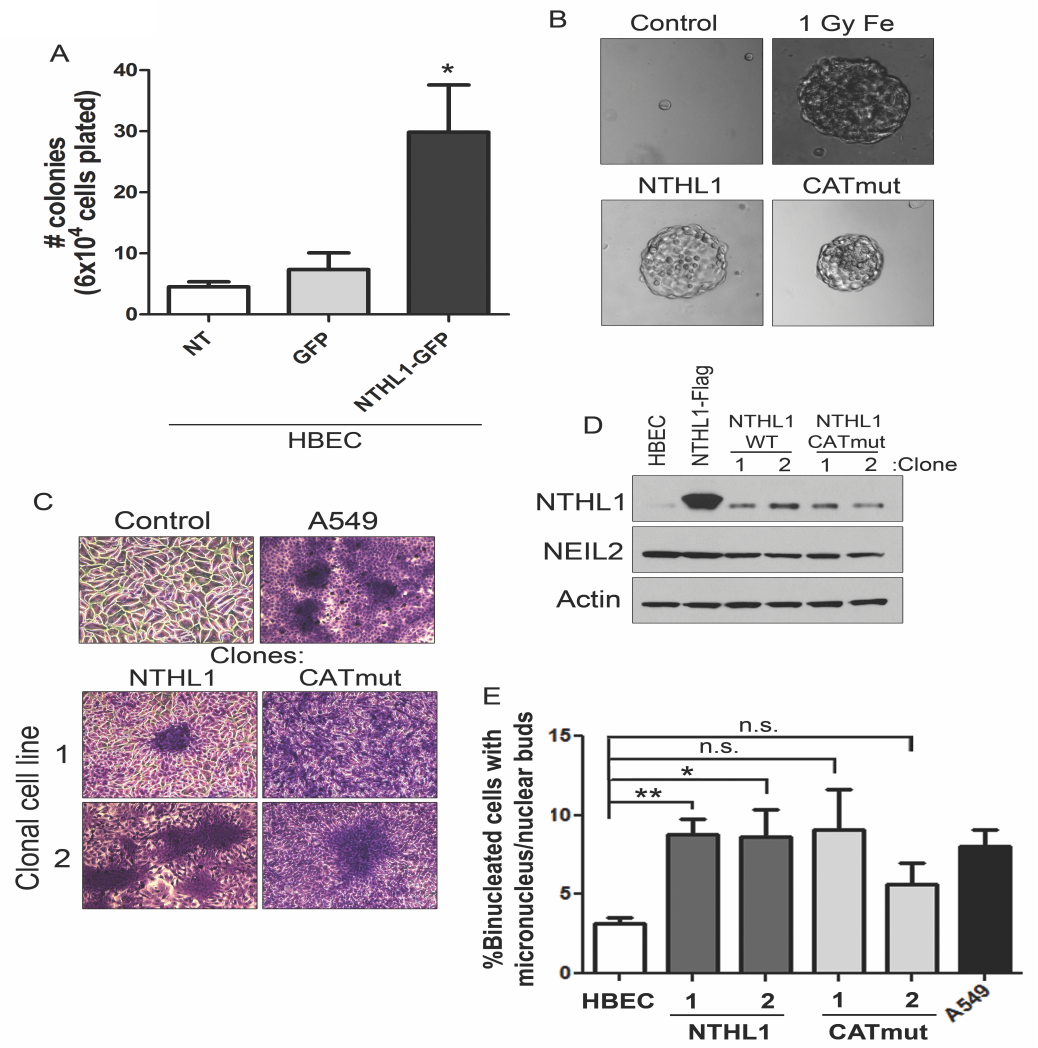

Figure 2.9: Enabling cancer hallmarks are conferred by wildtype NTHL1 and CATmut overexpression. A) HBEC cells were FACS sorted based on NTHL1-GFP expression. Cells were assayed for colony formation in soft agar. B) HBEC cells overexpressing NTHL1-Flag or CATmut-Flag were grown in soft agar. Parental HBEC cells (control) did not form soft agar colonies, while HBEC cells transformed with 1 Gy of Iron ion irradiation (1 Gy Fe) did form colonies. C) Cell lines derived from HBEC soft agar colonies transformed by NTHL1 or CATmut overexpression were crystal violet stained to display loss of contact inhibition, compared to the HBEC parental cells (control) monolayer growth and comparable to the A549 NSCLC line. Two independent cell lines were created from soft agar colonies for NTHL1 and CATmut. D) Immunoblot of NTHL1 levels in each of the isolated clonal cell lines. NTHL1-Flag overexpression in lane two is a transient overexpression. Protein expression levels of another BER glycosylase, NEIL2, do not change following colony formation. E) Clonal cell lines derived from soft agar demonstrate persistent genomic instability measured by the micronucleus assay compared to non-transfected (NT) HBEC cells and the A549 line as a positive control. NS= non-statistically significant; $p \leq 0.05$ * ; $p \leq 0.01$ ** ; $p \leq 0.001$ *** ; $p \leq 0.0001$ ****

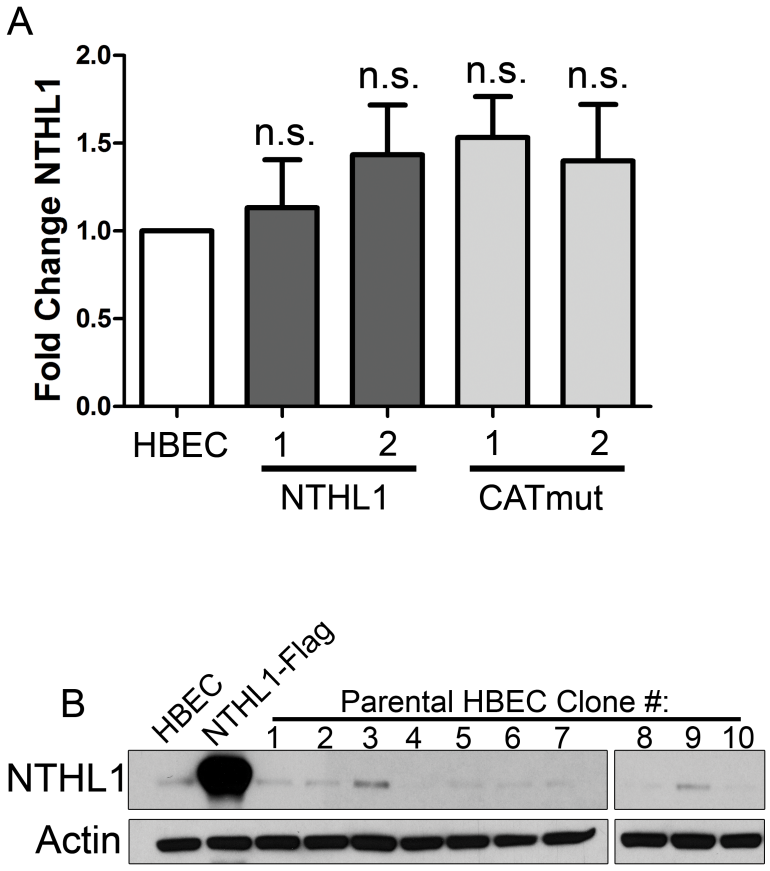

Figure 2.10: NTHL1 expression levels in soft agar clones and parental HBEC clones. A) Quantification of immunoblot of NTHL1 protein showing fold change of protein levels in soft agar clones compared to the parental HBEC line. B) Immunoblot of NTHL1 expression in the HBEC line, cells transiently overexpressing NTHL1, and ten different clonal derivatives of non-transfected HBECs. Immunoblots were run and developed at the same time, but were transferred onto different membranes in order to include all ten non-transfected HBEC clones.

2.6 Discussion

Our results demonstrate that overexpression of NTHL1 can cause in genomic instability and cellular transformation independent of catalytic activity. Previous studies identified a contribution of loss of NTHL1 to a colon cancer predisposition syndrome [46, 47], presumably through the accumulation of mutations and transformation over time (**Figure 2.11A**, left side). However, as we demonstrate, overexpression of NTHL1 can also contribute to transformation (**Figure 2.11A**, right side). Therefore, a proper balance of BER protein components is needed to protect genome stability, as either loss or overexpression of a BER component can negatively impact the genome.

In our system, transient overexpression of NTHL1 was employed to investigate the consequences of BER dysregulation. A range of NTHL1 overexpression was observed; however, the induction of DNA damage (**Figure 2.4**) and genomic instability (**Figure 2.6**) did not greatly vary between experimental replicates, suggesting that a minimum threshold for NTHL1 overexpression exists to trigger the observed phenotypes. Interestingly, the observed genomic instability and cellular transformation did not require sustained high levels of NTHL1 overexpression (**Figure 2.9**), suggesting that transient high levels of NTHL1 are able to initiate processes that can lead to transformation. In support of this concept, variants of the BER protein DNA polymerase β require only transient overexpression to generate loss of contact inhibition and growth in soft agar in non-transformed mouse mammary tissue [144], a result which parallels the transient NTHL1 overexpression analyzed here.

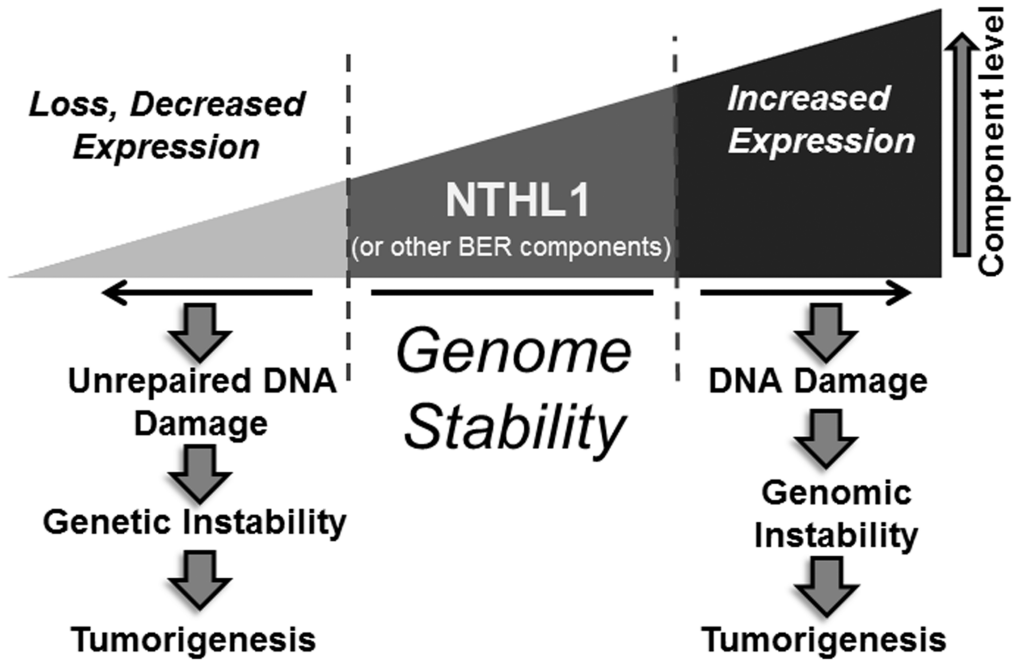

Figure 2.11A: Proposed models for NTHL1 regulation/dysregulation. A) Loss or decreased expression of *NTHL1* is linked to a novel cancer predisposition syndrome [46, 47], presumably through lack of DNA repair for NTHL1 substrates, yielding mutations and ultimately tumorigenesis. Alternatively, increased expression of NTHL1 protein levels can lead to DNA damage that drives genomic instability and contributes to tumorigenesis. Therefore, optimal protein levels of BER components are needed to balance the cellular need for DNA damage repair without initiating genomic instability.

