

SINGLE CELL ANALYSIS OF BIOCHEMICAL PHENOTYPES

by

AMANDA LUCINDA RICHER

B.A., Goshen College, 2010

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Molecular Biology Program

2020

This dissertation for the Doctor of Philosophy degree by

Amanda Lucinda Richer

has been approved for the

Molecular Biology Program

by

Heide Ford, Chair

Joshua Black

James DeGregori

Aaron Johnson

Katerina Kechris

Jay Hesselberth, Advisor

Date: August 14, 2020

Richer, Amanda Lucinda (PhD, Molecular Biology Program)

Single cell analysis of biochemical phenotypes

Thesis directed by Associate Professor Jay R. Hesselberth

ABSTRACT

Recent development of high throughput single cell methods has expanded our understanding of tissue heterogeneity, cell states, and developmental biology. Current single cell methods aim to understand cell function and phenotype by measuring a variety of cell features, like DNA sequence, mRNA abundance, chromatin accessibility, cell surface proteins, and histone modifications. However, cell phenotypes are regulated by a number of factors that escape the abundance measurements of current single cell methods.

To directly measure cell phenotypes, I developed a method to measure enzyme activities in single cells. To this end, I designed DNA repair substrates that can be used to measure strand incision events catalyzed by endogenous DNA repair enzymes and called the method “Haircut”. Haircut semi-quantitatively measures base excision repair and ribonucleotide excision repair and has been adapted to work simultaneously in a single cell mRNA sequencing experiment.

Using Haircut, I measured mRNA expression and DNA repair activities in primary human immune cells and found differences in several DNA repair enzyme activities between immune cell types. Some of the repair activity measurements were supported by mRNA abundance measurements in those cell types. While other activities, especially those catalyzed by multimeric proteins, did not correlate with gene expression measurements. Additionally, I used Haircut to measure mRNA expression and DNA repair heterogeneity in immune cells from a preliminary cohort of individuals with trisomy 21 and found no differences in DNA repair between individuals with trisomy 21 and individuals with disomy 21.

In summary, I developed a platform to measure enzymatic activities in single cells. I used Haircut to measure known and previously unknown differences in DNA repair across immune cell types. The platform I developed can be modified to measure other enzymatic activities and mRNA expression in thousands of single cells and can be expanded further to measure many enzymatic activities in millions of cells. Cell phenotypes are regulated by many complex mechanisms and are difficult to predict using DNA sequence or gene expression alone. Single cell analysis of biochemical phenotypes has the potential to bridge the gap between gene expression and cell function and provide direct functional readouts where other single cell methods cannot.

The form and content of this abstract are approved. I recommend its publication.

Approved: Jay R. Hesselberth

DEDICATION

To Eric for the daily love, support, laughs, and adventures

and

to Josh for making the roadmap for me to follow

ACKNOWLEDGEMENTS

It truly takes a village and I will do my best to thank all those who have helped and guided me during this time. First and foremost, I thank my advisor, Jay Hesselberth. I admire Jay's broad scientific curiosity and innovative approach to science. The possibility of working with Jay was one the reasons I attended to the University of Colorado Anschutz. I am grateful to have had the opportunity to work with and learn from Jay. Jay taught me to approach innovation through an "it's going to work" attitude that was extremely helpful in navigating the doctoral training process.

I want to thank my committee members, Heide Ford, Josh Black, James DeGregori, Katerina Kechris, and Aaron Johnson for their continued feedback and thoughtful questions.

The Hesselberth lab has been a wonderful place to work and I am grateful for the enthusiasm, humor, wisdom, and hard work that the lab embodies. My forever rotation buddy Laura White has been a constant presence during my graduate career. I admire Laura's pure curiosity-driven approach to science and her thoughtful approach to graduate school. I hope to someday have the work ethic of Rachel Ancar. Rachel single-handedly juggles several projects, clinicals, and committee duties with enthusiasm and determination. I want to express my extreme fondness and gratitude to both Laura and Rachel. I am grateful for their comradery through the highs and lows of graduate school and I could not have asked for a better set of classmates and labmates. Shannon Walsh has been an invaluable teacher in the lab. I am thankful to have benefited from her expertise. Several of the ideas in this thesis have been reduced to practice by her and I would like to explicitly thank her for the precision medicine figure. Shannon has been a wonderful bay-mate and I appreciated her tolerance for my many interrupting questions throughout the day. Patrick Cherry's methodical and detailed approach to science was an inspiration and I hope I absorbed some of his attention to detail. It would be remiss of me to not recognize Monica Ransom.

Monica Ransom is responsible for all of my training in next-generation sequencing and I am appreciative for all her teachings. Thank you to the “guys downstairs” who have patiently answered any and all my analysis questions. I especially want to thank Kent Riemondy for his contributions to the scrunchy software package and the DNA repair paper.

I want to express my gratitude to all of the administrative staff in the Biochemistry and Molecular Genetics Department, Molecular Biology Program, and the Biomedical Science Program including, Karen Vockrodt, Annie Vazquez, Shelly Fortner, Sabrena Heilman, Caitlin Moloney, and Liz Bowen. They have alleviated any and all non-science related stress in graduate school and I appreciate their help in navigating grant submissions, annual reporting, health insurance, and for all of the other tasks of which I am unaware. I also want to thank the CU Anschutz Genomics Core for the services they provide. In addition to the 271 libraries they sequenced for me, I appreciate their flexibility in using the 10x Genomics system.

My training was supported by several funding sources. My National Science Foundation Graduate Research Fellowship (DGE-1553798) supported me for three years and the Molecular Biology Program T32-INCLUDE (GM008730-20S1) supported me for one year. I thank the Linda Crnic Institute for their role in securing the T32-INCLUDE funding.

Finally, I want to thank all of my family and friends. Graduate school was a time of learning, hard work, and stress, but also a lot of fun. I used to say that each vacation day I took during school just got tacked on to the end, but it was worth it. Thank you for the vacations, happy hours, movie nights, camping, backpacking, climbing, rafting, ski trips, phone calls, tasty meals, game nights, and unconditional love and support. A special thanks to my parents for fostering my curiosity and education from the beginning. Thank you to each and every one of you for providing all the things that matter most to me during this time.

TABLE OF CONTENTS

CHAPTER

I.	INTRODUCTION.....	1
	DNA damage.....	2
	Heterogeneity in DNA repair capacity.....	4
	Base excision repair	6
	Uracil in the genome	11
	Ribonucleotides in the genome	24
	Abasic sites.....	32
	Oxidative DNA damage and repair.....	37
	Alkylation DNA damage and repair.....	50
	Additional DNA repair pathways.....	57
	Methods to measure DNA repair capacity	58
	Single cell methods	64
	Summary	71
II.	SIMULTANEOUS MEASUREMENT OF BIOCHEMICAL PHENOTYPES AND GENE EXPRESSION IN SINGLE CELLS	72
	Abstract	72
	Introduction	72
	Materials and methods	73

	Results and discussion.....	91
III.	SUMMARY AND FUTURE DIRECTIONS	114
	Improvements to Haircut.....	114
	Expanding the single-cell biochemistry toolkit.....	126
	Applications for single-cell biochemistry	132
	REFERENCES.....	143
	APPENDIX	
A.	MEASURING DNA REPAIR ACTIVITIES IN SINGLE CELLS.....	179
	Abstract	179
	Introduction	179
	Materials and methods	183
	Results	188
	Discussion	196
B.	DNA REPAIR HETEROGENEITY IN INDIVIDUALS WITH TRISOMY 21	199
	Abstract	199
	Introduction	199
	Methods.....	201
	Results and discussion.....	203

LIST OF TABLES

Table 1.1: Human glycosylases, their subcellular localization, and their substrates.	8
Table 2.1: Oligonucleotides for single cell DNA repair measurements	84
Table 2.2: DNA repair substrate experimental conditions	85
Table 2.3: Differential gene expression for base excision repair genes.....	86

LIST OF FIGURES

Figure 1.1: DNA damage and repair are diverse and essential for maintaining a stable genome...	3
Figure 1.2: The steps of base excision repair	7
Figure 1.3: Repair of single-base DNA lesions by BER and RER pathways.	16
Figure 1.4: DNA repair methods.....	59
Figure 1.5: Single-cell mRNA sequencing methods.....	66
Table 2.4. Differential repair calculated for biological repair positions.	87
Figure 2.1. Development and validation of a single-cell assay for measuring DNA repair capacity.	90
Figure 2.2: Schematic of single-cell Haircut library preparation.....	93
Figure 2.3: Determining biological DNA repair activity from single-cell Haircut signals.....	94
Figure 2.4. DNA repair measurements determine cell types in a cell mixing experiment.	96
Figure 2.5: Single-cell mRNA expression of RNASEH2C and UNG in Hap1 knockout cells. ...	98
Figure 2.6. Measuring DNA repair in single cells across multiple concentrations and time points.	104
Figure 2.7. Proportion of DNA substrates captured per cell.....	105
Figure 2.8: Analysis of DNA repair heterogeneity in human lymphocytes.....	106
Figure 2.9: Biological replication of DNA repair phenotypes in human PBMCs.	108
Figure 2.10. Gene expression of UNG and RNASEH2 in single cells and cell populations.....	110
Figure 2.11: Cell-type classification using DNA repair measurements.....	112
Figure 3.1: Improving Haircut specificity through translesion synthesis and computationally..	123
Figure 3.2: Improving Haircut sensitivity by removing unrepaired substrates.....	124
Figure 3.3: Combinatorial indexing protocol to measure DNA repair and RNA	125

Figure 3.4: Expanding the single-cell biochemistry toolkit	131
Figure 3.5: Single-cell biochemistry has the potential to contribute to the fields of precision medicine, pharmaceutical sciences, and developmental biology.....	136
Figure A.1: Multiplexed DNA repair assay to measure DNA repair in bulk cell extracts and single cells.....	182
Figure A.2: DNA repair activities are validated in bulk cell extract from knockout cell lines...	189
Figure A.3: Nonspecific signals in Haircut.....	192
Figure A.4: Repeatability and consistency of Haircut signals.	194
Figure A.5: Measuring DNA repair in single cells.	195
Figure B.1: Single cell analysis of primary peripheral blood mononuclear cells in people with trisomy 21.....	205
Figure B.2: Confirming known differences in cell populations and gene expression in individuals with T21 using single-cell RNA seq.	206
Figure B.3: Single-cell DNA repair measurements from PBMCs from individuals with T21...	209

CHAPTER I

INTRODUCTION

Each day cells incur hundreds to millions of DNA damage events that are efficiently and accurately repaired by an elegant set of DNA repair enzymes. Cell metabolism, cell division, and environmental exposures can cause identical alterations to DNA bases that can lead to strand distortions, mismatches, stalled replication forks, and DNA strand breaks. Given the mutagenic potential of many DNA damage events, it is only through highly efficient DNA repair mechanisms that cells are still able to maintain a stable genome with >99.99% accuracy (Nishant, Singh, and Alani 2009). However, this nearly perfect DNA repair system is not static; as it can differ between individuals, cell types, cell cycle phases, and can change with age and upon the onset of disease. Methods that stem from classical biochemical assays to high-throughput sequencing have been used to characterize, in detail, the key enzymatic factors that regulate DNA repair in live cells and in cell extracts. These methods have contributed to a wide body of literature describing the importance of DNA repair in cancer development, treatment and the heterogeneous nature of DNA repair in tumors and healthy individuals. In the following dissertation, I will describe my contribution to the DNA repair field in the development of a multiplexed, high throughput method to measure the activity of DNA repair enzymes in single cells.

Heterogeneity is a cornerstone of biological diversity. New methods to examine cells and tissues at single-cell resolution can measure mRNA, DNA, chromatin accessibility, and proteins at unprecedented resolution and throughput (Stuart and Satija 2019a). These methods have greatly expanded our understanding of cell heterogeneity in complex tissues, cell responses, development, and diseases; however, their reliance on measuring the abundance of molecules makes it difficult to fully understand biochemical phenotypic heterogeneity. Our understanding of phenotypic

heterogeneity through these methods often relies on one key assumption: RNA abundance is directly correlated to protein abundance that is directly correlated to protein activity. While this assumption may be generally true, it ignores several known cellular mechanisms for post transcriptional and post translational regulations that ultimately change protein activity independently of RNA and protein abundance. Directly measuring biochemical activities in single cells is a new modality in which we can measure phenotypic heterogeneity in complex cell populations. In this dissertation I will describe my contribution to the single-cell field in the development of a single-cell method to measure DNA repair enzyme activities within the context of a single-cell mRNA sequencing method.

DNA damage

DNA damage is ubiquitous and common in all cells. Hundreds to millions of DNA damage events can occur each day in human cells from endogenous and exogenous sources (**Figure 1.1**). Deoxyuracil, ribonucleotides, and mismatches can be incorporated by all replicative DNA polymerases into nascent DNA (Goulian, Bleile, and Tseng 1980a; Nick McElhinny et al. 2010; Kunkel, Roberts, and Sugino 1991). Products of cellular metabolism can create oxidative and alkylative lesions and deamination events. Spontaneous decay of DNA bases can lead to many of the DNA damage events that arise in cells (Lindahl 1993). Environmental exposure can also lead to single-base oxidation and alkylation events, bulky adducts, and accelerate the decay of DNA bases. To maintain a stable genome and cellular homeostasis, cells have multiple, and often redundant, DNA repair pathways. The efficiency of these DNA repair pathways leads to only ~50,000 steady-state DNA damage events (Swenberg et al. 2011) and 2-10 mutations per genome per cell division (Nishant, Singh, and Alani 2009).