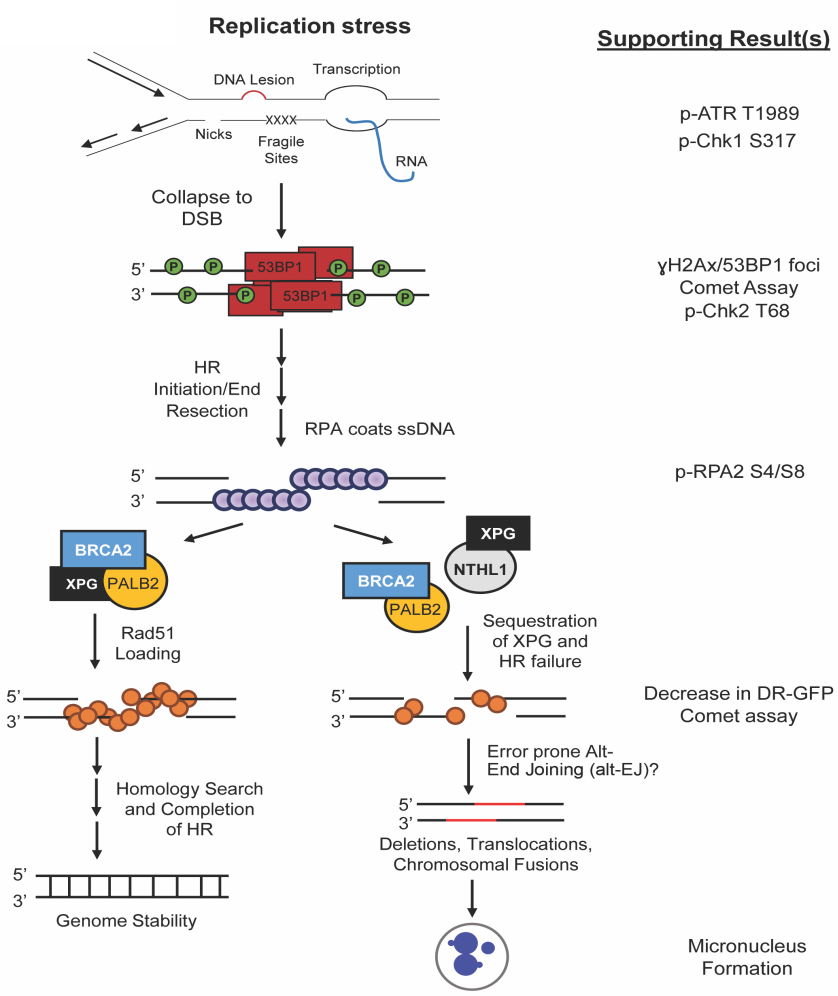

Figure 2.11B: Proposed models for NTHL1 regulation/dysregulation. B) Replication stress that arises during S phase may occur through multiple means, including unrepaired DNA lesions, accumulated DNA repair intermediates, nicks in the phosphodiester backbone, fragile sites, and encounters with transcription, among other causes. A collapsed replication fork results in a DSB that is repaired by homologous recombination (HR). Phosphorylation of γH2Ax and binding of 53BP1 signals the presence of a DSB. BRCA1 displaces 53BP1 in S phase to initiate HR by end resection. Single-stranded DNA is then coated by RPA, and RPA phosphorylation on serine 4 and serine 8 (pRPA S4/S8) signal for RAD51 loading. Initiation of HR requires formation of the RAD51 presynaptic filament through displacement of RPA by BRCA2 with PALB2 and DSS1, and efficient RAD51 loading is dependent on XPG. Completed HR results in genome stability, as a homologous template is used to repair the DSB. Upon NTHL1 overexpression, XPG may be titrated away from duties related to RAD51 loading. This results in chronic replication stress signaling, as evidenced by pATR T1989. Inefficient RAD51 loading onto DNA results in HR failure, and alt-EJ is initiated to salvage DSB repair, as DNA resection has already occurred. Alt-EJ is error prone and results in chromosomal deletions, translocations, and chromosome fusions that can lead to genome instability, as assessed by micronucleus formation.

Our data reveal a novel mechanism of inducing genomic instability and cellular transformation that does not depend on the catalytic activity of NTHL1 (**Figure 2.4, Figure 2.6, and Figure 2.9**). This surprising result suggests that a protein-protein interaction may underlie some consequences of NTHL1 overexpression (**Figure 2.11B**). One candidate interacting partner for NTHL1 that could affect HR activity is XPG [40, 145]. XPG is an endonuclease that functions in NER [146] and interacts with NTHL1 to enhance NTHL1 binding to and excision of DNA damage substrates [40]. XPG is also critical for HR activity and for recovery from replication fork collapse [110]. Cells deficient for XPG show increased DSBs as marked by γH2Ax and 53BP1 foci, elevated micronucleus formation, and decreased HR activity [110], consequences that are mirrored by NTHL1 overexpression in this study. The decrease in HR activity upon NTHL1 overexpression (**Figure 2.4**) is similar to the decrease in HR activity detected in cells that lack XPG when analyzed using the same DR-GFP reporter assay [110]. Further supporting disruption of HR by NTHL1 overexpression, we show that pRPA S4/S8 signaling is increased (**Figure 2.8B and C**), as it is upon loss of XPG. Given the similarity in phenotypes, we suggest that independent of its catalytic activity, excess NTHL1 could bind to and sequester XPG, impairing the loading of RAD51 (**Figure 2.11B**). This BER/HR pathway crosstalk would render cells vulnerable in S phase when HR is tasked with repairing DNA damage that might give rise to genomic instability [110, 138, 141].

Upon HR pathway disruption, DSBs can also be repaired by end joining processes, including classical non-homologous end joining (c-NHEJ) and alternative end joining

(alt-EJ) [106, 147]. While c-NHEJ preferentially functions during the G_1 phase [148], alt-EJ is active in S phase and is a highly error prone process [138, 147, 149]. Upon disruption of HR, an increase in alt-EJ activity may occur, and this HR/alt-EJ interplay is independent of c-NHEJ [139]. As end resection has already occurred in this scenario [139, 147], error prone alt-EJ may be utilized to complete DSB repair. We propose a model in which cells that overexpress NTHL1 sequester XPG protein and thus are deficient in the correct repair of DSBs that routinely arise in each cell cycle through collapse of stalled replication forks. These cells may then utilize alt-EJ to complete DSB repair, resulting in genomic instability (**Figure 2.11B**). Future work will need to examine whether alt-NHEJ plays a role in causing the cellular phenotypes observed upon NTHL1 overexpression.

A prediction of these results is that cancer cells overexpressing NTHL1 would be vulnerable to clinical agents that mechanistically induce DSBs. In support of this idea, a previous study found that lymphoblastoid cells in tissue culture display increased sensitivity to ionizing radiation upon NTHL1 overexpression [69]. This study concluded that this increase in sensitivity was due to processing of closely opposed NTHL1 substrates to generate DSBs. However, our results show that NTHL1 overexpression can directly interfere with HR repair of DSBs, which could contribute to radiation sensitivity.

In further support of our findings, a previous study reported cellular transformation as a result of expressing a distinct, catalytically-inactive germline variant of NTHL1 (D239Y) in the MCF10a breast cell line, in which DSBs, chromatid breaks, and chromosomal

fusions were observed [126]. However, these phenotypes were attributed to unrepaired cytotoxic lesions. The results thus differ from our findings, in which the resulting cellular transformation can occur independent of catalytic activity. The previous study did not evaluate wildtype NTHL1 or D239Y NTHL1 overexpression levels in comparison to endogenous NTHL1 levels, but the conditions used were stable expression in an immortal MCF10A breast epithelial cell line, in contrast to our high levels of transient overexpression in the HBEC line. Whether the different outcomes relate to differences in overexpression levels, differences in the cell lines used, or different consequences of the inactivating mutation tested are not clear. Nevertheless, both the previous study and our findings dramatically illustrate the importance of properly regulated NTHL1 to maintain genome stability.

While most outcomes in our study showed no statistically significant difference between NTHL1 and CATmut, we detected a difference for the number of DSB foci per cell (**Figure 2.4C**) and the number of micronuclei per binucleated cell (**Figure 2.6E**). This result is consistent with a potential contribution of NTHL1 catalytic activity to the generation of DNA damage. However, as we employed transient transfections and obtained a wide range of overexpression, we cannot formally exclude the possibility that differences in NTHL1 protein expression underlie this difference.

This study provides evidence suggesting that increases in BER glycosylase protein levels could contribute to tumorigenesis. Understanding proper regulation of BER components

(**Figure 7A**) will provide further insight into how dysregulation of BER is involved in cancer etiology.

2.7 Acknowledgements

This work was supported by National Institutes of Health grant number ES011163 (PWD), GM058728 (AHC), ES019935 (PKC). We would like to acknowledge the Doetsch and Corbett labs for their discussions and advice. "This study was supported in part by the Emory Flow Cytometry Core (EFCC) and the Emory Integrated Genomics Core (EIGC), part of the Emory Integrated Core Facilities (EICF), and is subsidized by the Emory University School of Medicine. Additional support was provided by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR000454. Research reported in this publication was also supported in part by the Winship Cancer Institute of Emory University and NIH/NCI under award number P30CA138292 and the Maiola Family Fund for lung cancer research. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health."

Chapter 3: General Discussion

Part of this chapter is adapted from the review article “BERing the burden of damage: Pathway crosstalk and posttranslational modification of base excision repair proteins regulate DNA damage management," authored by Kristin Limpose, Anita H. Corbett, and Paul W. Doetsch. DNA Repair, Accepted April 2017 for publication in August 2017.

Reproduced in accordance with Elsevier Publishing policy for dissertations (allowed to reuse for the purpose of dissertations or thesis):
https://www.elsevier.com/about/our-business/policies/copyright/permissions

KL and PWD generated the article concept. KL, AHC, and PWD all contributed to writing and editing the article for publication.

3.1 Summary

In this thesis, we provide evidence that in addition to the loss of BER as a contributing factor to tumorigenesis, an increase in the BER protein, NTHL1, can also cause genomic instability and cellular transformation. NTHL1 overexpression causes these phenotypes through the accumulation of DSBs via the inhibition of HR repair. We expected the glycosylase and AP lyase activity of NTHL1 to contribute to the accumulation of DNA damage due to accumulation of repair intermediates. Surprisingly, overexpression of NTHL1 can cause genomic instability independent of NTHL1 enzymatic activity. Furthermore, DSBs that are generated during S phase can exacerbate the genomic instability phenotype if unrepaired DSBs are present as cell enter mitosis. However, as the number of DSBs and micronuclei were elevated for wild-type NTHL1 compared to catalytically inactive NTHL1, at least two possible mechanisms may contribute to the genomic instability that we detect in cells that overexpress NTHL1: 1) NTHL1 may interact with and sequester key proteins required for maintaining genome stability. Currently, our data point toward a critical non-catalytic function for NTHL1, suggesting a protein-protein interaction may underlie the observed phenotypes, and 2) NTHL1 enzymatic activity may contribute to the accumulation of BER intermediates that can lead to DSBs. Quite likely, both of these mechanisms could be relevant and the type of tissue or tumor could dictate which mechanism is most relevant.

We have not yet defined how NTHL1 overexpression can generate cellular transformation independent of enzymatic activity. Our current hypothesis entails a protein-protein interaction whereby, NTHL1 titrates and sequesters key proteins required

for the HR pathway. One strong candidate is the XPG protein, which interacts with NTHL1 [40]. This sequestration would mean that XPG was unavailable to perform XPG-mediated RAD51 loading in conjunction with BRCA2 to complete HR repair [110]. This loss of XPG function could contribute to the genomic instability observed upon overexpression of NTHL1. Interestingly, cells that are depleted for the XPG protein display similar phenotypes to overexpression of NTHL1, namely accumulation of DNA damage, impaired HR activity, and genomic instability [110]. While we favor inhibition of HR through NTHL1-mediated titration of XPG as a component of cell transformation, sequestration of other key proteins could contribute to these phenotypes. For instance, the concentration of the single-strand binding protein, RPA, is rate limiting for replication and repair [150]. Depletion of the cellular pool of RPA due to DNA damage during S phase leads to "replication catastrophe". As a result, unprotected single-strand DNA is vulnerable to strand breaks during replication and generates genomic instability [150]. Future studies will address the mechanism(s) that are involved in cellular transformation induced by NTHL1 overexpression. Furthermore, we should examine whether the enzymatic activity of NTHL1 contributes to genomic instability through accumulation of BER intermediates. Teasing apart the mechanism(s) that contribute to NTHL1-induced transformation will yield a greater understanding of tumorigenic processes at the basic science level and potentially provide new targets for future clinical applications.

Importantly, our work adds a new paradigm to the regulation of BER glycosylases at the protein level. Previous studies identified a contribution of the loss of NTHL1 protein to a colon cancer predisposition syndrome [46, 47], presumably through the accumulation of

mutations and transformation over time. Loss of NTHL1 protein as a contributing factor to cancer appeared to solidify NTHL1 as a genuine tumor suppressor. However, as we demonstrate, overexpression of NTHL1 can also contribute to cellular transformation. Therefore, a proper balance of BER protein components is needed to protect genome stability, as either loss or overexpression of a BER component can negatively impact the genome. Glycosylase overexpression phenotypes may be effected through inappropriate protein-protein interactions between DNA repair pathways. This pathway crosstalk can subsequently influence the repair function(s) of other DNA repair pathways. Our work provides the first example of BER/HR pathway crosstalk in a cellular context to impact HR repair activity.

3.2 DNA Repair Pathway Crosstalk Implications

As each DNA repair pathway was characterized beyond the individual biochemical steps, it became apparent that coordination between DNA repair pathways is essential for proper cellular responses to DNA damage. We refer to such coordination as pathway crosstalk. To illustrate and examine DNA repair pathway crosstalk, we utilized the STRING protein-protein interaction network (www.string-db.org) [151]. Various types of protein interactions and databases are included in the STRING analysis. Visualization of the interaction map is straightforward; each protein is represented by a node (circle), while a protein interaction is represented by an edge (line) (**Figure 3.1**). We included a panel of DNA repair proteins from five repair pathways (BER, NER, MMR, NHEJ, and HR), which yields three main clusters after Kmeans clustering analysis with a high confidence (0.700) minimum interaction score.

By examining the three different clusters, one can appreciate the breadth of coordination not only within a specific pathway but also between DNA repair pathways (**Figure 3.1**). For example, the red cluster highlights BER. Dashed lines indicate crosstalk with the black supercluster containing components of NER, MMR, and NHEJ. One could postulate that a BER interaction may, in fact, influence the HR pathway (yellow) by modulating a common interaction highlighted within the black supercluster. As DNA repair is a tightly regulated process, perturbations in pathway crosstalk may have untoward consequences for multiple DNA repair pathways. Therefore, understanding the nuances of regulation at the protein level by identifying central interaction hubs could be an effective approach to identifying new targets for therapeutic development, or for predicting how a patient may respond to existing chemotherapeutic options that target DNA repair. For instance, if a protein coordinates functional interactions between multiple DNA repair pathways, such as MLH1, developing a targeted therapeutic towards MLH1 will have a broader impact on DNA repair as multiple repair pathways could be affected at once. Hypothetically, cancer cells that rely on MLH1 to coordinate these repair interactions may be rendered sensitive to further DNA damage from conventional chemotherapeutics.

To detect central interaction hubs, we propose a two-fold approach of assessing 1) the total number of edges that a node has, and 2) the number of dashed edges per node. In this way, interaction hubs for a specific pathway, or a hub that impacts the greatest number of interactions between pathways can be identified. For example, OGG1 has six solid edges to denote interactions with other BER components while nine dashed edges

represent crosstalk with various components of the supercluster for a total of fifteen edges. Furthermore, many MMR proteins have interactions connected to HR. One could hypothesize that dysregulation at the protein level anywhere along this string of interactions could impact the functions of BER, MMR, and/or HR simultaneously. The same logic can be applied along any node-edge pathway.

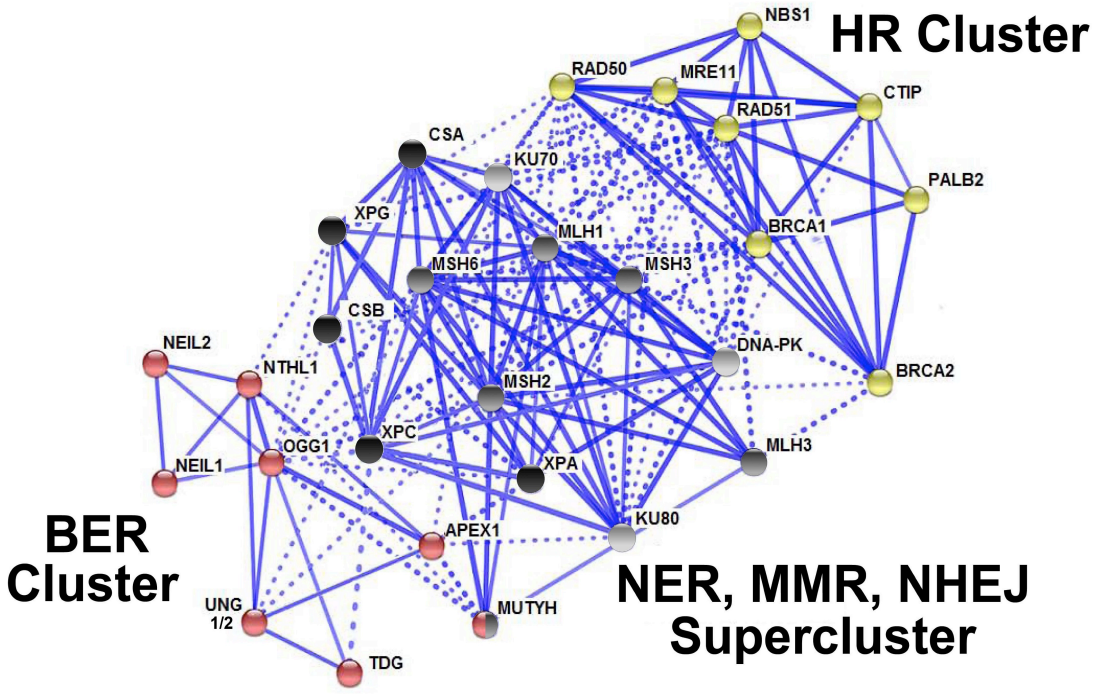

Figure 3.1 Human base excision repair protein interactions with other DNA repair pathway components. The STRING functional protein association network and Kmeans clustering reveal how DNA repair components interact [151]. A circle represents a node (protein), while edges (lines) indicate protein interactions. Solid edges denote interactions within the same cluster (pathway), while dashed lines are pathway crosstalk interactions that take place between proteins of different clusters. The line thickness is related to the strength of data that supports a particular interaction. Three clusters of interactions were generated and color-coded according to the canonical function of the protein- a BER cluster (red); a supercluster comprised of MMR (grey), NHEJ (light grey), NER (black), and an HR cluster (yellow). The MMR/NER/NHEJ supercluster was generated by the STRING algorithm during Kmeans clustering. The STRING map demonstrates that DNA repair pathways dynamically interact with each other as demonstrated by the dashed edges

Another striking observation is the large number of interactions that appear to be coordinated through the MMR proteins MLH1, MSH6, and MSH3 (**Figure 3.1**). These proteins emerge as central coordinators between certain BER proteins and HR. For example, an interesting case emerges with the MSH6 protein. MSH6 is implicated in promoting alt-NHEJ [109]. From the STRING interaction network, one can see that MSH6 interacts with RAD51 as well as components of the BER pathway (**Figure 3.1**). Whether MSH6 can promote alt-NHEJ while suppressing HR through the interaction with RAD51 remains to be determined. Alternatively, whether MSH6 interaction with BER components aids in the suppression of HR has not been investigated. The fact that NTHL1 is also a top candidate for promoting alt-NHEJ, suggests that BER and MMR could potentially influence HR functions.

A distinct subset of BER proteins is coordinated with various components of NER. One example is the interaction between NTHL1 and XPG. As previously noted, recent studies demonstrate that XPG is indispensable for proper HR, and that this function is independent of the NER functions of XPG [110]. In turn, XPG is also crucial for catalytic turnover of the NTHL1 glycosylase [40]. Thus, protein dysregulation of any of these DNA repair components has the potential to affect more than one DNA repair pathway. For example, dysregulation of NTHL1 could ultimately affect the efficiency of HR through unregulated interactions with XPG. Examination at the protein level will provide a starting point for investigating the overall impact of DNA repair protein dysregulation and pathway crosstalk.

3.3 Underlying mechanisms increasing steady-state NTHL1 protein

Examining the potential impact of BER protein levels and the corresponding effect on DNA repair pathway crosstalk, our study begs the question of how the *NTHL1* gene can become dysregulated to yield NTHL1 overexpression. We detected different steady-state levels of NTHL1 in the cell lines we examined showing that NTHL1 levels can vary among NSCLC cell lines. Such overexpression could result from genomic amplification, or from changes at multiple levels of gene expression. Indeed, even with our small sample of cell lines, we found that regulation could occur at multiple levels (**Figure 2.1**). In addition to gene amplification, expression of NTHL1 protein can be altered including regulation of transcription, RNA stability, translation, and/or protein stability. These mechanisms are not mutually exclusive.