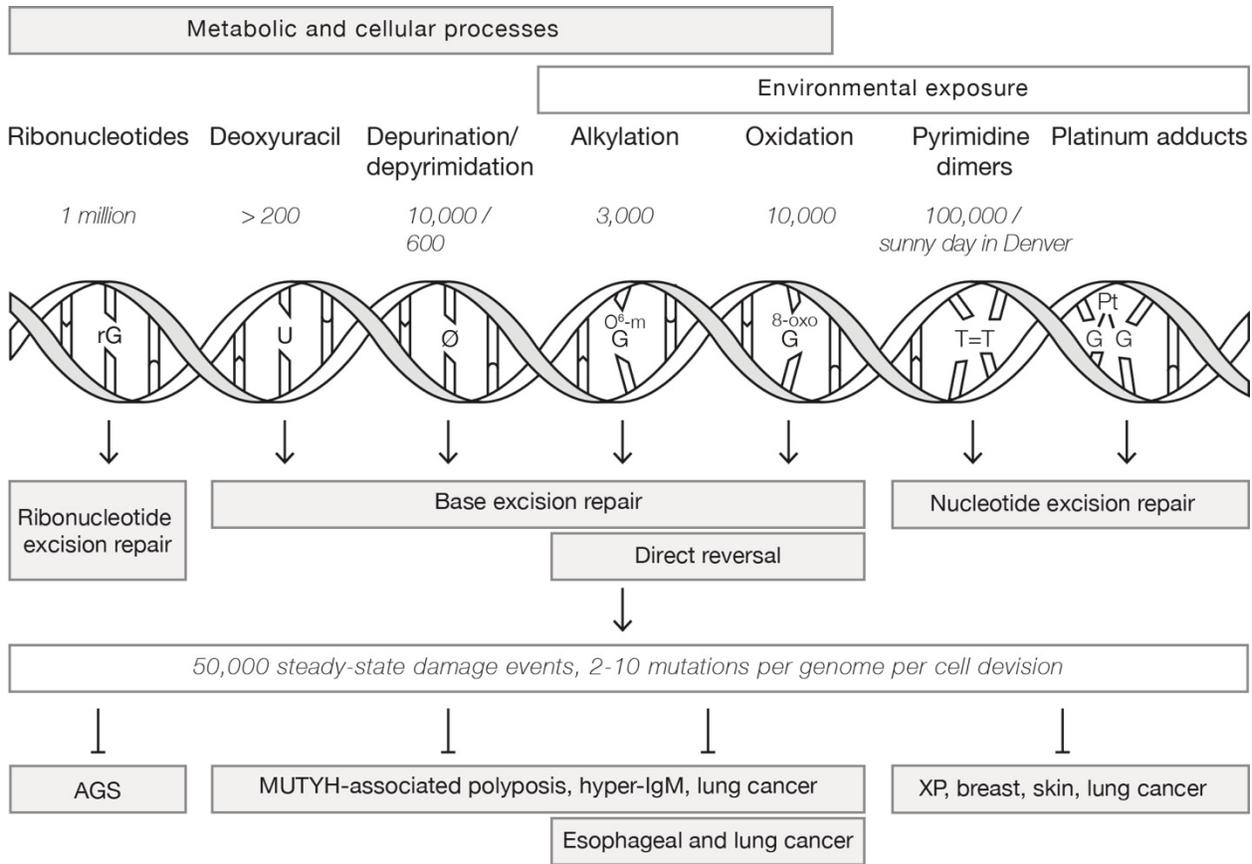


Figure 1.1: DNA damage and repair are diverse and essential for maintaining a stable genome

DNA damage events occur on the order of hundreds to one million events per day and can arise from environmental exposure and metabolic and cellular processes. DNA damage events are removed in a relatively error-free manner by one of several, somewhat overlapping, DNA repair pathways. DNA repair pathways are important to maintain a stable genome and preventing many human disorders and cancers.

Not surprisingly, if DNA repair factors expression or functions are altered, it can lead to a number of human diseases and cancers (Nagel, Chaim, and Samson 2014). When DNA repair factors are less effective it can lead to the development of cancer, however, given that many chemotherapies induce DNA damage, if DNA repair factors are overactive it can lead to chemotherapy resistance in many cancers. In fact, the expression or activity level of DNA repair factors can predict cellular sensitivity to DNA damaging chemotherapies (Nagel et al. 2017; Kitange et al. 2009). Given this double-edge role of DNA repair, it is not surprising that DNA repair is an important biomarker in cancer predisposition and chemotherapy response (Torgovnick and Schumacher 2015).

Heterogeneity in DNA repair capacity

While DNA repair remains an important cellular process, there is a lot of evidence that not all cell types or individuals repair DNA at the same rate. DNA repair capacity, or a cell's ability to repair its DNA from the numerous sources of DNA damage, can change with the cell cycle (Branzei and Foiani 2008), differ between tissue types, and change with age (Pons et al. 2010a), change upon the onset of disease (Nagel, Chaim, and Samson 2014), and differ between individuals (Andersson, Stenqvist, and Hellman 2007; Chaim et al. 2017a).

For example, many DNA repair factors are regulated by the cell cycle in order to couple some DNA repair activities with DNA replication and to prevent other DNA repair activities during DNA replication or mitosis (Orthwein et al. 2014; Mjelle et al. 2015). The expression of several DNA glycosylases is regulated by the cell cycle in addition to replicative DNA polymerases and the flap endonuclease Fen1 that play a role in long-patch base excision repair (BER). These cell cycle expression differences can partially explain why post-replicative cells have lower DNA repair capacity compared to dividing cells (Mansour Akbari et al. 2009).

In addition to cell cycle regulation, DNA repair genes have tissue-specific expression patterns (Uhlén et al. 2015). For example, there are cell type specific differences in DNA repair activities in peripheral blood mononuclear cells in healthy individuals (Amanda L. Richer et al. 2020a). Mutations in DNA repair genes often have tissue-specific phenotypes. For example, inactivating mutations in nucleotide excision repair genes lead to the disorder xeroderma pigmentosum. People with xeroderma pigmentosum have a 2000 fold greater risk of developing skin cancer (Kraemer, Lee, and Scotto 1984). Additionally, mutations in the BER factor MYTH1 cause a predisposition MUTYH-associated polyposis and colorectal cancer (Al-Tassan et al. 2002). Knockout mouse models have found that inactivation of DNA repair factors can lead to tissue-specific tumorigenesis. Other mutations in DNA repair factors can lead to neurological disorders, immunological disorders, and premature aging indicating that there are tissue-specific differences in DNA repair factor activity (Dion 2014).

Numerous studies have found significant inter-individual differences in DNA repair capacity. Some of these differences can be explained by genetic differences, either through single nucleotide polymorphisms or other mutations (Jalal, Earley, and Turchi 2011). Some differences in DNA damage between individuals can be explained by differences in diets, although there is little decisive evidence to support that diet influences DNA repair capacity (Caple et al. 2010; Chang et al. 2010; Slysikova, Lorenzo, et al. 2014). Additionally, normal tissue damage after radiation therapy differs greatly between patients, in part due to differences in DNA repair of radiation-induced DNA damage (Barnett et al. 2015). Sensitive measurement of DNA glycosylase and Ape1 activity in human cells from healthy donors indicate that interindividual variation in DNA repair capacity can be useful in predicting the cell's sensitivity to several DNA damage agents (Chaim et al. 2017a).

There is also evidence that DNA repair capacity is not stable. An imbalance of DNA repair can be a predictor for diseases (Sevilya et al. 2014; Leitner-Dagan et al. 2014; Obtulowicz et al. 2010; S. C. Lee and Chan 2015). DNA repair capacity can also change with age and disease (Pons et al. 2010a; S. Sauvaigo et al. 2007; Forestier, Douki, et al. 2012a). DNA repair is a constant and dynamic process in cells that is important to their homeostasis and function.

Base excision repair

Single-base lesions that do not distort the DNA backbone are repaired by base excision repair (BER) enzymes. BER occurs in four distinct repair steps: (1) The damaged base is excised from the DNA backbone by one of eleven DNA glycosylases. (2) The resulting abasic site is removed by strand incision. (3) The resulting nick is processed by repair synthesis and end-repair, and (4) ligation of the DNA backbone results in a complete repair (Krokan and Bjørås 2013) (**Figure 1.2**).

DNA glycosylases

BER is initiated by one of eleven DNA glycosylases in humans that recognize damaged DNA bases and cleave them from the DNA backbone. DNA glycosylases recognize one or multiple, often chemically similar, damaged bases (**Table 1**) and hydrolyze the bond between the base and the deoxyribose sugar creating an abasic site. DNA glycosylases can initiate repair on single- or double-stranded DNA, however, they will only cleave bases that are in an extra helical position from double-stranded DNA. DNA glycosylases scan the genome for damaged bases either by actively flipping damaged bases out of the major groove (Slupphaug et al. 1996) or by capturing damaged bases as they are transiently expelled from the helix (C. Cao et al. 2004). In addition to their glycosylase activity, bifunctional glycosylases also have AP lyase activity (described below).

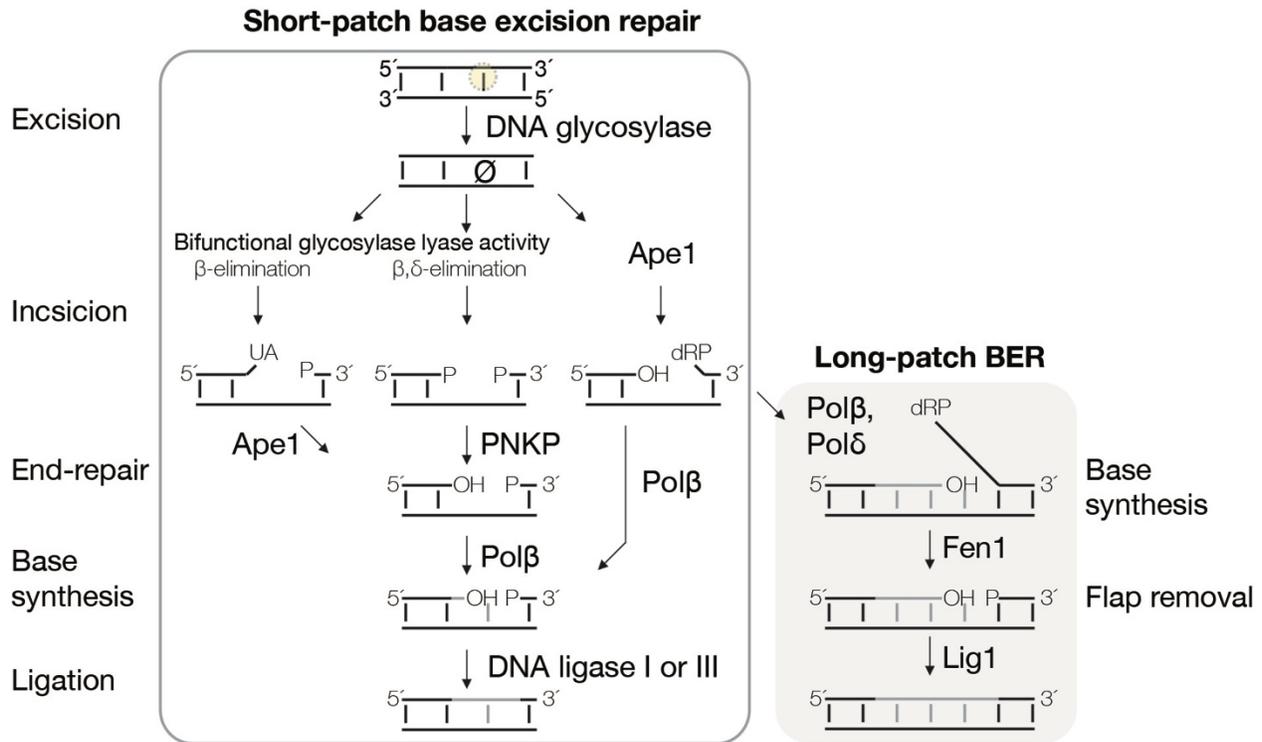


Figure 1.2: The steps of base excision repair

Base excision repair occurs in several steps. First, DNA damage is recognized by one of eleven human DNA glycosylases and excised from the DNA duplex creating an abasic site. The abasic site is then removed either by the lyase activity of bifunctional glycosylases or by the AP endonuclease, Ape1. For short-patch BER, the resulting strand incision event can be further processed by end-repair enzymes and the single-base gap is filled by the BER DNA polymerase, Pol β . The remaining nick is sealed by DNA ligase 1 or 3. For long-patch BER, following strand incision and removal of the abasic site, Pol β or the replicative DNA polymerase delta will extend several bases and create a single-strand flap. This flap is removed by the flap endonuclease, Fen1, and the resulting nick is sealed by DNA ligase 1.

Table 1.1: Human glycosylases, their subcellular localization, and their substrates

Enzyme	Subcellular localization	Mono/Bifunctional	Substrates
UNG1,2	Mitochondria, Nucleus	Monofunctional	U:A, U:G, 5-FU
SMUG1	Nucleus	Monofunctional	U:G, 5-hmU, 5-FU
MBD4	Nucleus	Monofunctional	U:G, hmU in CpG, T:G
TDG	Nucleus	Monofunctional	T:G, U:G
OOG1	Nucleus	Bifunctional	8-oxoG:C, Fapy:C
MUTYH	Nucleus	Monofunctional	A:8-oxoG/C/G
NTH1	Nucleus	Bifunctional	Tg, FapyG, 5-hC, 5-hU
NEIL1	Nucleus	Bifunctional	Tg, FapyG, FapyA, 8-oxoG, 5-hU
NEIL2	Nucleus	Bifunctional	Tg, FapyG, FapyA, 8-oxoG, 5-hU
NEIL3	Nucleus	Monofunctional (slow lyase activity)	FapyG, FapyA
MPG	Nucleus	Monofunctional	3-mA, 7-mG, 3-mG, inosine, ethenoA

(Krokan and Bjørås 2013)

DNA glycosylases are important for error-free repair of damaged bases, but many of them are not essential. Mice with homozygous knockout of most DNA glycosylases are viable and have only a mild accumulation of genomic mutations. This is possibly due to the redundant activity of DNA glycosylases, where several glycosylases recognize a single lesion. Moreover, other DNA repair pathways like nucleotide excision repair and mismatch repair can repair BER substrates; and translesion DNA polymerases can bypass DNA lesions allowing them to persist in the genome with relatively error-free consequences (Yoon et al. 2010a; Yoon, Prakash, and Prakash 2009).

Strand incision

Following the creation of an abasic site by DNA glycosylases or by spontaneous depurination/depyrimidination events, the DNA backbone is cleaved by an AP endonuclease. Ape1 is the major AP endonuclease in humans (discussed in detail below), however, bifunctional DNA glycosylases also have AP lyase activity. Ape1 activity cleaves on the 5' side of the abasic site creating a 3'-hydroxyl and a 5'-deoxyribose phosphate (5'-dRP). Bifunctional glycosylases can remove abasic sites in one of two ways: β -elimination (like NTHL1) or β - δ -elimination (like Neil1). β -elimination cleaves on the 3' side of the abasic site leaving a 3'-unsaturated aldehyde and a 5'-phosphate. Bifunctional glycosylases with δ -elimination activity will remove the 3'-unsaturated aldehyde leaving a 3'-phosphate (Krokan and Bjørås 2013). In any of these cases, the ends need to be processed in order to complete repair.

End-repair, base filling, and ligation

The 3'-unsaturated aldehyde product of β -elimination is removed by Ape1 leaving a one-base gap with 3'-OH and 5'-phosphate ends. The 3'-phosphate product of β - δ -elimination is removed by the polynucleotide kinase 3'-phosphatase (PNKP) also leaving a one-base gap with 3'-OH and 5'-phosphate ends. This one base gap is filled by the DNA repair polymerase Pol β .

Pol β also fills the one base gap and removes the 5'-dRP produced by Ape1 endonuclease activity. Pol β activity leaves a nick in the DNA backbone and a 3'-OH and 5'-phosphate. DNA ligase I or III will ligate the nick leaving an undamaged strand (Krokan and Bjørås 2013) (**Figure 1.2**). While both DNA ligase I and III ligate nicks during BER, DNA ligase I is the main nuclear ligase involved in BER (Gao et al. 2011).

To coordinate enzyme activity during BER, many repair factors are present in complex with the globular protein XRCC1. XRCC1 is a structural protein with no enzymatic activity, however, it coordinates the localization of the BER factors, Pol β , DNA ligase III, PARP1, REV1, and PNKP. XRCC1 has high affinity for nicked or gapped DNA and interacts with many DNA glycosylases and may help coordinate the steps of BER (London 2015).

Long-patch base excision repair

In contrast to the single-base synthesis BER mechanism described above, also known as short-patch BER, some damage events or cell cycle states favor the incorporation of 2-20 nucleotides in a process known as long-patch BER (Sattler et al. 2003; Mansour Akbari et al. 2009). The products of bifunctional glycosylase β - or β - δ -elimination are almost exclusively repaired by short-patch BER because the presence of an unmodified 5' nucleoside makes strand-displacement unlikely (Fortini et al. 1999). However, the presence of the 5'-dRP as a product of Ape1 activity can be repaired by either pathway (**Figure 1.2**). Following the strand incision events catalyzed mainly by Ape1, Pol β or another DNA polymerase will incorporate one base and then a replicative DNA polymerase will add several bases creating a small flap. The single-strand flap is removed by flap endonuclease 1 (Fen1), leaving a nick that is sealed by DNA ligase 1 (Krokan and Bjørås 2013; Sattler et al. 2003).

Determination of whether lesions are repaired by short-patch or long-patch BER is dependent on a number of factors. First, is the type of DNA damage. Lesions repaired by bifunctional glycosylases (like thymine glycol and other oxidative damage events) are almost exclusively repaired by short-patch BER, while lesions like ethenoA or abasic sites can be repaired by either pathway (Fortini et al. 1999). Second, the abundance of DNA polymerases can change the pathway selection. Pre-replicative cells undergo almost exclusively short-patch or two-base long-patch BER (Mansour Akbari et al. 2009). Additionally, cell cycle regulation of replicative DNA polymerases may lead to an increase in long-patch BER during some cell cycle phases. Finally, the abundance of the other cell cycle-regulated factors, like PCNA, promote long-patch BER when those factors are present in high abundance (Dogliotti et al. 2001).

Uracil in the genome

Sources of deoxyuracil in DNA

Deamination of cytosine

Deoxyuracil (dUMP) is one of the most commonly occurring DNA lesions and there are several mechanisms in which dUMP arises in the genome. Spontaneous and enzymatic deamination of cytosine is a major contributor of dUMP in genomic DNA and approximately 50-500 deamination events occur per genome per day (Lindahl 1993). Deamination of cytosine is inherently mutagenic as it leads to a U:G mismatch, and if not repaired, U:G mismatches will lead to C:T transitions during DNA replication. Alkaline hydrolysis of cytosine releasing ammonia results in deamination and conversion of cytosine to uracil. While deamination events are a common source of dUMP in the genome, the half-life of an individual cytosine is about 30,000 years in physiological conditions in double-stranded DNA (Lindahl 1993). Conversely, deamination rates occur ~100 times faster in single-stranded DNA compared to double-stranded

DNA (Lindahl 1993; Krokan, Drabløs, and Slupphaug 2002). This difference in stability in double-stranded vs single-strand DNA accounts for some of the differences in uracil accumulation between cell types. For example, cytosine deamination occurs on the order of 40 times more frequently in *S. cerevisiae* than *E. coli*, potentially due to slower rates of transcription in eukaryotes that results in more single stranded DNA being present in eukaryotic cells compared to *E. coli* (Lindahl 1993).

In addition to spontaneous deamination, activation-induced cytosine deaminase (AID) activity catalyzes the deamination of cytosines in a CpG context and is a major contributor to somatic hypermutation and recombination in immunoglobulin genes (Krokan, Drabløs, and Slupphaug 2002). AID activity is critical for antibody diversification and class switch recombination in B cells. AID catalyzes the deamination of cytosine in the immunoglobulin locus leading to an accumulation of C to T transitions after DNA replication (Di Noia and Neuberger 2007). Error-prone repair of U:G mismatches in immunoglobulin genes can lead to additional mutations. Base excision repair of U:G in class switching regions leads to the formation of double stranded breaks that are crucial for recombination events in the immunoglobulin locus. In fact, individuals with mutations in AID develop immunodeficiencies (Revy et al. 2000).

Some evidence also suggests that RNA editing enzyme, APOBEC, can also deaminate cytosine in DNA. Specifically, APOBEC3 family of enzymes have been characterized by their role in deamination of viral genomes (Willems and Gillet 2015). However, recent evidence suggests that some APOBEC3 enzymes act upon genome DNA (Landry et al. 2011). Mutations in APOBEC3 family members are associated with many cancers and result in APOBEC-specific mutational patterns in these cancers (Rebhandl et al. 2015). APOBEC3 also inhibits transposition of LINE-1 elements in deamination dependent and independent mechanisms by deaminating the

LINE-1 cDNA and promoting its degradation (Feng et al. 2017). It is still unknown how APOBEC3 activity is regulated in the nucleus and whether its activity on genomic DNA is regulated or an off-target effect of innate immune response (Narvaiza, Landry, and Weitzman 2012).

Incorporation of dUTP

Incorporation of dUTP during DNA replication can lead to additional dUMP accumulation throughout the genome. dUTP is a naturally occurring intermediate in dTTP biogenesis and unlike deamination of cytosine, incorporation of dUTP is not inherently mutagenic. In proliferating cells, it is estimated that one dUTP is incorporated for every 10^4 dTTPs (Goulian, Bleile, and Tseng 1980c, 1980a), meaning that on the order of 80,000 uracil bases are incorporated each cell division. Mammalian DNA polymerases do not discriminate between dUTP and dTTP and will incorporate both with equal affinity (Wardle et al. 2008). To limit the amount of uracil incorporated into healthy cells, the intracellular levels of dUTP are maintained at a low abundance compared to dTTP (0.2 and 37 μ M respectively) (Traut 1994). There are three pathways in cells that help maintain the dTTP:dUTP ratios: dUTP degradation by dUTPase, *de novo* dTMP biosynthesis, and salvage dTMP biosynthesis.

dUTP regulation by dUTPase

dUTPase is a ubiquitous enzyme that converts dUTP to dUMP. dUTPase plays an important role in maintaining low cellular levels of dUTP to reduce the rate of dUTP incorporation into DNA. The product of dUTPase activity, dUMP, feeds into the *de novo* dTTP biosynthesis pathway (Hirmondo et al. 2017). When dUTPase is knocked down in human cells, intracellular dUTP levels increase and cell proliferation rates decrease (Studebaker et al. 2005). High nuclear dUTPase activity can counteract the cytotoxicity of inhibiting the *de novo* dTTP biosynthesis

pathway, indicating that high dUTP level may be toxic to cells (Ladner et al. 2000; Tinkelenberg, Hansbury, and Ladner 2002).

De novo dTMP biosynthesis

De novo dTMP biosynthesis generates dTMP from dUMP through a multistep process involving several enzymes and a cofactor tetrahydrofolate (THF). Reduction of cellular folate pools by antifolates or by limiting dietary folate, results in more dUTP in cells and in the genome through reduction in *de novo* dTMP biosynthesis (Blount et al. 1997; Duthie et al. 2010; Goulian, Bleile, and Tseng 1980a). Direct inhibition of *de novo* dTMP biosynthesis enzymes also result in more dUTP in cells and in the genome (J. Chon, Field, and Stover 2019; Paone et al. 2014; MacFarlane et al. 2011; Blount et al. 1997). For example, heterozygous knockout of the *de novo* dTMP biosynthesis enzyme SHMT1 in mice leads to accumulation of uracil in some cell types and a neural tube developmental defect. This phenotype is exacerbated in mice fed a low-folate diet (MacFarlane et al. 2011). Many common chemotherapies (like 5-fluorouracil and pemetrexed) can inhibit *de novo* dTTP biosynthesis by sequestering or inhibiting thymidylate synthetase (J. Chon, Field, and Stover 2019). Thymidylate synthetase catalyzes the conversion of dUMP to dTMP and plays an important role in maintaining the ratio of dTTP to UTP. *De novo* dTMP biosynthesis helps maintain the dTTP:dUTP ratios by actively converting dUMP to dTMP and any disruption in this pathway can lead to increased dUTP accumulation and incorporation into DNA (J. Chon, Field, and Stover 2019).

Salvage dTMP biosynthesis

In addition to *de novo* dTMP biosynthesis, the salvage dTMP biosynthesis pathway also helps maintain cellular dTMP levels. Phosphorylation the thymidine nucleoside by thymidine kinase (TK1 in the cytoplasm and nucleus and TK2 in mitochondria) replenishes cellular dTMP

pools. Cytotoxicity of antifolates in cancer cells increases upon knockdown of TK1 because antifolates and TK1 knockdown inhibit redundant *de novo* and salvage dTMP biosynthesis pathways (Di Cresce et al. 2011). Thymidine kinase also plays a role in recovering from DNA damage. Cells deficient in thymidine kinase are more sensitive to UV and gamma radiation (McKenna and Hickey 1981; McKenna, McKelvey, and Frew 1988). Genotoxic stress also leads to an increase in nuclear localization of TK1 and better survival of cancer cells (Y.-L. Chen, Eriksson, and Chang 2010).

Repair of deoxyuracil in DNA

Uracil-DNA glycosylase - UNG

The uracil DNA glycosylase, UNG, is responsible for removing most of the uracil throughout the genome (**Figure 1.3a**). This monofunctional glycosylase recognizes uracil through a conserved leucine residue. UNG scans the genome through stochastic movements (Friedman, Majumdar, and Stivers 2009). When UNG is bound to a uracil base in a DNA chain, it hydrolyzes the N-glycosidic bond through the conical pinch–push–plug–pull mechanism, creating an abasic (AP) site (Jiang and Stivers 2002). The repair process is continued by the AP endonuclease, Ape1, that cleaves the 5' side of the AP site leaving a 3'-OH and a 5'-dRP. DNA polymerase beta (PolB) continues the processes by synthesizing a new base and removes the 5'-dRP flap, and thus filling in the 1 nucleotide gap and leaving a 3'-OH and a 5'-phosphate nick in the DNA backbone. The process is completed by DNA ligase 1, which seals the DNA backbone (**Figure 1.3a**). The above steps are canonical base excision repair steps that are explained in further detail above in the base excision repair section.

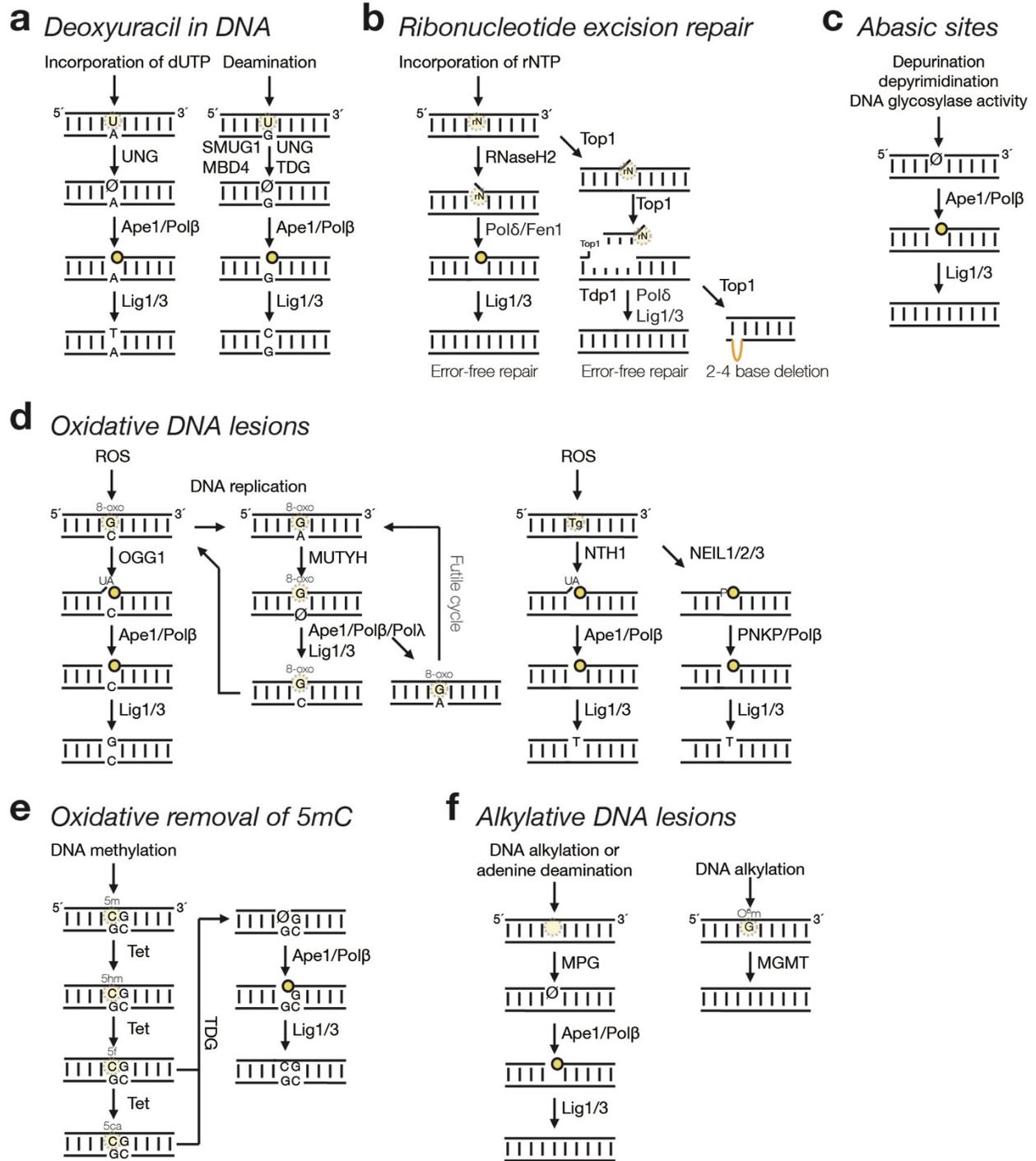


Figure 1.3: Repair of single-base DNA lesions by BER and RER pathways

a. Deoxyuracil is incorporated into the genome during DNA replication by DNA polymerases or created by deamination of cytosine residues. Uracil is removed by one of four glycosylases depending on the base-pairing and other factors. Repair is completed by BER enzymes resulting in clean DNA. Opaque yellow circles represent 5'-phosphates that can be captured in our single-cell DNA repair method, Haircut.

b. Ribonucleotides are incorporated into the genome during DNA replication by DNA polymerases. Ribonucleotide excision repair is initiated by RNaseH2, which cleaves on the 5' side of the ribonucleotide. The resulting danglantly 5'-rNP is processed by Ape1, Pol δ , and DNA ligase 1 or 3 for error-free repair. Top1 can also remove ribonucleotides through error-free and error-prone mechanisms.

c. Abasic sites created by depurination, depyrimidination, or DNA glycosylase activity are removed by Ape1 and subsequent BER enzymes.

d. The oxidative lesion 8-oxoG (left) is removed by the bifunctional DNA glycosylase OGG1 leaving a 3'-unsaturated aldehyde (UA) and a 5'-phosphate. The 3'-UA is removed by Ape1 and subsequent BER steps lead to complete repair. If 8-oxoG is mispaired with an adenine, the monofunctional DNA glycosylase MUTYH will remove the adenine and Ape1 will remove the abasic site (middle). Pol β or Pol λ will incorporate a cytosine across from the 8-oxoG and OGG1 can remove the OGG1 in an error-free manner. The oxidative lesion thymine glycol (Tg) is removed by bifunctional DNA glycosylases Nth1, NEIL1, NEIL2, or NEIL3. Nth1 leaves a 3'-UA and 5'-phosphate that is further processed by Ape1, Pol β , and DNA ligase 1 or 3. NEIL1, 2, and 3 have β,δ -lyase activity that leaves a 3'-phosphate, 5'-phosphate, and a single-base gap. The 3'-end is dephosphorylated by PNKP and Pol β and DNA ligase 1 or 3 will completely repair the strand break.

e. Oxidative removal of methylated cytosine residues at CpG sites is carried out by sequential oxidation of 5mC to 5hmC to 5fC to 5caC. 5fC and 5caC are removed by the monofunctional glycosylase TDG and repaired completely by BER enzymes.

f. Alkylative lesions like N3mA and deaminated adenines are removed by the monofunctional glycosylase MPG and other BER enzymes for complete repair. The alkylative lesion O6mG is removed by the single-use methyltransferase, MGMT, which removes the methyl group and leaves an undamaged guanine residue.