In light of the *NTHL1* gene amplification found in multiple cBioPortal datasets, mechanisms that generate *NTHL1* amplification should be explored in somatic tumor formation. Most obvious is aneuploidy through whole chromosome copy number gains, typically due to misegregation of chromosomes during mitosis [152]. Alternatively, focal amplifications, such as in the case of the *MYCN* gene*,* can occur from the formation of anaphase bridges that break unevenly, leading to one of the daughter cells containing additional copies of certain genes [132]. In the case of *MYC*, these amplifications can be extruded into small, circular chromosomes termed double minutes [152]. Amplified genes, such as double minutes, or even whole chromosomes can be expelled by the cell as a micronucleus in an attempt to maintain genetic homeostasis [132]. Additional amplifications can be generated from extruded micronuclei that are then recaptured by a

daughter cell [132]. Thus, the contents of a micronucleus can be reintroduced to the cell resulting in altered gene dosage [132, 153]. How *NTHL1* gene amplification is stochastically generated is unknown but is likely to occur in conjunction with other destabilizing events such as deregulation of the cell cycle and loss/mutation of p53 protein.

Outside of alterations at the gene level, regulation of *NTHL1* expression could occur at multiple levels of gene expression. One possible mode of regulation is via transcription. Interestingly, the promoter of *NTHL1* contains a putative ETS transcription factor consensus motif [45]. As the ETS family of transcription factors is involved in promoting all stages of tumorigenesis [154], dysregulation of ETS is an attractive candidate to explain increases in *NTHL1* transcript levels. ETS dysregulation occurs through gene amplification, novel gain-of-function transcription factor complexes, and ETS protein stabilization through PTMs [154]. The exact ETS family member and definitive evidence for ETS regulation of the *NTHL1* promoter is lacking. Future studies could address whether dysregulation of *NTHL1* can be added to the repertoire of ETS transformation potential.

Furthermore, once an *NTHL1* mRNA transcript is created, RNA stability can affect how much protein is produced. Stabilization of RNA can occur through processes that include but are not limited to RNA binding proteins and the presence or absence of miRNA [155]. Again, no data exist that describe potential *NTHL1* mRNA binding proteins or *NTHL1* targeting miRNA. The field of post-transcriptional regulation for DNA repair

genes, especially BER, remains relatively unexplored, even though RNA binding proteins are emerging as contributors to tumorigenesis, including invasion and metastasis [156]. Importantly, RNA binding proteins that regulate mRNA for DNA repair genes may be another unexplored area for targeted cancer therapy options in the future.

Control of translational processes also impact the steady-state protein levels. One method of translational regulation is through established oncogenic signaling. Classic examples include RAS and MAP kinase, PI3 kinase, loss of mTOR, and the translational protein, eIF-4E [157-159]. For a comprehensive review on how oncogenes impact steady-state protein levels through translational regulation, please refer to [157-159]. The RAS and MAP kinase signaling pathways influence translation by increasing the rates of initiation and elongation, while also simultaneously eliciting ribosome biogenesis. Curiously, RAS signaling can selectively recruit oncogenic RNA transcripts to polysomes to efficiently produce an increase of specific oncogenic proteins [158]. One means of controlling translation via oncogenic signaling is through the phosphorylation or overexpression of the translation initiating factor, eIF-4E [158, 159]. In fact, perturbation of eIF-4E activity is sufficient to cause cellular transformation in rodent fibroblasts [158, 159]. Experiments to assess the impact of translational control of NTHL1 will yield insight into the mechanism of increased NTHL1 steady-state protein levels.

Finally, protein stability can also regulate steady-state NTHL1 levels. One mechanism by which NTHL1 protein turnover could be impacted is via post-translational modifications (PTMs). For instance, SUMOylation protects proteins from ubiquitylation and

subsequent protein degradation, as SUMO and ubiquitin often compete for modifying the same lysine residues [160]. Our group recently mapped SUMOylation of the budding yeast ortholog to NTHL1, Ntg1 [53]. Additionally, we demonstrate that NTHL1 is also SUMO modified, although the lysine residue(s) that are modified are unknown. Our study suggests that SUMO modification of NTHL1 is evolutionarily conserved between the *S. cerevisiae* and mammalian NTHL1 glycosylases. SUMO modification of Ntg1 is proposed to function in initiating a cell cycle checkpoint between G_2 and M phases [53]. However, whether SUMOylation impacts the protein half-life of Ntg1 or NTHL1 has not yet been explored. Thus, if the SUMO conjugating pathway is dysregulated, a result could be aberrant SUMOylation of NTHL1. If SUMO modification plays a protective role, NTHL1 may not be degraded resulting in an apparent increase of steady-state NTHL1 protein levels. Future studies into the exact biological role this SUMO modification has on NTHL1 will need to be performed to assess whether PTMs can be added to the suite of NTHL1 regulatory elements.

3.4 BER glycosylase overexpression

Given the many ways that BER proteins can become dysregulated, examining other BER glycosylases and investigating how these BER components are also dysregulated is important. Indeed, perturbations of BER glycosylases are annotated in several cancer types. For instance, at least 30 truncating mutations and 30 different missense mutations of the *MUTYH* glycosylase can be found in familial colon cancers, all resulting in the inability to excise 8-oxoG:A mispairs [161]. Another example is the overexpression of the MPG glycosylase in gliobastoma [67, 68]. Patients with high expression of MPG

have poorer patient outcomes than patients that do not display elevated levels of MPG. MPG overexpression also leads to cellular resistance to the alkylating agent, temozolomide in glioblastoma [67, 68]. Further, SMUG1 glycosylase overexpression is correlated with poor disease-free survival in gastric cancer patients [70]. Our results expand the scope of knowledge to include an increase in steady-state NTHL1 protein as a potential contributing factor to generating genome instability and markers of cellular transformation.

Overexpression of other BER glycosylases are linked to various tumor types with clinical implications. High expression of the SMUG1 glycosylase, which repairs uracil in single strand DNA [162], is correlated with poor disease-free patient survival in gastric cancer patients [70]. Immunohistochemistry on tissue microarrays from patient biopsy cores was used for SMUG1 in this analysis, and results from this study investigated patient responses to surgery [70]. Overexpression of the OGG1 or NTHL1 glycosylases in lymphoblastoid cells sensitizes these cells to ionizing radiation [69]. This study led to a model where glycosylase overexpression exacerbates DSB formation through BER processing of closely opposed DNA base damage, a model which is dependent on the enzymatic activity of NTHL1 or OGG1 [69]. Our results show that overexpression of NTHL1 protein can trigger genomic instability through DSB accumulation that can induce cellular transformation. This mechanism does not depend on NTHL1 enzymatic activity, in contrast to the aforementioned study.

In addition to this focused study of NTHL1, we analyzed other BER glycosylases in cBioPortal and found that several additional glycosylases including *MPG*, *NEIL1*, and *OGG1* are amplified across various cancer types. These data suggest that overexpression of additional glycosylases could play a role in cellular transformation. Completion of these studies would expand the scope of knowledge of BER dysregulation in a cellular context.

3.5 Unanswered questions and predictions

3.5.1 Unanswered Questions

Our work establishes a new paradigm for BER glycosylase regulation and cellular transformation. Overexpression of NTHL1 is sufficient to initiate the transformation process through the accumulation of DSBs via inhibition of HR. Furthermore, cellular transformation is not dependent on NTHL1 enzymatic activity. This raises several unanswered questions: 1) whether NTHL1 is appropriate to classify as an oncogene; 2) whether alt-NHEJ compensates for disruption of HR; 3) whether the cells transformed by NTHL1 overexpression have the potential to form tumors; and 4) the potential clinical implications of NTHL1 overexpression.

1) NTHL1 as an oncogene. As NTHL1 overexpression is found in NSCLC cell lines and our study provides evidence that NTHL1 overexpression has potential to cause cellular transformation, we propose that NTHL1 can act as an oncogene upon overexpression. Our study provides the rationale needed to move into a mouse model of human NTHL1 overexpression to investigate the true oncogenic potential of NTHL1. Mice that

constitutively overexpress NTHL1 from the β-actin promoter would provide a physiologically relevant model to study the effects of NTHL1 perturbation. Alternatively, NTHL1 overexpression can be studied and controlled in a tissue dependent manner if viability proves to be an issue with global NTHL1 overexpression. NTHL1 overexpression phenotypes would be interesting to examine in lung, mammary, pancreatic, and prostate tissues, as cBioPortal shows *NTHL1* amplification in each of these respective cancer types. Necropsy of these animals and histology on any resulting tumors will provide insight into NTHL1 protein regulation, localization, and whether NTHL1 overexpression impacts the tumorigenic process in specific organ tissues.

2) NHEJ compensation. As we observed a preponderance of DNA damage and DSBs that were induced in a manner that was independent from NTHL1 enzymatic activity, coupled with a decrease of HR repair activity by ~50%, future studies will also aim to determine whether alt-NHEJ is promoted by HR inhibition via NTHL1 overexpression. Additional analysis of cytogenetic results may provide insight into the mechanism(s) of NTHL1-driven genomic instability based on the type of prevalent chromosomal aberrations observed. Sequencing studies could also determine if a DNA repair pathway signature such as alt-NHEJ microhomology occurs in NTHL1 overexpressing cells. However, completion of these experiments is dependent on the generation of stable NTHL1-inducible cell lines, which is currently underway in our laboratory.

3) Potential for NTHL1 to drive tumorigenesis. Our initial experiments describe early steps in the cellular transformation process via anchorage independent growth, loss of

contact inhibition, and persistent genomic instability (**Figure 2.9**). However, cellular transformation encompasses a number of criteria that cancers commonly display termed cancer hallmarks [142]. While a cell may attain one or a few cancer hallmarks, this is not a guarantee that the cell has acquired the capacity to establish a mature tumor in a biological setting. Therefore, we would like to assess the full transformation potential of the soft agar derived clonal cell lines in mice xenograft experiments. Results from this study would help determine the true oncogenic potential of NTHL1 overexpression.

4) Clinical implications of NTHL1 overexpression. A prediction of our model is that cancer cells overexpressing NTHL1 should be sensitized to cancer treatment modalities that generate DSBs such as ionizing radiation (IR) [163]. This prediction stems from our finding that NTHL1 overexpression results in decreased HR repair activity (**Figure 2.4**), and further overwhelming the cell with additional DSBs would result in cell death. Determining whether cancer cell sensitization to chemotherapeutics that induced DSBs can be achieved from NTHL1 overexpression could lead to clinical application of NTHL1 as a biomarker for specific subsets of NSCLC patients. Subdividing the NSCLC cell lines used in our study reveals that large cell and adenocarcinoma NSCLC subtypes are the predominate class of NSCLC cells that display increases in NTHL1 expression. The exception to this observation is the low NTHL1 expression found in the H1299 carcinoma line. In contrast, the H226 squamous cell line displays decreased levels of NTHL1, suggesting that decreased NTHL1 levels could be specific to squamous NSCLC. Investigation of a wider panel of squamous NSCLC cell lines would provide a more comprehensive examination of NTHL1 protein levels in this NSCLC subtype. Our initial

observation that NTHL1 is overexpressed involves the analysis of cell lines. Extension of our study into primary tumors is warranted to determine whether these phenotypes are recapitulated. Interestingly, breast, pancreas, and prostate cancer datasets in cBioPortal also display *NTHL1* amplification. Whether our observations can be extended into other cancer types is an area for future investigation. Extension of our studies for other BER glycosylases that are highly amplified in cancer datasets is also an area for future investigation to determine if genomic instability and HR inhibition is a general consequence of glycosylase overexpression.

3.5.2 Future Clinical Implications

Perhaps the best-characterized BER glycosylase with respect to impact of BER protein overexpression on chemotherapeutic sensitivities is the monofunctional MPG glycosylase. In human tumors, MPG repairs certain types of alkylation damage induced by temozolomide [164]. In cancers where overexpression of MPG confers sensitivity, the mechanism is attributed to the accumulation of abasic sites as MPG excises alkylated DNA bases [165]. On the contrary, glioblastomas that overexpress MPG show resistance to temozolomide, and MPG expression correlates with negative patient survival outcomes [67, 68]. In order to explain this discrepancy, one must also examine the expression levels of downstream BER components, such as APEX1 and DNA polymerase β (POLB). In a tissue culture system, complementation of a downstream BER protein, POLB, is required for cellular resistance to temozolomide in cells that also overexpress MPG. In this case, temozolomide resistance is due to POLB processing and repair of the accumulated AP site intermediates in order to complete the BER pathway [164, 166].

Analogously, overexpression of APEX1 in a glioblastoma cell culture system also overexpressing MPG confers resistance to temozolomide, in a similar mechanism to DNA polymerase β overexpression [167]. These results suggest that for temozolomide sensitivity, MPG must be overexpressed to generate BER intermediates, while downstream BER protein expression levels remain unchanged. Resistance to temozolomide on the other hand may be mediated by the concurrent overexpression of MPG and/or APEX1 and POLB. Whether this model for chemotherapeutic sensitivities is widely applicable to other BER glycosylases and other BER protein components is unknown.

This above concept is substantiated by our mining of cBioPortal cancer genomic datasets upon search for concurrent gene amplification of *MPG* with either *APEX1* or *POLB* [118]. *MPG*, *APEX1*, and *POLB* are highly amplified in a neuroendocrine prostate cancer dataset [168]. Mining of this prostate cancer dataset demonstrates that *MPG* and *APEX1* amplification have a tendency towards co-occurrence, and that this co-occurrence is statistically significant ($P<0.001$, log odds ratio 2.637) [168]. The same is true of *MPG* and *POLB* amplification co-occurrence ($P<0.001$, log odds ratio 2.399) [168]. Survival outcomes and responses to chemotherapeutics were not given for this study. In the TCGA glioblastoma study in cBioPortal [169], amplification of *MPG* tends towards co-occurrence with amplification of *APEX1* ($P<0.001$, log odds ratio >3), and overexpression of *POLB* ($P=0.001$, log odds ratio >3) [169]. Thus, examining a panel of BER steady state protein levels instead of focusing solely on *MPG* alone may be a better prognostic indicator for tumor responses to alkylating agents.

Contrary to the accumulation of BER intermediates induced by overexpression of MPG, overexpression of NTHL1 may leave a tumor vulnerable because these cells display an increase in DSBs and a reduction in HR activity. For instance, tumor cells overexpressing NTHL1 already have an increased baseline level of DSBs. Treatments that further increase the number of DSBs, such as radiation therapy, would cause cell death, specifically in cells overexpressing NTHL1 because of HR inhibition. Additional experiments will need to be performed to determine clinical outcomes based on NTHL1 expression patterns. Whether NTHL1 expression correlates with patient responses should also be examined in the context of the expression patterns of downstream BER components, such as APEX1 and/or POLB. In doing so, physicians may predict which patients would respond most effectively to DSB-inducing treatments through investigation of a panel of BER proteins instead of focusing on singular BER component.

Studies that have characterized and correlated glycosylase expression to radiation sensitivity report conflicting results depending on the tumor type and glycosylase [69, 70, 170-175]. Inactivating mutations in the NEIL1 glycosylase are found in primary tumor samples of gastric cancer from a Japanese population [172, 173]. This inactivation of NEIL1 correlates with increased sensitivity of gastric tumors to ionizing radiation therapy, possibly through lack of repair of cytotoxic and fork stalling DNA base lesions [172, 173]. In lung and renal cancer, OGG1 is down regulated; however, whether loss of OGG1 correlates with radiation sensitivity in these tumors was not determined [170, 171]. In yet another study, the down regulation of the NTHL1 glycosylase by siRNA in lymphoblastoid cells results in increased radiation sensitivity, presumably through lack of

repair of fork stalling DNA base lesions [175]. Results from these studies indicate that unrepaired cytotoxic lesions can directly impact radiation sensitivity in certain tumor types and in cultured cell systems. Studies on whether loss of BER glycosylases impact cancer treatment outcomes with respect to specific treatment modalities have yet to be comprehensively examined.

Whether the decrease or loss of a BER glycosylase at the protein level sensitizes cancer cells to radiation therapy or to various chemotherapeutics could possibly be predicted by the specific DNA base lesion that a glycosylase repairs. For example, NEIL1 repairs cytotoxic base lesions such as thymine glycol, which blocks DNA polymerases [23, 176]. Lesions such as thymine glycol pose an immediate risk to the cell because they delay replication progression and require bypass by certain error-prone polymerases [176]. The NEIL1 deficient cells would therefore be sensitized to agents that generate thymine glycol lesions, such as radiation [176]. DNA damage such as 8-oxoguanine is readily bypassed by polymerases and mismatched nucleotides, such as 8oxoG:A, may establish mutations in the genome [177]. Thus, loss of OGG1 leads to the accumulation of mutations in the genome [177], but whether OGG1 protein levels contribute to sensitivity of cancer treatment modalities, potentially through DNA repair pathway crosstalk, requires further study. Therefore, considering glycosylase steady-state expression and the type of lesion that is repaired may be an efficient method for biomarker optimization and designing targeted treatment modalities for the clinic.

We also propose that by focusing on central interaction hubs, as established in our

analysis of pathway crosstalk (**Figure 3.1**), a directed effort at future drug design and DNA repair protein screening will open avenues previously unexplored for responses to chemotherapeutics. For instance, if BER and/or MMR proteins are dysregulated through altered interactions, posttranslational modifications, or steady-state protein levels that suppress HR, could this scenario impact clinical response to chemotherapeutics? An analogous situation is found in breast cancer patients with a germline mutation in the *BRCA1/2* genes that result in inefficient HR function [178, 179]. As a consequence of decreased HR function, cells are sensitive to PARP inhibitors as PARP1 and BRCA1 are synthetically lethal [180, 181]. This lethality occurs because of the intersection of single strand break and DSB repair. By inhibiting PARP, single strand breaks accumulate in the genome, and when the replication machinery encounters these breaks, the result is a DSB. As HR is impaired due to the loss of BRCA1, these DSBs cannot be processed and ultimately result in cell death [182]. Conversely, if a patient does not have a *BRCA1/2* mutation, but instead has suppressed HR as a consequence of elevated protein levels in BER and/or MMR, this raises the issue of whether that patient would also be sensitive to PARP inhibitors. Data mining in cBioPortal with the TCGA invasive breast carcinoma dataset reveals that amplification of *NTHL1* is mutually exclusive with BRCA2 dysregulation (P=0.011, log odds ratio -1.145) [183]. This result implies that either *NTHL1* amplification or *BRCA2* mutation/loss is needed to negatively impact HR function in breast cancer, as either protein dysregulation is functionally redundant. Thus, patient tumors could also be screened for specific interactions or protein dysregulation as a potential biomarker for tumor responsiveness to currently employed chemotherapeutics in breast cancers.

As studies reveal additional functional interactions within and among the various DNA repair pathways, the interaction networks displayed in **Figure 3.1** will need to further evolve to accurately reflect this new information. As a consequence, pathway crosstalk through protein-protein interactions may reveal a potential therapeutic avenue to sensitize a tumor previously thought to be unresponsive to certain treatment options. We must expand our knowledge of DNA repair in order to derive a clinical benefit from BER protein expression. Future work needs to be conducted to determine if dysregulated BER can guide clinical treatment options and to explore the value of BER glycosylases as a biomarker for patient outcomes and responses to cancer treatment modalities.

In this thesis, we introduce a new paradigm of BER dysregulation through the overexpression of the NTHL1 glycosylase. Several markers of cellular transformation are gained through the inhibition of DSB repair including genomic instability, anchorage independent growth, and the loss of contact inhibition. We also favor the idea of determining the oncogenic potential of NTHL1 upon overexpression. From our data, we suggest that examining the protein level of BER glycosylases, in addition to RNA transcript levels and gene mutations will allow for a more comprehensive analysis of DNA repair dysregulation in contributing to cancer. Finally, further exploring how pathway crosstalk effects DNA damage repair could open new treatment strategies for clinical applications.

Chapter 4: References

1. Bauer, N.C., A.H. Corbett, and P.W. Doetsch, *The current state of eukaryotic DNA base damage and repair.* Nucleic Acids Res, 2015. **43**(21): p. 10083-101.
2. Brand, M.D., *The proton leak across the mitochondrial inner membrane.* Biochim Biophys Acta, 1990. **1018**(2-3): p. 128-33.
3. Nicholls, D.G., *The influence of respiration and ATP hydrolysis on the proton-electrochemical gradient across the inner membrane of rat-liver mitochondria as determined by ion distribution.* Eur J Biochem, 1974. **50**(1): p. 305-15.
4. Fahl, W.E., et al., *DNA damage related to increased hydrogen peroxide generation by hypolipidemic drug-induced liver peroxisomes.* Proc Natl Acad Sci U S A, 1984. **81**(24): p. 7827-30.
5. Lambeth, J.D., *NOX enzymes and the biology of reactive oxygen.* Nat Rev Immunol, 2004. **4**(3): p. 181-9.
6. Babior, B.M., *The respiratory burst of phagocytes.* J Clin Invest, 1984. **73**(3): p. 599-601.
7. Wheeler, M.D., *Endotoxin and Kupffer cell activation in alcoholic liver disease.* Alcohol Res Health, 2003. **27**(4): p. 300-6.
8. Beckman, K.B. and B.N. Ames, *Oxidative decay of DNA.* J Biol Chem, 1997. **272**(32): p. 19633-6.
9. Cooke, M.S., et al., *Oxidative DNA damage: mechanisms, mutation, and disease.* FASEB J, 2003. **17**(10): p. 1195-214.
10. Krwawicz, J., et al., *Bacterial DNA repair genes and their eukaryotic homologues: 1. Mutations in genes involved in base excision repair (BER) and*

DNA-end processors and their implication in mutagenesis and human disease. Acta Biochim Pol, 2007. **54**(3): p. 413-34.