Symbols: Transparent yellow circles are sites of DNA damage. Opaque yellow circles indicate 5'-phosphates that can be captured in our single-cell DNA repair assay, Haircut. \emptyset is an abasic site.

UNG is the primary glycosylase that removes dUMP that is incorporated during DNA replication. UNG is differentially transcribed from the *Ung* gene locus to make two isoforms - UNG1 and UNG2 (Slupphaug et al. 1995). UNG1 and UNG2 have the same catalytic domain, but different N-terminal sequences lead to differential localization. The N-terminus of UNG1 contains a mitochondrial localization sequence and the N-terminus of UNG2 contains a nuclear localization sequence (Nilsen et al. 1997). UNG2 localizes to the nucleus and colocalizes with the DNA replication complex. Specifically, the N-terminus of UNG2 contains binding motifs for the DNA replication factors PCNA and RPA (Otterlei et al. 1999). UNG2 is also located diffusely throughout the nucleoplasm, indicating it plays a role in the removal of deaminated cytosines outside of DNA replication (Kavli et al. 2002). UNG2 transcription is cell-cycle regulated and peaks during S-phase, indicating its importance in removing dUMP incorporated during DNA replication (Olinski, Jurgowiak, and Zaremba 2010). UNG1, however, is not cell cycle regulated and is also lacking PCNA and RPA binding domains. However, UNG1 expression increases after oxidative stress, indicating mitochondrial BER may be important in responding to oxidative damage (M. Akbari et al. 2007).

UNG1 and UNG2 have identical catalytic domains and their activity will be described under the umbrella name, UNG. UNG will remove uracil from single-stranded or double-stranded DNA and has a high turnover rate that is compatible with repairing uracil simultaneously with DNA replication as it travels with the rapidly moving replication fork (Krokan, Drabløs, and Slupphaug 2002). Additionally, UNG will remove uracil from a U:A base pair or U:G mismatch with similar affinity, although the sequence of surrounding bases can effect UNG cleavage rates (Nilsen et al. 1995; Slupphaug et al. 1995). Cells with inactivating mutations in UNG (*UNG*^{-/-}) accumulate approximately 10 fold more uracil compared to wild-type cells (Nilsen et al. 2000).

Conversely, the spontaneous mutation rates in *UNG*^{-/-} mice are only modestly increased 1.3-1.5 fold, most likely due to the compensating activity of other uracil-DNA glycosylases (SMUG1, MBD4, and TDG). In contrast, *ungΔ S. cerevisiae* and *UNG*^{-/-} immortalized cell lines accumulate 3-20 times more mutations than wildtype cells, possibly due to the increased growth rate (Radany et al. 2000; Impellizzeri, Anderson, and Burgers 1991). Since most dUMP is incorporated during DNA replication, it is likely that as cells undergo more cell divisions, they accumulate more and more uracil that can lead to increased mutation rates with each cell division.

Other uracil glycosylases - SMUG1, MBD4, and TDG

In addition to UNG, there are three other uracil DNA glycosylases (**Figure 1.3a**). Single-strand-selective monofunctional uracil-DNA glycosylase 1 (SMUG1) is the main enzyme responsible for removing the oxidation lesion, 5'-hydroxymethyluracil (5hmU), from single-stranded or double-stranded DNA (An et al. 2005). SMUG1 will also remove uracil from U:G mismatches but will not recognize U:A base pairs (Kavli et al. 2007). In the absence of UNG, there is a significant increase in uracil in the genome, however, the cells are not inherently mutagenic due to the removal of mutagenic U:G mismatches by SMUG1. When both UNG and SMUG1 are knocked out or knocked down in cells, the number of mutations increases and cells are more sensitive to gamma radiation (An et al. 2005). While SMUG1 and UNG have similar substrates, they still have nonredundant functions in cells.

Two additional glycosylases recognize uracil in double-stranded DNA, thymine-DNA glycosylase (TDG) and methyl-CpG binding domain 4, DNA glycosylase (MBD4). TDG removes T and U from T:G and U:G mismatches. TDG will also remove 5hmU from double-stranded DNA (Jacobs and Schär 2012). Like SMUG1, TDG is a very low turnover enzyme because it binds tightly to the AP site product (Visnes et al. 2009). Its activity is enhanced with the addition of

Ape1 because Ape1 will displace TDG from the AP site (Waters et al. 1999). MBD4 will recognize the same substrates as TDG, however it binds much stronger to T:G mismatched in methylated CpG sites indicating its role in repairing deaminated 5'-meC in methylated CpG sites (Hendrich et al. 1999). MBD4 also interacts with mismatch repair (MMR) machinery, indicating it may play a distinct role in repair (Bellacosa et al. 1999).

Uracil repair and genome stability and cancer

While U:A base pairs are not inherently mutagenic, the removal of uracil from DNA by base excision repair enzymes creates single-stranded breaks. An accumulation of single-stranded DNA breaks throughout the genome can lead to regions that have double-stranded breaks that are inherently mutagenic and cytotoxic. UNG^{-/-} mice only have ~1.5 fold increase in spontaneous mutational rates compared to wildtype mice, indicating that the majority of uracil accumulation in the genome is non-mutagenic or removed by other uracil DNA glycosylases (SMUG1, MBD4, and TDG) (Nilsen et al. 2000). However, yeast with a deletion in dUTPase have a mutator phenotype that is exacerbated by inactivation AP endonucleases (Guillet, Van Der Kemp, and Boiteux 2006). Additionally, the mutator phenotype in dUTPase null yeast is alleviated by codeletion in the uracil DNA glycosylase UNG (Guillet, Van Der Kemp, and Boiteux 2006). This indicates that U:A base pairs are not mutagenic, but their repair by UNG can lead to genotoxic stress and mutagenesis.

U:A base pairs can also alter cell homeostasis when uracil is present within the coding regions of the genome. Using ectopic EGFP plasmids that contain or do not contain uracil in the coding region of EGFP, Lühnsdorf, et. al (2014), found that EGFP expression was reduced in the uracil-containing plasmid, but the expression levels recovered in UNG knockdown cells. The authors concluded that RNA polymerases do not stall at uracil in the template, but repair of the

uracil and creation of transient abasic sites and strand breaks does impair transcription rates (Lühnsdorf, Epe, and Khobta 2014).

Genomic uracil may also impair the affinity of DNA binding proteins. *In vitro* evidence indicates that some transcription factors have lower affinity for U:A base pairs present in transcription factor binding sites (Verri et al. 1990). U:A base pairs may also change secondary DNA structures that could alter DNA binding protein affinity (N. Yan et al. 2011; Marathe and Bansal 2010). However, there has not been significant *in vivo* data to suggest that genomic uracil significantly changes the functions of DNA binding proteins (J. Chon, Field, and Stover 2019).

Uracil and cancer

Intracellular dUTP levels are altered with several classes of chemotherapies, including thymidine synthetase inhibitors and antifolates. Alterations in dUTP levels either through inhibition of dTTP synthesis or alterations in folate metabolism leads to DNA damage and cell death mostly through a P53 and PARP1 dependent manner (J. Chon, Field, and Stover 2019).

TS inhibitors

5-fluorouracil (5-FU) is a common chemotherapy. 5-FU is cytotoxic through multiple mechanisms. One mechanism is that 5-FU will bind and inhibit thymidine synthetase, which converts dUMP to dTMP. This leads to an accumulation of dUMP that can be phosphorylated to dUTP and incorporated into the genome (Wyatt and Wilson 2009). 5-FU toxicity can be partially mitigated by overexpressing dUTPase, which will keep cellular dUTP levels lower by dephosphorylating dUTP to dUMP (Canman et al. 1994; Parsels et al. 1998). For a long time, the pervasive theory was that removal of uracil from the genome resulted in an accumulation of single-strand and double-strand breaks that overwhelm cellular DNA repair pathways and lead to cell death (Dianov et al. 1991). However, the toxicity of 5-FU is only mildly dependent on UNG

expression (Andersen et al. 2005) and 5-FU toxicity is higher in combination with UNG inhibition (Y. Yan et al. 2016) indicating that 5-FU toxicity is not solely due to increased genomic uracil nor is it dependent on UNG-induced single-strand breaks.

In addition to altering nucleotide pools, 5-FU and its deoxynucleotide metabolite, 5-fluorodeoxyuridine (5-dFU) can be incorporated into DNA (Wyatt and Wilson 2009). The literature behind the removal of 5-FU from genomic DNA is long and complicated, however, all uracil glycosylases have also been shown to remove 5-FU from DNA (Wyatt and Wilson 2009). Regardless of which glycosylase is the most important, if not repaired, 5-FU causes DNA polymerases to stall, leading to stalled or collapsed DNA replication forks. These 5-FU induced stalled replication forks are one of the main modes of 5-FU toxicity in cells (Y. Yan et al. 2016; Huehls et al. 2016).

Antifolates

Antifolates, like methotrexate and pemetrexed, generally inhibit folate metabolism, purine synthesis, and pyrimidine synthesis in cells. Methotrexate binds to dihydrofolate reductase (DHFR). DHFR catalyzes the conversion of dihydrofolate to tetrahydrofolate, which is a cofactor in *de novo* thymidine synthesis (Gonen and Assaraf 2012). Folate is also involved in other purine and pyrimidine metabolism pathways, and therefore methotrexate will broadly alter nucleotide pools. Antifolates also lead to oxidative DNA damage and an accumulation of 8-oxoG lesions. These lesions are lethal in cells that are null for the mismatch repair protein MSH2 (Martin et al. 2009). Pemetrexed inhibits thymidine synthetase, which leads to a depletion in dTTP and an accumulation of dUMP (Berger, Pittman, and Wyatt 2008). An imbalance of dTTP:dUTP levels leads to accumulation of uracil in the genome that can result in cytotoxic and mutagenic events discussed above.

Toxicity of methotrexate may depend on the accumulation of uracil in the genome and DNA repair mechanisms may play a role in reducing methotrexate toxicity. UNG expression, but not protein abundance or activity, increased in developing rats upon exposure to methotrexate (Vinson and Hales 2002). Folate deficiency (induced by a folate-deficient diet) in UNG^{-/-} mice, but not UNG^{+/+} mice, leads to neurodegeneration (Kronenberg et al. 2008). Folate deficiency induced by methotrexate treatment increased cell death in UNG^{-/-} cell lines (Kronenberg et al. 2008). Treatment with methotrexate leads to uracil being incorporated in the genome (Goulian, Bleile, and Tseng 1980a). Together, these data indicate that methotrexate-induced genomic uracil is toxic and removal of genomic uracil by UNG mitigates some of the toxic effects of methotrexate. Pemetrexed toxicity is also increased in UNG^{-/-} cells (L. D. Weeks et al. 2014) and UNG expression determines pemetrexed sensitivity in lung cancer (Lachelle D. Weeks, Fu, and Gerson 2013). These are just a few examples that highlight the unique interplay between DNA damage and DNA repair in cancer treatments.

Heterogeneity in uracil repair

There are several examples that highlight the heterogeneity of uracil repair. First, uracil repair is not the same across individuals. Chaim, IA, *et. al.* (2017) found that uracil repair capacity (on a U:G substrate) differed between a panel of cell lines and between individual donors. The differences in repair capacity was also used to predict cell line sensitivity to pemetrexed (Chaim et al. 2017a). Second, UNG expression alone can be used to determine sensitivity to pemetrexed (L. D. Weeks et al. 2014), indicating that UNG expression differences may be a source of inter- or intratumoral heterogeneity. Third, UNG expression and activity can vary across human PBMC cell types (Uhlén et al. 2015; Amanda L. Richer et al. 2020a). Finally, in addition to UNG heterogeneity, the other uracil DNA glycosylases also have tissue and cell cycle regulation that

can contribute to heterogeneity in uracil repair (Uhlén et al. 2015; Di Noia, Rada, and Neuberger 2006).

Ribonucleotides in the genome

Sources of ribonucleotides in DNA

The concentration of intracellular ribonucleotides (rNTPs) are much higher than dNTPs by ~ 30-200 fold in budding yeast (Nick McElhinny et al. 2010) and 10-130 fold in human cells (Traut 1994). To prevent the incorporation of rNTPs into DNA, DNA replicative polymerases exclude the 2'-hydroxyl (2'-OH) moiety of rNTPs with a conserved tyrosine residue (Brown and Suo 2011). The tyrosine residue acts as a 'steric gate' and inhibits rNTPs from entering the active site of DNA polymerases. While this exclusion favors the binding and incorporation of dNTPs into nascent DNA chains, all DNA polymerases can accommodate and incorporate rNTPs into DNA (Nick McElhinny et al. 2010). In budding yeast, *in vitro* evidence suggests that at least 13,000 ribonucleotides are embedded into DNA during DNA replication (Nick McElhinny et al. 2010), and *in vivo* evidence suggests that one ribonucleotide is incorporated every 6,500 bases (Lujan et al. 2013). In human cells, on the order of one million ribonucleotides are incorporated in DNA during each cell division (Clausen et al. 2013a), including ~5% of the lagging strand in the form of RNA primers called Okazaki fragments (L. Zheng and Shen 2011). These measurements make ribonucleotides the single most common lesion in the DNA duplex in cells (Kellner and Luke 2020).