11. Saxowsky, T.T. and P.W. Doetsch, *RNA polymerase encounters with DNA damage: transcription-coupled repair or transcriptional mutagenesis?* Chem Rev, 2006. **106**(2): p. 474-88.
12. Valko, M., et al., *Free radicals and antioxidants in normal physiological functions and human disease.* Int J Biochem Cell Biol, 2007. **39**(1): p. 44-84.
13. Alexandrov, L.B., et al., *Signatures of mutational processes in human cancer.* Nature, 2013. **500**(7463): p. 415-21.
14. Nik-Zainal, S., et al., *Mutational processes molding the genomes of 21 breast cancers.* Cell, 2012. **149**(5): p. 979-93.
15. Swartzlander, D.B., et al., *Regulation of base excision repair: Ntg1 nuclear and mitochondrial dynamic localization in response to genotoxic stress.* Nucleic Acids Res, 2010. **38**(12): p. 3963-74.
16. Marullo, R., et al., *Cisplatin induces a mitochondrial-ROS response that contributes to cytotoxicity depending on mitochondrial redox status and bioenergetic functions.* PLoS One, 2013. **8**(11): p. e81162.
17. Shuck, S.C., E.A. Short, and J.J. Turchi, *Eukaryotic nucleotide excision repair: from understanding mechanisms to influencing biology.* Cell Res, 2008. **18**(1): p. 64-72.
18. Hoeijmakers, J.H., *Nucleotide excision repair. II: From yeast to mammals.* Trends Genet, 1993. **9**(6): p. 211-7.

19. Drablos, F., et al., *Alkylation damage in DNA and RNA--repair mechanisms and medical significance.* DNA Repair (Amst), 2004. **3**(11): p. 1389-407.

20. Tchounwou, P.B., et al., *Heavy metal toxicity and the environment.* EXS, 2012. **101**: p. 133-64.

21. Riley, P.A., *Free radicals in biology: oxidative stress and the effects of ionizing radiation.* Int J Radiat Biol, 1994. **65**(1): p. 27-33.

22. Maynard, S., et al., *Base excision repair of oxidative DNA damage and association with cancer and aging.* Carcinogenesis, 2009. **30**(1): p. 2-10.

23. Kim, Y.J. and D.M. Wilson, 3rd, *Overview of base excision repair biochemistry.* Curr Mol Pharmacol, 2012. **5**(1): p. 3-13.

24. Lari, S.U., et al., *Quantitative determination of uracil residues in Escherichia coli DNA: Contribution of ung, dug, and dut genes to uracil avoidance.* DNA Repair (Amst), 2006. **5**(12): p. 1407-20.

25. Limpose, K., Corbett, Anita H, and Doetsch, Paul W, *DNA Damage*. 2014, John Wiley & Sons Ltd, Chichester: Encyclopedia of Life Sciences (eLS).

26. Mitra, S., *DNA glycosylases: specificity and mechanisms.* Prog Nucleic Acid Res Mol Biol, 2001. **68**: p. 189-92.

27. Loeb, L.A. and B.D. Preston, *Mutagenesis by apurinic/apyrimidinic sites.* Annu Rev Genet, 1986. **20**: p. 201-30.

28. Boiteux, S. and M. Guillet, *Abasic sites in DNA: repair and biological consequences in Saccharomyces cerevisiae.* DNA Repair (Amst), 2004. **3**(1): p. 1-12.

29. Demple, B. and J.S. Sung, *Molecular and biological roles of Ape1 protein in mammalian base excision repair.* DNA Repair (Amst), 2005. **4**(12): p. 1442-9.

30. Fortini, P. and E. Dogliotti, *Base damage and single-strand break repair: mechanisms and functional significance of short- and long-patch repair subpathways.* DNA Repair (Amst), 2007. **6**(4): p. 398-409.

31. Frosina, G., et al., *Two pathways for base excision repair in mammalian cells.* J Biol Chem, 1996. **271**(16): p. 9573-8.

32. Robertson, A.B., et al., *DNA repair in mammalian cells: Base excision repair: the long and short of it.* Cell Mol Life Sci, 2009. **66**(6): p. 981-93.

33. Zharkov, D.O., *Base excision DNA repair.* Cell Mol Life Sci, 2008. **65**(10): p. 1544-65.

34. Klungland, A. and T. Lindahl, *Second pathway for completion of human DNA base excision-repair: reconstitution with purified proteins and requirement for DNase IV (FEN1).* EMBO J, 1997. **16**(11): p. 3341-8.

35. Petermann, E., M. Ziegler, and S.L. Oei, *ATP-dependent selection between single nucleotide and long patch base excision repair.* DNA Repair (Amst), 2003. **2**(10): p. 1101-14.

36. Fouquerel, E. and R.W. Sobol, *ARTD1 (PARP1) activation and NAD(+) in DNA repair and cell death.* DNA Repair (Amst), 2014. **23**: p. 27-32.

37. Wright, R.H., et al., *ADP-ribose-derived nuclear ATP synthesis by NUDIX5 is required for chromatin remodeling.* Science, 2016. **352**(6290): p. 1221-5.

38. Horton, J.K., et al., *Hypersensitivity of DNA polymerase beta null mouse fibroblasts reflects accumulation of cytotoxic repair intermediates from site-specific alkyl DNA lesions.* DNA Repair (Amst), 2003. **2**(1): p. 27-48.

39. Sobol, R.W., et al., *The lyase activity of the DNA repair protein beta-polymerase protects from DNA-damage-induced cytotoxicity.* Nature, 2000. **405**(6788): p. 807-10.

40. Klungland, A., et al., *Base excision repair of oxidative DNA damage activated by XPG protein.* Mol Cell, 1999. **3**(1): p. 33-42.

41. Shimizu, Y., et al., *Xeroderma pigmentosum group C protein interacts physically and functionally with thymine DNA glycosylase.* EMBO J, 2003. **22**(1): p. 164-73.

42. Aamann, M.D., et al., *Cockayne Syndrome group B protein stimulates NEIL2 DNA glycosylase activity.* Mech Ageing Dev, 2014. **135**: p. 1-14.

43. Tuo, J., et al., *Functional crosstalk between hOgg1 and the helicase domain of Cockayne syndrome group B protein.* DNA Repair (Amst), 2002. **1**(11): p. 913-27.

44. Muftuoglu, M., et al., *Cockayne syndrome group B protein stimulates repair of formamidopyrimidines by NEIL1 DNA glycosylase.* J Biol Chem, 2009. **284**(14): p. 9270-9.

45. Imai, K., et al., *Genomic structure and sequence of a human homologue (NTHL1/NTH1) of Escherichia coli endonuclease III with those of the adjacent parts of TSC2 and SLC9A3R2 genes.* Gene, 1998. **222**(2): p. 287-95.

46. Rivera, B., et al., *Biallelic NTHL1 Mutations in a Woman with Multiple Primary Tumors.* N Engl J Med, 2015. **373**(20): p. 1985-6.

47. Weren, R.D., et al., *A germline homozygous mutation in the base-excision repair gene NTHL1 causes adenomatous polyposis and colorectal cancer.* Nat Genet, 2015. **47**(6): p. 668-71.

48. *Identification and characterization of the tuberous sclerosis gene on chromosome 16.* Cell. **75**(7): p. 1305-1315.

49. Strausberg, R.L., et al., *Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.* Proc Natl Acad Sci U S A, 2002. **99**(26): p. 16899-903.

50. Aspinwall, R., et al., *Cloning and characterization of a functional human homolog of Escherichia coli endonuclease III.* Proc Natl Acad Sci U S A, 1997. **94**(1): p. 109-14.

51. Ikeda, S., et al., *Differential intracellular localization of the human and mouse endonuclease III homologs and analysis of the sorting signals.* DNA Repair (Amst), 2002. **1**(10): p. 847-54.

52. Ikeda, S., et al., *Purification and characterization of human NTH1, a homolog of Escherichia coli endonuclease III. Direct identification of Lys-212 as the active nucleophilic residue.* J Biol Chem, 1998. **273**(34): p. 21585-93.

53. Swartzlander, D.B., et al., *Identification of SUMO modification sites in the base excision repair protein, Ntg1*. DNA Repair (Amst), 2016. **48**: p. 51-62.

54. Katafuchi, A., et al., *Differential specificity of human and Escherichia coli endonuclease III and VIII homologues for oxidative base lesions.* J Biol Chem, 2004. **279**(14): p. 14464-71.

55. Luna, L., et al., *Cell-cycle regulation, intracellular sorting and induced overexpression of the human NTH1 DNA glycosylase involved in removal of formamidopyrimidine residues from DNA.* Mutat Res, 2000. **460**(2): p. 95-104.

56. Liu, X. and R. Roy, *Truncation of amino-terminal tail stimulates activity of human endonuclease III (hNTH1).* J Mol Biol, 2002. **321**(2): p. 265-76.

57. Liu, X., S. Choudhury, and R. Roy, *In vitro and in vivo dimerization of human endonuclease III stimulates its activity.* J Biol Chem, 2003. **278**(50): p. 50061-9.

58. Marenstein, D.R., et al., *Stimulation of human endonuclease III by Y box-binding protein 1 (DNA-binding protein B). Interaction between a base excision repair enzyme and a transcription factor.* J Biol Chem, 2001. **276**(24): p. 21242-9.

59. Marenstein, D.R., et al., *Substrate specificity of human endonuclease III (hNTH1). Effect of human APE1 on hNTH1 activity.* J Biol Chem, 2003. **278**(11): p. 9005-12.

60. Takao, M., et al., *Mitochondrial targeting of human DNA glycosylases for repair of oxidative DNA damage.* Nucleic Acids Res, 1998. **26**(12): p. 2917-22.

61. Chan, M.K., et al., *Targeted deletion of the genes encoding NTH1 and NEIL1 DNA N-glycosylases reveals the existence of novel carcinogenic oxidative damage to DNA.* DNA Repair (Amst), 2009. **8**(7): p. 786-94.

62. Alseth, I., et al., *The Saccharomyces cerevisiae homologues of endonuclease III from Escherichia coli, Ntg1 and Ntg2, are both required for efficient repair of spontaneous and induced oxidative DNA damage in yeast.* Mol Cell Biol, 1999. **19**(5): p. 3779-87.

63. Griffiths, L.M., et al., *Dynamic compartmentalization of base excision repair proteins in response to nuclear and mitochondrial oxidative stress.* Mol Cell Biol, 2009. **29**(3): p. 794-807.

64. Goto, M., et al., *Altered expression of the human base excision repair gene NTH1 in gastric cancer.* Carcinogenesis, 2009. **30**(8): p. 1345-52.

65. Koketsu, S., T. Watanabe, and H. Nagawa, *Expression of DNA repair protein: MYH, NTH1, and MTH1 in colorectal cancer.* Hepatogastroenterology, 2004. **51**(57): p. 638-42.

66. Albertson, D.G., *Gene amplification in cancer.* Trends Genet, 2006. **22**(8): p. 447-55.

67. Agnihotri, S., et al., *Alkylpurine-DNA-N-glycosylase confers resistance to temozolomide in xenograft models of glioblastoma multiforme and is associated with poor survival in patients.* J Clin Invest, 2012. **122**(1): p. 253-66.

68. Bobola, M.S., et al., *Repair of 3-methyladenine and abasic sites by base excision repair mediates glioblastoma resistance to temozolomide.* Front Oncol, 2012. **2**: p. 176.