While DNA polymerases generally incorporate single rNMPs into nascent DNA chains, short RNA-DNA hybrids occur naturally during DNA replication in the form of RNA primers (Okazaki fragments) that are essential for lagging strand DNA synthesis. Okazaki fragments are efficiently removed by the DNA replication machinery by short- or long-flap pathways where

DNA polymerase δ displaces the RNA primer and flap endonuclease 1 (Fen1) cleaves the resulting flap (L. Zheng and Shen 2011). RNA-DNA hybrids also transiently occur during transcription; however, the nascent RNA chain does not stably interact with the DNA duplex due to the presence of RNA binding proteins and topoisomerase 1 (Top1) at the site of transcription (Santos-Pereira and Aguilera 2015). In some cases, the stable interaction between nascent (or other) RNA chains and DNA create RNA-DNA hybrids that are stable. These hybrids create a displaced ssDNA section and are called R-loops. R-loops play an essential role in mitochondrial DNA replication and class-switch recombination in immunoglobulin genes (Aguilera and García-Muse 2012; Santos-Pereira and Aguilera 2015). However, R-loops are also a source of genome instability (Castellano-Pozo, García-Muse, and Aguilera 2012).

The presence of single ribonucleotides in DNA are also a source of genome instability. First, the presence of the 2'-OH on RNA makes it more unstable than DNA and prone to strand cleavage via hydrolysis of the phosphodiester bond (Y. Li and Breaker 1999). Second, DNA polymerases are inhibited by a ribonucleotide in the template DNA, which can lead stalled DNA replication forks (Nick McElhinny et al. 2010; McElhinny et al. 2010). To maintain a stable genome, it is important for cells to have an efficient mechanism to remove ribonucleotides from DNA.

Repair of ribonucleotides in DNA

RNaseH removal of rNMPs

Cells have two ribonuclease (RNase) enzymes, RNaseH1 (RNaseHI in prokaryotes) and RNaseH2 (RNaseHII in prokaryotes), that are responsible for removing rNMPs from the DNA duplex. The monomeric RNaseH1 is responsible for the removal of at least four sequential rNMPs in a DNA duplex (Cerritelli and Crouch 2009). Human RNaseH1 interacts with 11 bases of the

RNA/DNA substrate (Nowotny et al. 2007) and recognizes both strands of the complex (Lima, Rose, Nichols, Wu, Migawa, Wyrzykiewicz, Vasquez, et al. 2007; Lima, Rose, Nichols, Wu, Migawa, Wyrzykiewicz, Siwkowski, et al. 2007). RNaseH1 interacts with four 2'-OH molecules of the RNA strand, indicating that RNaseH1 does not have sequence preference for cleavage (Lima, Rose, Nichols, Wu, Migawa, Wyrzykiewicz, Vasquez, et al. 2007; Lima, Rose, Nichols, Wu, Migawa, Wyrzykiewicz, Siwkowski, et al. 2007). RNaseH1 is the only enzyme responsible for removal of rNPMs in mitochondria (Cerritelli et al. 2003), and plays a role, along with RNaseH2, in removal of R-loops in the nucleus (Cerritelli and Crouch 2016).

In contrast to the monomeric RNaseH1, the heterotrimeric RNaseH2 cleaves single rNMPs from DNA (**Figure 1.3b**). Each subunit of the heterotrimeric RNaseH2 enzyme is necessary for RNase activity (Jeong et al. 2004). While RNaseH2 activity is conserved from yeast to humans, the sequences of each subunit have diverged significantly, indicating the subunits have coevolved (Cerritelli and Crouch 2009). RNaseH2 is composed of three subunits: the catalytic RNaseHA, and two structural subunits RNaseH2B and RNaseH2C. RNaseH2B interacts with PCNA in proliferating cells, however, disruption of the RNaseH2B - PCNA interaction does not affect RNaseH2 activity or processivity (H. Chon et al. 2009), and only mildly disrupts RNaseH2 localization to the replication fork or PCNA during DNA replication (Bubeck et al. 2011). Taken together, these data indicate that RNaseH2 may help regulate removal of rNMPs during DNA replication. RNaseH2C is also a scaffolding subunit, and it may coordinate the orientation of the catalytic RNaseH2A and PCNA-interacting RNaseH2B subunits, however there is little direct evidence supporting this hypothesis. The roles of RNaseH2B and RNaseH2C are not completely understood, but their presence is essential for RNaseH2 activity and mutations in any subunit can affect RNaseH2 activity (Crow et al. 2006).

Much of our understanding of human RNaseH2 comes from studying mutations found in patients with the genetic disorder Aicardi-Goutières syndrome (AGS) that affects the brain, spinal cord, and immune system. Several mutations in all RNaseH2 subunits have been identified in AGS patients; however, there have been no reported complete loss-of-function mutations, indicating that RNaseH2 activity is essential in humans (Crow et al. 2006). Substitutions of G37 of the RNaseH2A subunit reduce the activity of RNaseH2 because this mutation limits the accessibility of the catalytic aspartic acid residue (Rohman et al. 2008). AGS mutations in RNaseH2A (G37S), RNaseH2B (K162T, A177T, V185) and RNaseH2C (K143I) all reduce the enzymes activity and processivity. Deletion of RNaseH2 subunits in mice is embryonic lethal in a P53 DNA damage response dependent mechanism (Reijns et al. 2012). Some evidence suggests that organisms have a ribonucleotide tolerance, where the abundance of rNMPs in the genome below the threshold activates the innate immune activation and above the threshold results in cell death (Uehara et al. 2018). However, *S. cerevisiae* and human cells lines can tolerate deletion of individual RNaseH2 subunits or in combination and the result is only a minor sensitivity to genotoxic stress (Cerritelli and Crouch 2009).

RNaseH2 has a wide variety of substrates that it can process, including removing single rNMPs from a DNA duplex, Okazaki fragments that were improperly removed leaving a residual ribonucleotide attached to the DNA strand (Murante, Henricksen, and Bambara 1998), stretches of four or more rNMPs in DNA (Jeong et al. 2004), and removal of R-loops (Lockhart et al. 2019). RNaseH2 has ribonucleotide excision repair (RER) activity that is responsible for removing most of the single rNMPs in the genome (Sparks et al. 2012). The RER mechanism has been reconstituted *in vitro*. RNaseH2 recognizes the 5' -rNMP - 3'dNMP junction, positioning the 5'-phosphate of the rNMP and the 2'-hydroxyl of the ribonucleotide in the active site of RNaseH2A

(Rychlik et al. 2010). The coordinated metal-ion dependent hydrolysis of the phosphate bond on the 5' side of the rNMP results in a single-strand DNA break with a 3'-hydroxyl and 5'-rNMP (Rychlik et al. 2010; Sparks et al. 2012). Pol δ (or Pol ϵ) will displace the 5'-rNMP by one base synthesis creating a rNMP flap. The single nucleotide flap is removed by the flap endonuclease (Fen1) or the exonuclease, Exo1. Finally, the remaining nick in the DNA is sealed by DNA ligase III (Sparks et al. 2012) (**Figure 1.3b**).

Top1 removal of rNMPs

In addition to RNaseH2 RER, topoisomerase 1 (Top1) can remove single rNMPs from DNA *in vitro* (Sekiguchi and Shuman 1997) and *in vivo* (Williams et al. 2013) (**Figure 1.3b**). Top1 removes ribonucleotides incorporated by the leading strand polymerase Pol ϵ (Williams et al. 2015). To remove rNMPs, Top1 can catalyze the transesterification of the 3' phosphate of the rNMP, however, instead of re-ligation (as Top1 does on the canonical supercoiled DNA substrate), the Top1-phosphate bond will often attack the 2'-OH group of the ribose resulting in Top1 release and a 2'-3' cyclic phosphate ribonucleotide (Kellner and Luke 2020). Top1 cannot re-ligate the strand nick of the resulting 2'-3' cyclic phosphate and the rNMP requires additional processing in order to be repaired. Namely, Top1 can cleave two bases upstream of the rNMP and with the addition of Tdp1 can repair the ribonucleotide error free (Sparks and Burgers 2015). However, the 2-5 nucleotide gap created by Top1 can also be realigned and ligated by Top1 that results in a 2-5 base deletion (Sparks and Burgers 2015) (**Figure 1.3b**). These 2-5 base deletions are characteristic of Top1 mediated slippage mutations (McElhinny et al. 2010; N. Kim et al. 2011). Top1 is the main source of genome instability in the absence of RNaseH2 (Kellner and Luke 2020). It is still unknown how Top1 recognizes ribonucleotides in DNA. It may be a byproduct of its canonical function of relieving supercoiling stress, or it could be a targeted mechanism.

Genomic ribonucleotide repair and genome instability and cancer

While rNMPs in the genome are not inherently mutagenic, rNMPs are subject to spontaneous hydrolysis due to the 2-OH group on the ribose that results in single-strand breaks (Y. Li and Breaker 1999). However, much of the genomic instability caused by RNaseH2 depletion is not due to spontaneous hydrolysis of the rNMP, but due to error-prone Top1 mediated removal of ribonucleotides and the lack of R-loop processing (Cornelio et al. 2017). In fact, codeletion of *Δtop1* and *Δrnh201* (RNaseH2B in mammals) in yeast ameliorates the mutator phenotype that is present in *Δrnh201* strain (Williams, Gehle, and Kunkel 2017). Codeletion of Top1 and RNaseH2 subunits in human cell lines decreases the number of single strand breaks and inhibits S phase cell cycle arrest caused by RNaseH2 deletion alone (Zimmermann et al. 2018).

RNaseH2 has two major functions: ribonucleotide excision repair activity that removes single rNMPs and processive RNase activity that removes stretches of rNMPs. Studies in yeast and mice have been used to tease out the importance of each of these functions *in vivo*. These studies have found that both activities are important for maintaining a stable genome.

To determine if the single rNMPs are responsible for the lethality of RNaseH2 deletion in mice, Uehara, *et al.* (2018) used a mouse RNaseH2-RED mutant that cannot remove single rNMPs but retains its processive rNMP removal activity that removes stretches of rNMPs. The authors determined that mouse development in RNaseH2-RED mice halted at the same stage of development as the RNaseH2A homozygous deletions (E 8.5) in a P53-dependent manner (Uehara et al. 2018). P53 and RNaseH2A double deleted embryos had slightly longer development, suggesting that single rNMPs processed by TOP1 may cause strand breaks in DNA that lead to cell death (Uehara et al. 2018). Uehara, *et al.* (2018) also proposed that there is a threshold of ribonucleotide accumulation throughout the genome before p53 induces apoptosis during

development. If the level of rNMPs is below the threshold, then cells induce an innate immune response via type I interferon signaling. If the level of rNMPs is above this threshold, then it leads to p53 mediated cell death (Uehara et al. 2018). This theory is somewhat supported by the observation in AGS patients where the level of interferon signaling decreases with age as loss of mental function declines (Rice et al. 2007). These findings suggest that cell viability decreases due to an accumulation of rNMPs that results in cell death as opposed to activation of an innate immune response in AGS patients. Additionally, In the absence of RNaseH2, inefficient processing of single rNMPs by TOP1 can sequester PARP1, and many cancers with deletions of RNaseH2 subunits are sensitive to PARP inhibitors (Zimmermann et al. 2018). All this evidence supports that RER activity of RNaseH2 plays an important role in maintaining genome integrity and cell viability.

Stretches of rNMPs from nascent transcription in DNA can form R-loops. R-loops can promote recombination, a process important in immune cell maturation. However, the presence of R-loops also increases the frequency of unregulated recombination events, leading to genome instability. R-loop processing by RNaseH2 is not necessary in yeast, possibly due to the redundant activity of RNaseH1 (Lockhart et al. 2019). Studies in yeast indicate that R-loops, and not single rNMPs, contribute to recombination events and genome instability (Zimmer and Koshland 2016). Zimmer and Koshland (2016) found recombination hotspots throughout the genome in RNase deficient yeast at probable R-loop formation sites. Additionally, RNaseH2 is required during G2 to remove R-loops for viability in the absence of RNase1 (Lockhart et al. 2019). These results indicate the removal of stretches of rNMP by RNaseH2 also contributes to genome stability.

RER and cancer

Like many DNA repair pathways, mutations in the RER pathway have been linked to poor prognosis and aggressive disease progression. In colorectal cancer, co-occurring inactivating mutations in RNaseH2 and P53 is associated with poor prognosis (Aden et al. 2019). RNaseH2B copy number is altered in chronic lymphocytic leukemia (CCL) and prostate cancer. Additionally, two loci on chromosome 13q14 are frequently lost in both CCL and prostate cancer. These loci are coincidentally in close proximity to RNaseH2B, so deletion of these loci can often co-occur with deletion of RNaseH2B (Kellner and Luke 2020). Cancers deficient in RNASEH2 activity are sensitive to PARP1 inhibition, most likely by trapping PARP1 on intermediates caused by Top1 removal of rNMPs (Zimmermann et al. 2018). Cells with RNaseH2 deletions are also sensitive to ATR inhibitors most likely due to the increased replicative stress caused by stalled replication forks at genomic rNMPs (Hustedt et al. 2019).

Hydroxyurea (HU) is a common chemotherapeutic. One major target of HU is ribonucleotide reductase (RNR). RNR converts ribonucleotide diphosphates into their respective deoxyribonucleotides (Nordlund and Reichard 2006). By blocking RNR, HU significantly reduces the abundance of dNTPs, which causes cells to arrest in S phase (Singh and Xu 2016). By reducing the abundance of dNTPs in cells, HU also increases rNMP incorporation during DNA replication. Deletions or inactivation of RNaseH2 leads to hypersensitivity to HU (Reijns et al. 2012; Arudchandran et al. 2000).

Heterogeneity in RER

There are several examples of heterogeneity in RER. First, in cancers, which are inherently heterogenous, there are several polymorphisms in RNaseH2 that correlate to poor prognosis or metastasis risk (Aden et al. 2019; Deasy et al. 2019). Second, while RNaseH2 is expressed in all

cells, mutations or deletions in RNaseH2 can lead to tissue-specific tumorigenesis, tissue specific disorders (e.g. AGS), and tissue-specific alterations in the innate immune response (Hiller et al. 2018; Aden et al. 2019; Kellner and Luke 2020; Mackenzie et al. 2016). Third, RNase activity plays an important role in immune cell maturation by resolving R-loops to promote class-switch recombination in immunoglobulin genes (Maul et al. 2017). Finally, RNaseH2 activity may play different roles in repair depending on the cell cycle phase by processing R-loops during G1 and removing single rNMPs during S phase (Lockhart et al. 2019). All of these examples show that RER can vary across individuals, tissues, and cell cycle phases.