69. Yang, N., H. Galick, and S.S. Wallace, *Attempted base excision repair of ionizing radiation damage in human lymphoblastoid cells produces lethal and mutagenic double strand breaks.* DNA Repair (Amst), 2004. **3**(10): p. 1323-34.

70. Abdel-Fatah, T., et al., *Are DNA repair factors promising biomarkers for personalized therapy in gastric cancer?* Antioxid Redox Signal, 2013. **18**(18): p. 2392-8.

71. Ciccia, A. and S.J. Elledge, *The DNA damage response: making it safe to play with knives.* Mol Cell, 2010. **40**(2): p. 179-204.

72. Sinha, R.P. and D.P. Hader, *UV-induced DNA damage and repair: a review.* Photochem Photobiol Sci, 2002. **1**(4): p. 225-36.

73. Reardon, J.T., et al., *In vitro repair of oxidative DNA damage by human nucleotide excision repair system: possible explanation for neurodegeneration in xeroderma pigmentosum patients.* Proc Natl Acad Sci U S A, 1997. **94**(17): p. 9463-8.

74. Kovtun, I.V. and C.T. McMurray, *Crosstalk of DNA glycosylases with pathways other than base excision repair.* DNA Repair (Amst), 2007. **6**(4): p. 517-29.

75. Hardeland, U., et al., *Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis.* J Biol Chem, 2000. **275**(43): p. 33449-56.

76. Hanawalt, P.C. and G. Spivak, *Transcription-coupled DNA repair: two decades of progress and surprises.* Nat Rev Mol Cell Biol, 2008. **9**(12): p. 958-70.

77. Waters, T.R., et al., *Human thymine DNA glycosylase binds to apurinic sites in DNA but is displaced by human apurinic endonuclease 1.* J Biol Chem, 1999. **274**(1): p. 67-74.

78. Menoni, H., J.H. Hoeijmakers, and W. Vermeulen, *Nucleotide excision repair-initiating proteins bind to oxidative DNA lesions in vivo.* J Cell Biol, 2012. **199**(7): p. 1037-46.

79. Boulon, S., et al., *The nucleolus under stress.* Mol Cell, 2010. **40**(2): p. 216-27.

80. Fusser, M., et al., *Spontaneous mutagenesis in Csb(m/m)Ogg1(-)(/)(-) mice is attenuated by dietary resveratrol.* Carcinogenesis, 2011. **32**(1): p. 80-5.

81. Khobta, A., et al., *8-Oxoguanine DNA glycosylase (Ogg1) causes a transcriptional inactivation of damaged DNA in the absence of functional Cockayne syndrome B (Csb) protein.* DNA Repair (Amst), 2009. **8**(3): p. 309-17.

82. Tuo, J., et al., *Primary fibroblasts of Cockayne syndrome patients are defective in cellular repair of 8-hydroxyguanine and 8-hydroxyadenine resulting from oxidative stress.* FASEB J, 2003. **17**(6): p. 668-74.

83. Shell, S.M., et al., *Xeroderma pigmentosum complementation group C protein (XPC) serves as a general sensor of damaged DNA.* DNA Repair (Amst), 2013. **12**(11): p. 947-53.

84. Aamann, M.D., et al., *Multiple interaction partners for Cockayne syndrome proteins: implications for genome and transcriptome maintenance.* Mech Ageing Dev, 2013. **134**(5-6): p. 212-24.

85. Dou, H., S. Mitra, and T.K. Hazra, *Repair of oxidized bases in DNA bubble structures by human DNA glycosylases NEIL1 and NEIL2.* J Biol Chem, 2003. **278**(50): p. 49679-84.

86. Evans, M.D., M. Dizdaroglu, and M.S. Cooke, *Oxidative DNA damage and disease: induction, repair and significance.* Mutat Res, 2004. **567**(1): p. 1-61.

87. Cooper, P.K., et al., *Defective transcription-coupled repair of oxidative base damage in Cockayne syndrome patients from XP group G.* Science, 1997. **275**(5302): p. 990-3.

88. Chovanec, M., et al., *Management of stage I testicular germ cell tumours.* Nat Rev Urol, 2016. **13**(11): p. 663-673.

89. Group, T.I.A.L.C.T.C., *Cisplatin-Based Adjuvant Chemotherapy in Patients with Completely Resected Non-Small-Cell Lung Cancer.* New England Journal of Medicine, 2004. **350**: p. 351-360.

90. Rose, P.G., et al., *Concurrent cisplatin-based radiotherapy and chemotherapy for locally advanced cervical cancer.* N Engl J Med, 1999. **340**(15): p. 1144-53.

91. McGuire, W.P., et al., *Cyclophosphamide and cisplatin compared with paclitaxel and cisplatin in patients with stage III and stage IV ovarian cancer.* N Engl J Med, 1996. **334**(1): p. 1-6.

92. Jamieson, E.R. and S.J. Lippard, *Structure, Recognition, and Processing of Cisplatin-DNA Adducts.* Chem Rev, 1999. **99**(9): p. 2467-98.

93. Fichtinger-Schepman, A.M., et al., *Adducts of the antitumor drug cis-diamminedichloroplatinum(II) with DNA: formation, identification, and quantitation.* Biochemistry, 1985. **24**(3): p. 707-13.

94. Quasthoff, S. and H.P. Hartung, *Chemotherapy-induced peripheral neuropathy.* J Neurol, 2002. **249**(1): p. 9-17.

95. Kim, H.S., et al., *APE1, the DNA base excision repair protein, regulates the removal of platinum adducts in sensory neuronal cultures by NER.* Mutat Res, 2015. **779**: p. 96-104.

96. Kelley, M.R., et al., *Role of the DNA base excision repair protein, APE1 in cisplatin, oxaliplatin, or carboplatin induced sensory neuropathy.* PLoS One, 2014. **9**(9): p. e106485.

97. Damia, G., et al., *Sensitivity of CHO mutant cell lines with specific defects in nucleotide excision repair to different anti-cancer agents.* Int J Cancer, 1996. **66**(6): p. 779-83.

98. Sancar, A., et al., *Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints.* Annu Rev Biochem, 2004. **73**: p. 39-85.

99. Jiang, Y., et al., *Role of APE1 in differentiated neuroblastoma SH-SY5Y cells in response to oxidative stress: use of APE1 small molecule inhibitors to delineate APE1 functions.* DNA Repair (Amst), 2009. **8**(11): p. 1273-82.

100. Wong, H.K., et al., *Cockayne syndrome B protein stimulates apurinic endonuclease 1 activity and protects against agents that introduce base excision repair intermediates.* Nucleic Acids Res, 2007. **35**(12): p. 4103-13.

101. Licht, C.L., T. Stevnsner, and V.A. Bohr, *Cockayne syndrome group B cellular and biochemical functions.* Am J Hum Genet, 2003. **73**(6): p. 1217-39.

102. Tuo, J., et al., *The Cockayne Syndrome group B gene product is involved in general genome base excision repair of 8-hydroxyguanine in DNA.* J Biol Chem, 2001. **276**(49): p. 45772-9.

103. de Waard, H., et al., *Cell type-specific hypersensitivity to oxidative damage in CSB and XPA mice.* DNA Repair (Amst), 2003. **2**(1): p. 13-25.

104. de Waard, H., et al., *Different effects of CSA and CSB deficiency on sensitivity to oxidative DNA damage.* Mol Cell Biol, 2004. **24**(18): p. 7941-8.

105. Kanaar, R., J.H. Hoeijmakers, and D.C. van Gent, *Molecular mechanisms of DNA double strand break repair.* Trends Cell Biol, 1998. **8**(12): p. 483-9.

106. Ceccaldi, R., B. Rondinelli, and A.D. D'Andrea, *Repair Pathway Choices and Consequences at the Double-Strand Break.* Trends Cell Biol, 2016. **26**(1): p. 52-64.

107. Kiraly, O., et al., *DNA glycosylase activity and cell proliferation are key factors in modulating homologous recombination in vivo.* Carcinogenesis, 2014. **35**(11): p. 2495-502.

108. Bunting, S.F. and A. Nussenzweig, *End-joining, translocations and cancer.* Nat Rev Cancer, 2013. **13**(7): p. 443-54.

109. Howard, S.M., D.A. Yanez, and J.M. Stark, *DNA damage response factors from diverse pathways, including DNA crosslink repair, mediate alternative end joining.* PLoS Genet, 2015. **11**(1): p. e1004943.

110. Trego, K.S., et al., *Non-catalytic Roles for XPG with BRCA1 and BRCA2 in Homologous Recombination and Genome Stability.* Mol Cell, 2016. **61**(4): p. 535-46.

111. Prakash, R., et al., *Homologous recombination and human health: the roles of BRCA1, BRCA2, and associated proteins.* Cold Spring Harb Perspect Biol, 2015. **7**(4): p. a016600.

112. de Souza-Pinto, N.C., et al., *The recombination protein RAD52 cooperates with the excision repair protein OGG1 for the repair of oxidative lesions in mammalian cells.* Mol Cell Biol, 2009. **29**(16): p. 4441-54.

113. New, J.H., et al., *Rad52 protein stimulates DNA strand exchange by Rad51 and replication protein A.* Nature, 1998. **391**(6665): p. 407-10.

114. Sugiyama, T., J.H. New, and S.C. Kowalczykowski, *DNA annealing by RAD52 protein is stimulated by specific interaction with the complex of replication protein A and single-stranded DNA.* Proc Natl Acad Sci U S A, 1998. **95**(11): p. 6049-54.

115. Swanson, R.L., et al., *Overlapping specificities of base excision repair, nucleotide excision repair, recombination, and translesion synthesis pathways for DNA base damage in Saccharomyces cerevisiae.* Mol Cell Biol, 1999. **19**(4): p. 2929-35.

116. Fu, D., J.A. Calvo, and L.D. Samson, *Balancing repair and tolerance of DNA damage caused by alkylating agents.* Nat Rev Cancer, 2012. **12**(2): p. 104-20.

117. Cerami, E., et al., *The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data.* Cancer Discov, 2012. **2**(5): p. 401-4.

118. Gao, J., et al., *Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal.* Sci Signal, 2013. **6**(269): p. pl1.

119. Ramirez, R.D., et al., *Immortalization of human bronchial epithelial cells in the absence of viral oncoproteins.* Cancer Res, 2004. **64**(24): p. 9027-34.

120. Werner, E., H. Wang, and P.W. Doetsch, *Opposite roles for p38MAPK-driven responses and reactive oxygen species in the persistence and resolution of radiation-induced genomic instability.* PLoS One, 2014. **9**(10): p. e108234.

121. Pierce, A.J., et al., *XRCC3 promotes homology-directed repair of DNA damage in mammalian cells.* Genes Dev, 1999. **13**(20): p. 2633-8.

122. Leung, S.W., et al., *Splice variants of the human ZC3H14 gene generate multiple isoforms of a zinc finger polyadenosine RNA binding protein.* Gene, 2009. **439**(1-2): p. 71-8.

123. Morreall, J., et al., *Inactivation of a common OGG1 variant by TNF-alpha in mammalian cells.* DNA Repair (Amst), 2015. **26**: p. 15-22.

124. Marullo, R., et al., *HPV16 E6 and E7 proteins induce a chronic oxidative stress response via NOX2 that causes genomic instability and increased susceptibility to DNA damage in head and neck cancer cells.* Carcinogenesis, 2015. **36**(11): p. 1397-406.

125. Wang, J., et al., *GPRC5A suppresses protein synthesis at the endoplasmic reticulum to prevent radiation-induced lung tumorigenesis.* Nat Commun, 2016. **7**: p. 11795.

126. Galick, H.A., et al., *Germ-line variant of human NTH1 DNA glycosylase induces genomic instability and cellular transformation.* Proc Natl Acad Sci U S A, 2013. **110**(35): p. 14314-9.

127. Olive, P.L. and J.P. Banath, *The comet assay: a method to measure DNA damage in individual cells.* Nat Protoc, 2006. **1**(1): p. 23-9.

128. Shaposhnikov, S., E. Frengen, and A.R. Collins, *Increasing the resolution of the comet assay using fluorescent in situ hybridization--a review.* Mutagenesis, 2009. **24**(5): p. 383-9.

129. Ryan, A.J., et al., *Camptothecin cytotoxicity in mammalian cells is associated with the induction of persistent double strand breaks in replicating DNA.* Nucleic Acids Res, 1991. **19**(12): p. 3295-300.

130. Fenech, M., *Cytokinesis-block micronucleus cytome assay.* Nat Protoc, 2007. **2**(5): p. 1084-104.

131. Castagnola, P. and W. Giaretti, *Mutant KRAS, chromosomal instability and prognosis in colorectal cancer.* Biochim Biophys Acta, 2005. **1756**(2): p. 115-25.

132. Fenech, M., et al., *Molecular mechanisms of micronucleus, nucleoplasmic bridge and nuclear bud formation in mammalian and human cells.* Mutagenesis, 2011. **26**(1): p. 125-32.

133. Xu, B., et al., *Replication stress induces micronuclei comprising of aggregated DNA double-strand breaks.* PLoS One, 2011. **6**(4): p. e18618.

134. Palmitelli, M., M. de Campos-Nebel, and M. Gonzalez-Cid, *Progression of chromosomal damage induced by etoposide in G2 phase in a DNA-PKcs-deficient context.* Chromosome Res, 2015. **23**(4): p. 719-32.

135. Liu, S., et al., *ATR autophosphorylation as a molecular switch for checkpoint activation.* Mol Cell, 2011. **43**(2): p. 192-202.

136. Zhao, H. and H. Piwnica-Worms, *ATR-mediated checkpoint pathways regulate phosphorylation and activation of human Chk1.* Mol Cell Biol, 2001. **21**(13): p. 4129-39.

137. Smith, J., et al., *The ATM-Chk2 and ATR-Chk1 pathways in DNA damage signaling and cancer.* Adv Cancer Res, 2010. **108**: p. 73-112.

138. Gelot, C., I. Magdalou, and B.S. Lopez, *Replication stress in Mammalian cells and its consequences for mitosis.* Genes (Basel), 2015. **6**(2): p. 267-98.

139. Ahrabi, S., et al., *A role for human homologous recombination factors in suppressing microhomology-mediated end joining.* Nucleic Acids Res, 2016. **44**(12): p. 5743-57.

140. Shao, R.G., et al., *Replication-mediated DNA damage by camptothecin induces phosphorylation of RPA by DNA-dependent protein kinase and dissociates RPA:DNA-PK complexes.* EMBO J, 1999. **18**(5): p. 1397-406.

141. Zeman, M.K. and K.A. Cimprich, *Causes and consequences of replication stress.* Nat Cell Biol, 2014. **16**(1): p. 2-9.

142. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.

143. Franken, N.A., et al., *Clonogenic assay of cells in vitro.* Nat Protoc, 2006. **1**(5): p. 2315-9.

144. Sweasy, J.B., et al., *Expression of DNA polymerase {beta} cancer-associated variants in mouse cells results in cellular transformation.* Proc Natl Acad Sci U S A, 2005. **102**(40): p. 14350-5.

145. Oyama, M., et al., *Human NTH1 physically interacts with p53 and proliferating cell nuclear antigen.* Biochem Biophys Res Commun, 2004. **321**(1): p. 183-91.

146. O'Donovan, A., et al., *XPG endonuclease makes the 3′ incision in human DNA nucleotide excision repair.* Nature, 1994. **371**(6496): p. 432-5.

147. Grabarz, A., et al., *Initiation of DNA double strand break repair: signaling and single-stranded resection dictate the choice between homologous recombination, non-homologous end-joining and alternative end-joining.* Am J Cancer Res, 2012. **2**(3): p. 249-68.

148. Takata, M., et al., *Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells.* EMBO J, 1998. **17**(18): p. 5497-508.

149. Truong, L.N., et al., *Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells.* Proc Natl Acad Sci U S A, 2013. **110**(19): p. 7720-5.

150. Toledo, L.I., et al., *ATR prohibits replication catastrophe by preventing global exhaustion of RPA.* Cell, 2013. **155**(5): p. 1088-103.

151. Szklarczyk, D., et al., *STRING v10: protein-protein interaction networks, integrated over the tree of life.* Nucleic Acids Res, 2015. **43**(Database issue): p. D447-52.

152. Grade, M., M.J. Difilippantonio, and J. Camps, *Patterns of Chromosomal Aberrations in Solid Tumors.* Recent Results Cancer Res, 2015. **200**: p. 115-42.

153. Rao, X., et al., *Multiple origins of spontaneously arising micronuclei in HeLa cells: direct evidence from long-term live cell imaging.* Mutat Res, 2008. **646**(1-2): p. 41-9.

154. Sizemore, G.M., et al., *The ETS family of oncogenic transcription factors in solid tumours.* Nat Rev Cancer, 2017. **17**(6): p. 337-351.

155. Liu, E.Y., C.P. Cali, and E.B. Lee, *RNA metabolism in neurodegenerative disease.* Dis Model Mech, 2017. **10**(5): p. 509-518.

156. Kudinov, A.E., et al., *Musashi RNA-Binding Proteins as Cancer Drivers and Novel Therapeutic Targets.* Clin Cancer Res, 2017. **23**(9): p. 2143-2153.

157. Clemens, M.J., *Targets and mechanisms for the regulation of translation in malignant transformation.* Oncogene, 2004. **23**(18): p. 3180-8.

158. Holland, E.C., *Regulation of translation and cancer.* Cell Cycle, 2004. **3**(4): p. 452-5.

159. Holland, E.C., et al., *Signaling control of mRNA translation in cancer pathogenesis.* Oncogene, 2004. **23**(18): p. 3138-44.

160. Zhao, J., *Sumoylation regulates diverse biological processes.* Cell Mol Life Sci, 2007. **64**(23): p. 3017-33.

161. Wallace, S.S., D.L. Murphy, and J.B. Sweasy, *Base excision repair and cancer.* Cancer Lett, 2012. **327**(1-2): p. 73-89.

162. Kavli, B., et al., *hUNG2 is the major repair enzyme for removal of uracil from U:A matches, U:G mismatches, and U in single-stranded DNA, with hSMUG1 as a broad specificity backup.* J Biol Chem, 2002. **277**(42): p. 39926-36.

163. Jasin, M. and R. Rothstein, *Repair of strand breaks by homologous recombination.* Cold Spring Harb Perspect Biol, 2013. **5**(11): p. a012740.

164. Tang, J.B., et al., *N-methylpurine DNA glycosylase and DNA polymerase beta modulate BER inhibitor potentiation of glioma cells to temozolomide.* Neuro Oncol, 2011. **13**(5): p. 471-86.

165. Fishel, M.L., et al., *Manipulation of base excision repair to sensitize ovarian cancer cells to alkylating agent temozolomide.* Clin Cancer Res, 2007. **13**(1): p. 260-7.

166. Trivedi, R.N., et al., *Human methyl purine DNA glycosylase and DNA polymerase beta expression collectively predict sensitivity to temozolomide.* Mol Pharmacol, 2008. **74**(2): p. 505-16.

167. Fishel, M.L. and M.R. Kelley, *The DNA base excision repair protein Ape1/Ref-1 as a therapeutic and chemopreventive target.* Mol Aspects Med, 2007. **28**(3-4): p. 375-95.

168. Beltran, H., et al., *Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer.* Nat Med, 2016. **22**(3): p. 298-305.

169. Brennan, C.W., et al., *The somatic genomic landscape of glioblastoma.* Cell, 2013. **155**(2): p. 462-77.

170. Audebert, M., et al., *Alterations of the DNA repair gene OGG1 in human clear cell carcinomas of the kidney.* Cancer Res, 2000. **60**(17): p. 4740-4.

171. Kohno, T., et al., *Genetic polymorphisms and alternative splicing of the hOGG1 gene, that is involved in the repair of 8-hydroxyguanine in damaged DNA.* Oncogene, 1998. **16**(25): p. 3219-25.

172. Rosenquist, T.A., et al., *The novel DNA glycosylase, NEIL1, protects mammalian cells from radiation-mediated cell death.* DNA Repair (Amst), 2003. **2**(5): p. 581-91.

173. Shinmura, K., et al., *Inactivating mutations of the human base excision repair gene NEIL1 in gastric cancer.* Carcinogenesis, 2004. **25**(12): p. 2311-7.

174. Vens, C. and A.C. Begg, *Targeting base excision repair as a sensitization strategy in radiotherapy.* Semin Radiat Oncol, 2010. **20**(4): p. 241-9.

175. Yang, N., M.A. Chaudhry, and S.S. Wallace, *Base excision repair by hNTH1 and hOGG1: a two edged sword in the processing of DNA damage in gamma-irradiated human cells.* DNA Repair (Amst), 2006. **5**(1): p. 43-51.

176. Yoon, J.H., et al., *Error-free replicative bypass of thymine glycol by the combined action of DNA polymerases kappa and zeta in human cells.* Proc Natl Acad Sci U S A, 2010. **107**(32): p. 14116-21.

177. Freudenthal, B.D., W.A. Beard, and S.H. Wilson, *DNA polymerase minor groove interactions modulate mutagenic bypass of a templating 8-oxoguanine lesion.* Nucleic Acids Res, 2013. **41**(3): p. 1848-58.

178. Antoniou, A., et al., *Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies.* Am J Hum Genet, 2003. **72**(5): p. 1117-30.

179. Ford, D., et al., *Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium.* Am J Hum Genet, 1998. **62**(3): p. 676-89.

180. Bryant, H.E., et al., *Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase.* Nature, 2005. **434**(7035): p. 913-7.

181. Farmer, H., et al., *Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy.* Nature, 2005. **434**(7035): p. 917-21.

182. Helleday, T., *The underlying mechanism for the PARP and BRCA synthetic lethality: clearing up the misunderstandings.* Mol Oncol, 2011. **5**(4): p. 387-93.

183. Ciriello, G., et al., *Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer.* Cell, 2015. **163**(2): p. 506-19.