Abasic sites

Sources of abasic sites

Apurinic or apyrimidinic sites (abasic sites or AP sites) are caused by the cleavage of the N-glycosidic bond of the deoxyribonucleoside. The N-glycosidic bond of deoxyribonucleosides are much more likely to undergo hydrolysis compared to ribonucleosides due to the labile N-glycosidic bonds that is a consequence of the deoxy-2' position of ribose (Lindahl 1993). Cellular DNA undergoes 3×10^{-11} depurinations per nucleotide residue per second and the rate of depurination is 100-500 times faster than that of depyrimidination (Lindahl and Nyberg 1972; Lindahl and Karlström 1973). Depurination/depyrimidination measurements *in vitro* would predict that there are approximately 10,000 spontaneous abasic sites formed each day in mammalian cells and one study even suggested that there up to ~50,000 steady state abasic sites in liver cells (J. Nakamura and Swenberg 1999).

Depurination rates can increase with some chemical treatments. For example, exposure to the alkylating agent methanesulfonate (MMS) can lead to alkylation adducts on purines that destabilize the nucleoside leading to spontaneous depurination (Strauss and Hill 1970). Oxidized

DNA lesions caused by oxidizing agents also produce unstable nucleosides that decompose to form abasic sites (Loeb and Preston 1986). Hydroxyl radicals formed during ionizing radiation can also cause depurination events (Loeb and Preston 1986). In addition to spontaneous hydrolysis of the N-glycosidic bond of nucleosides, abasic sites are formed by DNA glycosylases that remove damaged bases and leave an abasic repair intermediate (Krokan and Bjørås 2013).

If unrepaired, abasic sites are inherently mutagenic. Abasic sites can halt DNA replication machinery and transcription machinery. Bypassing abasic sites during DNA replication can result in one-base substitutions or deletions. Abasic sites are unstable and can facilitate cleavage of the phosphate backbone ~100 times faster than sites containing a purine or pyrimidine (Loeb and Preston 1986). Abasic sites are also chemically reactive and can lead to protein-DNA and DNA-DNA crosslinks that can be highly mutagenic and toxic to cells (Sczepanski et al. 2010).

Repair of abasic sites

The AP endonuclease, Ape1, is the main enzyme responsible for removing AP sites from the genome (**Figure 1.3c**). Ape1 generates a 3'-hydroxyl and 5'-deoxyribose phosphate (5'-dRP). This strand break is fully repaired either by short-patch or long-patch BER (discussed above, **Figure 1.2**). Ape1 is an important enzyme for base excision repair, but also may act as a signaling protein in response to oxidative stress and as a proofreading exonuclease (Whitaker and Freudenthal 2018).

Ape1 is composed of a rigid C-terminal nuclease domain and a flexible N-terminal domain. The C-terminal domain of Ape1 binds directly to the DNA and slides along the strand (Whitaker and Freudenthal 2018). Interacting with the phosphodiester backbone, Ape1 searches for abasic sites. Ape1 interacts with both the major and minor grooves of the abasic site and will 'flip out' the abasic site, which is then stabilized within the active site by two domains that interact with the

5' and 3' ends of the abasic site (Whitaker and Freudenthal 2018). Structural studies have elucidated a catalytic mechanism in which the coordinated hydrolysis of the phosphate backbone can occur within the active site (Whitaker and Freudenthal 2018).

Ape1 kinetics are predicted to be important for genome stability and are characterized by a rapid catalysis and slow product release. The rapid catalysis quickly initiates repair on potentially unstable and mutagenic lesions and the slow release may help facilitate coordination with other BER enzymes to shelter the strand break initiated by Ape1 and to ensure the break is fully repaired (Whitaker and Freudenthal 2018). Supporting this idea, Ape1 activity is stimulated by the addition of other BER proteins: PARP-1, XRCC1, DNA ligase III, PNKP, and Tdp1 (Prasad et al. 2015a).

In addition to Ape1, there are other enzymes and pathways that have been shown to remove AP sites, such as TDP1 (Lebedeva et al. 2012a) and PARP1 (Khodyreva et al. 2010a; Prasad et al. 2015a). Cleavage of abasic sites can also be catalyzed by bi-functional DNA glycosylases such as OGG1 (Dyrkheeva, Lebedeva, and Lavrik 2016). Together these enzymes contribute to error-free removal of abasic sites.

Other activities of Ape1

In addition to removing abasic sites, Ape1 has several other activities in cells. Ape1 has 3' to 5' exonuclease activity that can remove damaged bases, chain terminating drugs, blocked termini, and mismatched bases on the 3' end (Whitaker and Freudenthal 2018; Wong, DeMott, and Demple 2003; Chou and Cheng 2002, 2003). While the exonuclease activity of Ape1 is much slower than its endonuclease activity, the exonuclease activity occurs at a similar rate to other end-processing enzymes in the cell (namely Ape2, Tdp1, aprataxin, PNK, and the lyase activity of Pol β), indicating that the exonuclease activity of Ape1 is biologically relevant (Whitaker and Freudenthal 2018). The exonuclease activity of Ape1 can remove several damage events present

in cells. For example, bifunctional glycosylases leave a 3' α,β -unsaturated aldehyde group that Ape1 removes in order for complete repair to occur (**Figure 1.2**). Additionally, Ape1 can enhance bifunctional glycosylase (e.g. OGG1) activity by facilitating the release of the glycosylase from the incised product (Dyrkheeva, Lebedeva, and Lavrik 2016; Sidorenko, Nevinsky, and Zharkov 2008). Ape1 removes 3' 8-oxoguanine that can form due to oxidative damage or as a product of OGG1 activity. Ape1 can also remove mismatches that are inserted by the BER Pol β , which can incorporate one mismatch for every 4,000 nucleotides (Osheroff et al. 1999). Pol β itself does not have proofreading activity, but it is predicted that Ape1 can serve as an extrinsic proofreader for BER since Pol β fidelity increases in the presence of other BER factors (Matsuda et al. 2003) and recent structural evidence has shown that the catalytic pocket of Ape1 can accommodate a mismatched base and can function as a BER proofreader (Whitaker and Freudenthal 2018). Ape1 can also incise certain DNA adducts produced by ionizing radiation like 5,6-dihydrothymidine (Gros et al. 2004). In these cases, the reaction yields a 3'-OH and a dangling 5' base that can be further removed by BER enzymes.

In addition to its BER roles, Ape1 is also involved in stimulating the DNA binding activity of several transcription factors (A. R. Evans, Limp-Foster, and Kelley 2000). The N-terminal of Ape1 is essential for its redox activity and has been associated with redox activation of several transcription factors including AP-1 that regulates a number of cellular processes (Xanthoudakis and Curran 1992; Hirota et al. 1997). Ape1 transcription also increases in the presence of oxidative stress, indicating it plays a role in cellular responses to oxidative stress, in part due to endonuclease activity (Grösch, Fritz, and Kaina 1998; Fritz et al. 2003).

Ape1 may also play a role in RNA processing (Antoniali, Malfatti, and Tell 2017). HeLa cells with Ape1 knockout express two times more c-myc mRNA (T. Barnes et al. 2009).

Additionally, Ape1 may also be involved in cleaving some microRNAs, other mRNAs, and part of the SARS-coronavirus (Antoniali, Malfatti, and Tell 2017; W.-C. Kim, King, and Lee 2010). Taken together, Ape1 has an essential role in mammalian cells that extends past its endonuclease activity.

Heterogeneity of Ape1 activity

While Ape1 is ubiquitously expressed in all cell types, its expression level and localization patterns differ between tissue types and neighboring cells (A. R. Evans, Limp-Foster, and Kelley 2000). Expression levels of Ape1 protein and mRNA differ between cell types and tissues. For example, within peripheral blood mononuclear cells (PBMCs), Ape1 is most highly expressed in dendritic cells compared to all other cell types, including the dendritic precursors, monocytes (Uhlén et al. 2015). Additionally, Ape1 nuclear localization differs in normal versus malignant tissue (Kakolyris et al. 1998). Ape1 associates with chromatin in cerebellum tissues but not in epidermal cells (Duguid et al. 1995). Paradoxically, Ape1 is found almost exclusively in the cytoplasm in some cell types, including motor neurons and macrophages (Duguid et al. 1995). Cytoplasmic localization of Ape1 may indicate that Ape1 plays a role in relieving oxidative stress in those cell types, rather than DNA repair (Kakolyris et al. 1998). Repair of abasic sites also decreases with age, possibly due to changes in Ape1 expression or activity (Pons et al. 2010a). Taken together, the differential expression and cellular localization indicate that Ape1 may play a multifunctional, complex, and heterogenous role in alleviating oxidative stress and DNA repair.

Oxidative DNA damage and repair

Sources and types of oxidative DNA damage

Endogenous sources of oxidative damage

Endogenous or environmental sources can cause oxidative DNA damage. Identical DNA lesions can form regardless of the source of oxidative damage. Endogenous sources of oxidative stress come from cell metabolism and induction of the inflammatory response. Metabolic reactions can form reactive oxygen species (ROS) in the form of superoxide, hydrogen peroxide, peroxide radicals, and singlet oxygen (De Bont and van Larebeke 2004). Most ROS is made during mitochondrial oxidative phosphorylation in which 1-5% of oxygen undergoes single-electron transfer that generates superoxide anion radicals (Chance, Sies, and Boveris 1979). This superoxide is quickly converted to hydrogen peroxide and further to water by superoxide dismutase, catalase, and glutathione peroxidase. However, hydrogen peroxide can form a peroxide radical if reacted with transition metals. Peroxide radicals are highly reactive and will cause oxidative damage to surrounding molecules (De Bont and van Larebeke 2004). These radicals, when reacted with DNA, lead to the formation of damaged bases and DNA strand breaks. Some cells, like neutrophils and eosinophils produce radicals during an immune response to fight infection. Cells exposed to activated neutrophils and eosinophils incur oxidative DNA damage (Z. Shen, Wu, and Hazen 2000). While the cell has many mechanisms to reduce cellular ROS, oxidative DNA damage events are still very common (De Bont and van Larebeke 2004).

To prevent damage caused by endogenous sources of ROS, cells are equipped with enzymatic and nonenzymatic antioxidants (Birben et al. 2012). In many cancers and inflammatory diseases, there is a disruption in the balance between cellular ROS and antioxidants that can lead to an increase in oxidative stress and accumulation of oxidative DNA damage (De Bont and van

Larebeke 2004). Additionally, oxidative DNA lesions accumulate with age, indicating that cellular ROS and antioxidant metabolism may change with age (Y. J. Wang et al. 1995).

Exogenous sources of oxidative damage

There are several exogenous sources of ROS that can lead to oxidative DNA damage. In addition to producing bulky DNA lesions, exposure to UV light, mainly UVA, can lead to ROS through excitation of cellular photosensitizers (Cadet, Douki, and Ravanat 2015). Cigarette smoke can also increase cellular ROS (Church and Pryor 1985). Additionally many drugs and chemotherapies increase ROS through altering mitochondrial metabolism or altering antioxidant synthesis (H. Yang et al. 2018).

Types of oxidative DNA damage

Oxidative DNA damage can take many forms that are summarized below.

Oxidized purines: 8-oxoG and 8-oxoA

Guanine is highly susceptible to oxidative damage and one of the most common oxidative lesions on guanine is 7,8-dihydro-8-oxoguanine (8-oxoG). 8-oxoG is inherently mutagenic and DNA polymerases will readily incorporate an adenine across from 8-oxoG leading to G to T transversions (Neeley and Essigmann 2006; Shibutani, Takeshita, and Grollman 1991). In addition to 8-oxoG, the free guanine nucleotide can be oxidized to 8-oxodGTP and incorporated into nascent DNA across from either cytosine or adenine (Tajiri, Maki, and Sekiguchi 1995). 8-oxodGTP has also been used as a biomarker for oxidative stress in tissues and fluids (Hiroshi Kasai 1997).

Adenine can also be oxidized to form 7,8-dihydro-8-oxoadenine (8-oxoA). Some studies indicate that 8-oxoA occurs just as frequently as 8-oxoG under oxidative stress, although 8-oxoA

has been studied less frequently (Y. J. Wang et al. 1995). 8-oxoA can base pair with both cytosine and guanine leading to A to G transitions and A to C transversions (Kamiya and Kasai 1995).

FapyG

Other common lesions that form under oxidative stress are 2,6-diamino-4-hydroxy-5-formamidopyrimidine (FapyG) and 4,6-diamino-5-formamidopyrimidine (FapyA). FapyG and FapyA are formed at equal or higher levels than 8-oxoG after oxidative stress, although are studied much less frequently (Cadet, Douki, and Ravanat 2008). Some measurements indicate that FapyG is two times more likely to form than 8-oxoG in cellular DNA exposed to gamma radiation (Pouget et al. 2002). FapyG and FapyA result from a hydroxy radical attack on guanine and adenine rings, that lead to an opening of the imidazole ring. Both FapyG and FapyA are mutagenic and adenine is incorporated opposite of both lesions *in vitro* (Delaney, Wiederholt, and Greenberg 2002; Wiederholt and Greenberg 2002).

Oxidized pyrimidines

Oxidized pyrimidines can take many forms. Some of the most commonly studied oxidative pyrimidine lesions are thymine glycol and hydroxyuracil (M. D. Evans, Dizdaroglu, and Cooke 2004). Oxidation of the C5 of cytosine and thymine followed by the addition of a hydroxyl ion leads to the formation of cytosine glycol and thymine glycol (Tg) (M. D. Evans, Dizdaroglu, and Cooke 2004). Although cytosine glycol and Tg can form, Tg is much more commonly studied. Tg is a mutagenic (Basu et al. 1989) and DNA polymerase blocking lesion, however, translesion synthesis polymerases can bypass Tg adducts in an error-free manner (Yoon et al. 2010a).

Cytosine glycol is unstable and spontaneously deaminates and dehydrates to form hydroxyuracil (hU), this process is sometimes called oxidative deamination of cytosine (M. D.

Evans, Dizdaroglu, and Cooke 2004). hU can form base pairs with any of the four nucleotides, giving it high mutagenic potential (Thiviyathan et al. 2005).

Oxidation of 5-methylcytosine

5mC is a stable nucleotide modification that is mostly present in silenced regions of the genome (Arne Klungland and Robertson 2017). While most of 5mC is persistent throughout cell divisions, cytosine methylation status can change during development and in response to environmental changes (Arne Klungland and Robertson 2017). However, the mechanism 5mC removal is still under investigation. One mechanism in which 5mC is removed from the genome is through the stepwise enzymatic oxidation of 5mC. A series of oxidation reactions are catalyzed by the ten-eleven-translocation (Tet) enzymes and result in oxidation of 5mC to 5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-fC), and 5-carboxylcytosine (5-caC) (Ito et al. 2011). Both 5-mC and 5-hmC are stable in the genome, play a role in regulating gene expression, and their regulation is often mis-controlled in cancers (Y. Xu et al. 2011; Lian et al. 2012). Further oxidation of 5-hmC leads to the sequential formation of 5-fC and 5-caC, which are both lowly abundant in the genome and therefore thought to be less stable (Ito et al. 2011). Additionally, both 5-fC and 5-caC are actively removed from the genome via base excision repair enzymes, however, the biological roles, if any, for 5-fC and 5-caC are still under investigation.

Cyclopurines

Reactions between hydroxyl radicals and purines can create cyclopurines (cyc-dG and cyc-dG) (Becker and Sevilla 1993). Cyclopurines were originally identified as oxidation products induced by ionizing radiation, but have since been found in normal tissues and cells (Brooks 2017). Cyclopurines contain an additional 5'-8 carbon-carbon bond that causes them to block DNA polymerases (Kuraoka et al. 2000) and slow or halt transcription (You et al. 2012).

In response to ionizing radiation, cyclopurines are less likely to form than other oxidative lesions (Terzidis, Ferreri, and Chatgialloglu 2015), however, they are quite stable (Theruvathu et al. 2007). Like other bulky lesions, cyclopurines are removed by nucleotide excision repair machinery (Kuraoka et al. 2000; Brooks et al. 2000).

Repair of oxidative DNA damage

Removal of 8-oxoG

There are three conserved DNA repair enzymes that mitigate the mutagenic potential of 8-oxoG lesions: MTH1, OGG1, MUTYH (MutT, MutY, MutM in bacteria) (D. E. Barnes and Lindahl 2004). MTH1 hydrolyzes the nucleotide, 8-oxodGTP, and reduces its abundance in the nucleotide pool. The DNA glycosylase OGG1 removes 8-oxoG from the 8-oxoG:C base pairs. If an 8-oxoG lesion undergoes DNA replication creating 8-oxoG:A base pair, MUTYH will excise the adenine. Upon hydrolysis of adenine from 8-oxoG:A by MUTYH, creating an abasic site. Ape1 and Pol β or Pol λ can incorporate a cytosine across from 8-oxoG giving OGG1 another opportunity to repair the 8-oxoG lesion in an error-free manner (D. E. Barnes and Lindahl 2004).

MTH1 reduces 8-oxodGTP in the nucleotide pool

Compared to oxidized bases in the DNA duplex, oxidized nucleotides are more likely to form and can be incorporated into replicating DNA leading to DNA damage or mutations. 8-oxodG is a common oxidative lesion and its nucleotide 8-oxodGTP is also formed under oxidative stress (H. Kasai and Nishimura 1984). To prevent incorporation of oxidized nucleotides, mut-T homolog 1 (MTH1) hydrolyzes oxidized purine nucleoside triphosphates (both 8-oxodGTP and 8-oxodATP) to monophosphates and pyrophosphates (K. Sakumi et al. 1993; Sakai et al. 2002). 8-oxodGMP is processed further to the nucleoside 8oxodG to prevent its incorporation into DNA (Nakabeppu 2014). MTH1 is mitochondrial and nuclear localized indicating it is important for

surveying nucleotide pools throughout the cell (Kang et al. 1995). MTH1 knockout mice have an increase in liver and stomach cancers and an increase in spontaneous mutation rates in embryonic stem cells (Nakabeppu 2014). Overexpression of MTH1 in cancer cells increases their tolerance to oxidative stress (Kennedy et al. 1998) and MTH1 inhibitors have been shown to increase sensitivity to oxidative chemotherapies in cancer cells (Gad et al. 2014).

OGG1 removes 8-oxoG from the DNA duplex

The bifunctional glycosylase OGG1 is responsible for the error-free removal of 8-oxoG from the DNA duplex (**Figure 1.3d**). OGG1 will also excise 8-oxoA from 8-oxoA:C base pairs (Jensen et al. 2003). Structural and kinetic studies have shown that OGG1 binds to the DNA helix and rapidly scans short sections of the genome (~400 bases) for 8-oxoG lesions by probing the helix for weak interactions (David, O'Shea, and Kundu 2007; Blainey et al. 2006). Probing unmodified residues results in a small buckle of the base pair, however, probing 8-oxoG:C causes the 8-oxoG to rapidly flip from the helix and be captured by the base-binding site and processed by the active site in OGG1. Occasionally an unmodified guanine will flip into the base-binding site, but it cannot be processed by the active site and is released (A. Banerjee et al. 2005).

OGG1, like all other glycosylases that remove oxidized bases, is a bifunctional glycosylase. Bifunctional glycosylases hydrolyze the N-glycosidic bond of the DNA base and the phosphate backbone of the DNA strand creating a single-strand break (Wallace 2013). OGG1 is a β -elimination bifunctional glycosylase, meaning that it will first excise the 8-oxoG base and then cleave the phosphodiester bond on the 3' side of the newly created abasic site. This results in a 5'-phosphate and blocking 3'- α,β unsaturated aldehyde. The blocking 3' group is further processed and removed by Ape1 (Krokan and Bjørås 2013). While OGG1 has lyase activity *in vitro*, its activity is enhanced by Ape1 (Hill et al. 2001), and *in vivo* it has been suggested that OGG1 acts

mainly as a monofunctional glycosylase and Ape1 releases OGG1 the AP site (Dalhus et al. 2011).

OGG1 activity is modulated by a number of factors. First, other protein-protein interactions besides its interaction with Ape1 can enhance or impair OGG1 function. For example, the activity of OGG1 is enhanced when bound to the scaffolding protein XRCC1 (Marsin et al. 2003). XRCC1 also enhances many other glycosylases' activities (London 2015; Campalans et al. 2005). Second, posttranslational modifications can regulate OGG1 activity. Acetylation of OGG1 by histone acetyltransferases enhances its activity *in vivo* (Bhakat et al. 2006). Phosphorylation of OGG1 can either enhance its glycosylase activity or leave it unchanged (Dantzer et al. 2002; J. Hu, Imam, et al. 2005). Finally, the redox-sensitive cysteine residue of OGG1 is sensitive to oxidation and OGG1 activity is reduced when this residue is in its oxidized state (A. Bravard, 2006). Paradoxically, this causes OGG1 activity to be reduced under conditions of oxidative stress (Bravard et al. 2010; Morreall et al. 2015).

OGG1 inactivation during transient oxidative stress may play an important role in gene expression. Under oxidative stress induced by TNF- α or NF- κ B, 8-oxoG residues accumulate in gene promoters (Pan et al. 2016) and studies done in OGG1^{-/-} mice suggest that OGG1 may play a role in the immune response (Ba and Boldogh 2018; Vlahopoulos et al. 2019). There is also some evidence that accumulation of 8-oxoG under oxidative stress may change transcription factor binding in the genome to regulate cellular response to oxidative stress (Ba and Boldogh 2018).

MUTYH gives OGG1 another chance at error-free removal of 8-oxoG

The oxidative DNA lesion, 8-oxoG, readily forms a base pair with adenine and can lead to G to T transversions. To combat the mutagenic potential of 8-oxoG:A base pairs, the glycosylase MUTYH will excise the adenine from 8-oxoG:A base pairs and gives OGG1 another change at

removing 8-oxoG (Markkanen, Dorn, and Hübscher 2013) (**Figure 1.3d**). Humans have both nuclear and mitochondrial MUTYH isoforms that repair nuclear and mitochondrial DNA respectively.

Nuclear MUTYH associates with PCNA, RPA, and other replication fork factors and is responsible for removing adenine opposite of 8-oxoG in nascent, but not template DNA. Following removal of adenine, Pol β or Pol λ will efficiently incorporate a cytosine opposite of 8-oxoG (Krahn et al. 2003; van Loon and Hübscher 2009). Structural studies suggest that Pol β is able to incorporate a cytosine across from 8-oxoG (Krahn et al. 2003). However, *in vitro* and *in vivo* evidence suggests that Pol λ is the likely polymerase to synthesize error-free A:8-oxoG base pairs (van Loon and Hübscher 2009). Under oxidative stress, however, excision of adenine:8-oxoG can lead to a futile BER cycle where adenine is re-incorporated across from 8-oxoG and re-removed by MUTYH (Hashimoto et al. 2004) (**Figure 1.3d**, middle). This futile cycle can lead to persistent abasic sites or strand breaks in the genome that can lead to PARP-mediated cell death (Sugako Oka et al. 2008).

MUTYH also repairs mitochondrial DNA in the same manner as nuclear DNA. Under excessive oxidative stress and the futile cycle of A:8-oxoG repair by MUTYH also occurs in mitochondrial DNA. MUTYH activity can then lead to mitochondrial DNA depletion via an accumulation of single-strand breaks. Mitochondrial DNA depletion causes mitochondrial dysfunction and cell death (Sugako Oka et al. 2008; Nakabeppu 2014)

NTH1 removes oxidized pyrimidines

NTH1 (homologue of bacterial Nth protein) removes oxidized pyrimidines from the genome (**Figure 1.3d**, right). NTH1 is a bifunctional glycosylase that removes the damaged base and cleaves the 3' side of the resulting AP site leaving a 5'-phosphate and a 3'-unsaturated

aldehyde (Wallace 2013). NTH1 has a wide array of substrate specificity and will broadly excise a number of oxidized pyrimidines. One lesion removed by NTH1 is thymine glycol (Tg) (Dizdaroglu et al. 1999). NTH1 will also remove 5-hydroxyuracil and 5-hydroxycytosine, preferentially in base pairs with guanine (Dizdaroglu et al. 1999; Eide et al. 2001). NTH1 is also able to remove oxidized purines, like 8-oxoG when mispaired with guanine (Matsumoto et al. 2001). NTH1 will also remove FapyG lesions from genomic DNA (J. Hu, de Souza-Pinto, et al. 2005). NTH1 activity is enhanced by its association with the nucleotide excision repair factor, XPG, and may play a role in transcription-coupled repair of Tg (Bessho 1999; Cooper et al. 1997).

NEIL1, NEIL2, and NEIL3

The Nei like proteins, NEIL1, 2, and 3 (named after their bacterial homologues), are bifunctional glycosylases that remove oxidized pyrimidines, although NEIL3 has weak AP lyase activity (Wallace 2013). Contrary to NTH1 and OGG1, NEIL proteins have β,δ -elimination AP lyase activity (**Figure 1.3d**). The glycosylase and β,δ -elimination activity of NEIL proteins result in a one base gap with a 3' - and 5' -phosphate (Wallace 2013). The 3' -phosphate is removed by PNK leaving a 3' -hydroxyl that allows Pol β to fill the gap and DNA ligase III to seal the nick (**Figure 1.2** and **Figure 1.3d**, right). NEIL1, NEIL2, and NEIL3 recognize most of the same damaged bases. The NEIL proteins are responsible for removing Tg and other oxidized pyrimidines (Bandaru et al. 2002; Wallace 2013). NEIL1 will also remove FapyG and FapyA (Jaruga et al. 2004).

Even though their substrates overlap, the activities of NEIL1, NEIL2, and NEIL3 do not seem to be redundant (Rolseth et al. 2017). NEIL1, 2, and 3 have different expression patterns in tissues, cell cycle, and subcellular localizations indicating that they have unique activities (Wallace 2013). For example, during development, NEIL1 and NEIL2 protect against mitochondrial DNA

oxidation during neural crest differentiation (Han et al. 2019). While NEIL3 is not expressed in adult neural tissue (Uhlén et al. 2015), it is responsible for oxidative DNA repair in neural progenitor cells (Regnell et al. 2012). NEIL1 is highly expressed during S phase and is responsible for repairing oxidized bases during DNA replication (Hegde et al. 2013) and NEIL2 is expressed throughout the cell cycle and colocalizes to transcribing genes indicating it may play a role in transcription-coupled repair (D. Banerjee et al. 2011). NEIL1 and NEIL2 also interact with TDG to remove products of TET oxidation products of 5-mC (Schomacher et al. 2016).

Active removal of 5-mC through oxidation of 5-mC

5-methylcytosine is a stable DNA modification that is essential for cell differentiation. DNA methylation patterns rapidly change during gametogenesis and the rate of demethylation in maternal and paternal genomes cannot be explained by passive demethylation alone (demethylation that occurs due to DNA replication) (Rasmussen and Helin 2016). Therefore, it has been proposed that active demethylation pathways exist to control DNA methylation patterns. While the factors controlling active demethylation are not fully understood, a prevailing theory is that products of enzymatic oxidation of 5mC and 5hmC by TET enzymes are hydrolyzed and removed by the DNA glycosylase TDG (Wu et al. 2014; Wu and Zhang 2014). The consequence of 5hmC and other active demethylation intermediates, 5fC and 5caC, are still unknown, however, there is some evidence that 5fC and 5caC may play a role in gene expression independent of their role as active demethylation intermediates (Rasmussen and Helin 2016).

The TET enzyme family is comprised of three enzymes, TET1, TET2, and TET3 that catalyze the conversion of 5-mC to 5-hmC to 5-fC to 5-caC (He et al. 2011; Ito et al. 2011) (**Figure 1.3e**). TET enzymes bind to CpG sites, but do not have other sequence specificity (S. Hu et al. 2013). Some kinetic studies indicate that TET1 oxidation of 5mC to 5hmC is a fairly rapid reaction

(Ito et al. 2011; L. Hu et al. 2015). Further oxidation of 5hmC to 5fC and 5caC by TET enzymes is fairly slow (Ito et al. 2011; L. Hu et al. 2015), however, the removal of 5fC and 5caC by TDG may be quite rapid (Maiti and Drohat 2011). Together, the kinetics of TET enzymes and TDG leads to a somewhat stable accumulation of 5-hmC in the genome that has led some researchers to postulate that 5hmC may be a stable epigenetic mark (Rasmussen and Helin 2016; Ito et al. 2011; L. Hu et al. 2015).

The thymine DNA glycosylase (TDG) removes thymine from T:G mismatches, 5fC, 5caC (Maiti and Drohat 2011) and 5hmU from CpG sites (Bennett et al. 2006) (**Figure 1.3e**). T:G mismatches occur through passive or enzymatic deamination of 5mC (Métivier et al. 2008). TDG base excision repair activity is essential for development, which may indicate it plays an important role in regulating epigenetic marks during development (Cortellino et al. 2011). TDG is a monofunctional glycosylase that requires subsequent BER steps for complete removal of modified bases. Interestingly, TDG expression is negatively regulated by the DNA damage response, which may indicate that TDG expression is important for epigenetic regulation rather than DNA repair (T. Nakamura et al. 2017). The products of TET oxidation of 5mC, 5-fC and 5-caC, are efficiently removed by TDG, however, there is some evidence that the carboxyl group from 5-caC can also be removed by an unknown methyltransferase (Schiesser et al. 2012).

Oxidative DNA damage and repair in cancer

Oxidative DNA damage and repair are involved in cancer development and treatments. OGG1 is frequently lost in lung cancer and OGG1 knockout mice develop spontaneous lung carcinoma, however, double-knockout of OGG1 and MTH1 suppresses this phenotype (Kunihiko Sakumi et al. 2003). However, both OGG1 deficient lung cancer and OGG1^{-/-}/MTH1^{-/-} mice accumulate genomic 8-oxoG lesions. It is possible that the genomic stress caused by OGG1^{-/-}

/MTH1^{-/-} double-knockout leads to cell death in a MUTYH-dependent mechanism and thus prevents tumorigenesis (Nakabeppu 2014). This prediction was confirmed in a triple knockout mouse OGG1^{-/-}/MTH1^{-/-}/MUTYH^{-/-} that resulted in significant genomic 8-oxodG accumulation, a shorter lifespan, and developed of tumors in several tissues (Ohno et al. 2014).

The futile BER of A:8-oxoG repair by MUTYH may be used as a general sensor of oxidative stress and may have tumor suppression roles in cell division. In humans, mutations in MUTYH cause autosomal recessive familial adenomatous polyposis (MUTYH-associated polyposis) (Sieber et al. 2003; Al-Tassan et al. 2002). MUTYH-null mice develop spontaneous adenocarcinoma in the small intestine and colon (Sakamoto et al. 2007). Additionally, MUTYH is transcriptionally regulated by the tumor suppressor P53 (S. Oka et al. 2014) and MUTYH can act directly as a tumor suppressor by both preventing oxidative DNA mutations and by inducing cell death under high oxidative stress that is likely to lead to mutagenesis (Sugako Oka and Nakabeppu 2011).

Many cancers have altered cell metabolisms and redox regulation. Some theories indicate that during oncogenesis, cells experience an increase in ROS that creates a mutagenic environment (H. Yang et al. 2018). Additionally, alterations in cell metabolism caused by oncogenic mutations are correlated to an imbalance of radicals and antioxidants (H. Yang et al. 2018). Moreover, many cancer therapies increase cellular ROS either directly or indirectly. For example, ionizing radiation induces ROS through a variety of direct and indirect creation of oxidative radicals (Azzam, Jay-Gerin, and Pain 2012). Cisplatin and other platinum-based chemotherapies induce a mitochondrial ROS response (Marullo et al. 2013). Alkylating agents and topoisomerase inhibitors also increase cellular ROS (H. Yang et al. 2018). The toxicity of these drugs relies on the accumulation of oxidation damage throughout the cell, including oxidative DNA damage.

Recently, targeted inhibition of MTH1 has been studied as a potential chemotherapeutic. Overexpression of MTH1 reduces genomic alterations in mismatch repair-deficient cells (Russo et al. 2004). Knockdown of MTH1 increases genomic 8-oxoG and decreases survival in cancer cells (Gad et al. 2014). Several inhibitors of MTH1 also inhibit growth of cancer cells, but their mechanism of action is not yet known, although it may not rely on high intracellular ROS (van der Waals et al. 2019). Overall, MTH1 may play a role in cleansing the nucleotide pool in rapidly dividing cancer cells and inhibition of this leads to cell death (Gad et al. 2014).

Heterogeneity of oxidative DNA repair

Oxidative DNA damage is heterogeneous and repair enzymes respond to that damage through complementary and overlapping functions. Cells incur heterogeneous levels of DNA damage when exposed to bleomycin (Ostling and Johanson 1987) and malignant cells also experience heterogeneous levels of DNA damage when exposed to ionizing radiation (Olive, Banáth, and Durand 1990). In modern radiotherapy, damage to healthy tissue varies greatly between people, however, finding biomarkers to predict the toxicity has been challenging (Barnett et al. 2015). However, there is evidence that oxidative DNA repair factors are differentially expressed or active between tissue types, with age, and between individuals (Wallace 2013; Uhlén et al. 2015; Pons et al. 2010a; Chaim et al. 2017a).

Homozygous deletion in OGG1 in mice leads to accumulation of 8-oxoG lesions only in some cell/tissue types and is dependent on the age of the mice (Osterod et al. 2001). This indicates that OGG1 expression and oxidative damage may differ between tissues and change with age. In humans, repair of 8-oxoG in human fibroblasts is reduced with (Pons et al. 2010a) and polymorphisms in OGG1 have been identified as a risk for lung cancer (Z. Xu, Yu, and Zhang 2013). Mutations in MUTYH have been associated with hereditary colorectal cancer indicating

that MUTYH plays an important role in maintaining genome integrity in colorectal tissue. Polymorphisms identified in MUTYH lead to decreased or abolished glycosylase activity leading to an increase in oxidative DNA damage in these cells (Ruggieri et al. 2013). The NEIL enzymes and NTH1 have redundant substrate specificity; however, they have different tissue-specific expression patterns, cell cycle regulation, and cell localization patterns (Wallace 2013). Additionally, nuclear and mitochondrial glycosylase activity of oxidative lesions differs between tissues (Karahalil et al. 2002).

Alkylation DNA damage and repair

Sources and types of alkylation DNA damage

Alkylating agents are ubiquitous in nature and cells are constantly exposed to a number of different alkylating agents from the environment, our food, and pollutants (Fu, Calvo, and Samson 2012). Cellular processes can also produce alkylating agents. Specifically, the cellular methyl donor S-adenosylmethionine (SAM) can react with DNA leading to methylated guanine and adenine residues (Rydberg and Lindahl 1982). Alkylating agents most commonly react with the nitrogen and oxygen atoms of purines. Specifically, ~75% of alkylating DNA damage is in the form of N7-methylguanine (N7mG) (Beranek 1990). N7mG is not inherently mutagenic, but it may alter sequence recognition and binding of some DNA binding proteins (Jianli Cao and Revzin 1993; Siebenlist and Gilbert 1980). Additionally, N7mG undergoes depurination about 60 times faster than unmodified guanine residues (Hemminki, Peltonen, and Vodicka 1989), leaving abasic sites that can result in mutagenic single strand or double strand breaks. The second-most common alkylating DNA lesion is N3-methyladenine (N3mA), which accounts for 10-20% of the alkylative lesions on DNA (Drabløs et al. 2004). N3mA residues block replicative DNA polymerases and RNA polymerases (R. E. Johnson et al. 2007; Engelward et al. 1998). Finally, methylation of the

O6 position of guanine can lead to the mutagenic O6-methylguanine lesion (O6mG). O6mG readily base pairs with thymine and can cause G to A transitions during DNA replication (Loechler, Green, and Essigmann 1984). Several other alkylative lesions have been described; however, they naturally occur 100-fold less frequently than the DNA lesions described above (Fu, Calvo, and Samson 2012).

Etheno-adducts

Alkylating agents come in two flavors, the monofunctional alkylating agents that lead to the lesions described above, and bifunctional alkylating agents. Bifunctional alkylating agents can cause bulky DNA lesions, DNA crosslinks, and other biomolecule crosslinks. Environmental chemicals, like vinyl chloride, and intracellular products of ROS-induced lipid peroxidation are some examples of bifunctional alkylating agents that lead to DNA damage (Basu et al. 1993; Blair 2008). Some lesions produced by bifunctional alkylating agents are 1,N6-ethenoadenine (ϵ A), 3,N4-ethenocytosine, N2,3-ethenoguanine, and other bulky DNA lesions and crosslinks (Basu et al. 1993; Blair 2008). Etheno-DNA adducts are mutagenic since they can have ambiguous base pairing properties and they block replicative DNA polymerases causing stalled or collapsed DNA replication forks (Pourquier, Bjornsti, and Pommier 1998; Basu et al. 1993; Nair et al. 1999).

Temozolomide

Temozolomide (TMZ) is a common chemotherapeutic for treating glioblastomas. TMZ leads to methylation of O6 and N7 of guanine and N3 of adenine (Tentori and Graziani 2009). These DNA lesions are the main mechanism in which TMZ promotes cell death and of the various DNA lesions formed by TMZ, O6mG is thought to be the most toxic (Kaina et al. 1997; Roos et al. 2007; Meikrantz et al. 1998). O6mG does not halt replicative DNA polymerases, but it readily pairs with thymine and is highly mutagenic (Loechler, Green, and Essigmann 1984). O6mG:T

base pairs are recognized and removed by mismatch repair machinery creating single-strand breaks (Kat et al. 1993). High abundance of O6mG in the genome causes mismatch repair machinery to signal through ATR and P53 DNA damage response pathway and can lead to cell death (Yoshioka, Yoshioka, and Hsieh 2006; M. J. Hickman and Samson 1999). In fact, TMZ triggers apoptosis through several signaling pathways. Sensitivity to TMX is dependent on the levels of the repair factor, MGMT, cell proliferation rates, P53, and the abundance of double-strand DNA breaks (Roos et al. 2007; Fan et al. 2013; Mark J. Hickman and Samson 2004).

Deamination of purines

While purine deamination is not a consequence of alkylation DNA damage, the products of deaminated purines are repaired by the methyl-purine DNA glycosylase (MPG) and therefore are important DNA damage events to discuss within the alkylation DNA damage section.

Deamination of adenine in single stranded DNA occurs much slower than deamination of cytosine in single stranded DNA (Karran and Lindahl 1980). As free bases, both adenine and cytosine have similar deamination rates in physiological conditions (Levy and Miller 1998). Deamination of purines occurs much more rapidly under oxidative stress and in the presence of nitric oxide (Shapiro and Shiuey 1969). In cells exposed to cigarette smoke or hydrogen peroxide, deamination of adenine and guanine to hypoxanthine and xanthine respectively were among the most common DNA lesions produced (Spencer et al. 1995; Toyokuni, Mori, and Dizdaroglu 1994). Additionally, cells exposed to sodium nitrite or other reactive nitrogen species created during inflammation accumulated a significant number of deaminated purines (Spencer et al. 2000). Alterations in purine biosynthesis can also lead to an increase in deaminated purines in DNA and RNA (Pang et al. 2012). Deaminated purines are inherently mutagenic because hypoxanthine

readily base pairs with cytosine and causes A to G transitions after DNA replication. Xanthine will base pair with thymine causing G to A transitions (Kow 2002).

Repair of alkylation DNA lesions

Enzymatic and nonenzymatic removal of N7mG and N3mA

The repair of alkylation DNA lesions, like many other DNA lesions, can occur through multiple and overlapping DNA repair pathways. The removal of N7mG and N3mA is initiated by the methylpurine DNA glycosylase (MGP or AAG) (Engelward et al. 1996) or by spontaneous depurination. Both enzymatic and nonenzymatic removal of N7mG and N3mA result in the formation of an abasic site that is then repaired by subsequent BER steps as discussed above. While both N7mG and N3mA have much faster depurination than unmodified purines, *in vivo* removal of N3mA is faster than *in vitro* depurination due to MPG activity (Lawley and Warren 1976). Additionally, *in vitro* MPG rapidly removes N3mA and N7mG from duplex DNA, although MPG removal of N7mG varies *in vivo* in a sequence dependent manner, most likely due to chromatin structure (Ye, Holmquist, and O'Connor 1998).

MPG: A glycosylase with many substrates

While the methyl-purine glycosylase is named after its methyl-purine substrates, MPG has a broad range of substrate specificity (C.-Y. I. Lee et al. 2009; O'Brien and Ellenberger 2004). MPG is a monofunctional glycosylase that recognizes and removes deaminated adenine (aka hypoxanthine or deoxyinosine) and the bulkier alkylation product ethenoA (O'Brien and Ellenberger 2004) (**Figure 1.3f**). This diverse substrate recognition is abnormal for DNA glycosylases that generally recognize a single substrate or structurally similar substrates. MPG has a non-specific interaction with DNA bases and can bind both modified and unmodified purines, however, its catalytic activity is specific for modified purines (O'Brien and Ellenberger 2004).

MPG most rapidly excises hypoxanthine from DNA, which indicates that MPG may have evolved to recognize deaminated adenine (O'Brien and Ellenberger 2004; Aamodt et al. 2004). However, MPG still recognizes and excises methyl-purines like ethenoA, N3mA, and N7mG, although at a slower rate than hypoxanthine (O'Brien and Ellenberger 2004). Some evidence suggests that MPG achieves its broad substrate specificity by recognizing instability in the DNA duplex. Bases from unstable base pairs or mismatches are flipped into the recognition and catalytic pockets of MPG and will stably interact with MPG if the DNA duplex is unstable, evidenced by the fact that MPG will also remove mismatched purines at a low rate. Namely, MPG will remove adenine from the least stable mismatch, A:C, ten-fold more efficiently than A:T base pair (O'Brien and Ellenberger 2004). The pervasive theory of how MPG achieves its broad specificity is through several mechanisms. First, MPG's base-flipping mechanism is unfavorable for conical Rosalind Franklin base pairs. Second, the binding pocket of MPG actively excludes the exocyclic amino group of unmodified purines, which allows for its ability to excise a wide variety of modified purines. Since N3mA and N7mG are already chemically unstable, they're still able to be processed by MPG. Third, MPG uses an acid-catalysis that discriminates against pyrimidines (O'Brien and Ellenberger 2003). Evidence suggests that MPG has evolved to select against unmodified purines and pyrimidines, allowing it to remove a diverse array of modified bases (O'Brien and Ellenberger 2004; C.-Y. I. Lee et al. 2009; Lau et al. 2000; Connor and Wyatt 2002).

Removal of xanthine

MGP will readily remove hypoxanthine from DNA, however, it does not efficiently remove xanthine and there are currently no known mammalian glycosylase that recognize xanthine (Kow 2002). It has been hypothesized that xanthine is removed by alternative excision repair catalyzed by endonuclease V, nucleotide excision repair machinery, or is passively removed by

spontaneous hydrolysis (Kow 2002; Morita et al. 2013; S. Wang and Hu 2016; Suzuki et al. 1997).

Direct reversal of O6mG

Unlike N-alkylation products, the oxygen alkylation product, O6mG, is not removed by MPG or by other BER enzymes. O6mG is recognized and removed by the single-use methyltransferase, O6-methylguanine-DNA methyltransferase (MGMT) (Fu, Calvo, and Samson 2012) (**Figure 1.3f**, right). MGMT is the only enzyme responsible for the error-free removal of O6mG. MGMT is a single-use enzyme, meaning that one MGMT molecule removes one O6mG adduct. MGMT covalently transfers the alkyl group to a cysteine in the active site (Sedgwick et al. 2007). This mechanism leaves an unmodified guanine and an alkylated and inactivated MGMT that undergoes ubiquitin-mediated degradation (Sedgwick et al. 2007).

Alkylation DNA damage and repair in cancer

TMZ is often a very successful first-line therapy in gliomas; however, TMZ resistance is very common in recurring disease (Oliva et al. 2010). One of the main mechanisms of TMZ resistance is an increase in MGMT expression. High expression of MGMT has been measured in resistant glioma samples (Cai et al. 2005) and is a predictor of TMZ resistance (Nagel et al. 2019). MGMT expression is directly associated with TMZ resistance and methylation status of the MGMT promoter inversely correlates with expression and directly correlates with TMZ sensitivity (Christmann et al. 2010; Wick et al. 2012; Everhard et al. 2006). In addition to MGMT promoter methylation, activation of punitive MGMT enhancers by H3K4 methylation and H3K27 acetylation marks also correlate with TMZ resistance (X. Chen et al. 2018). Chromatin remodeling may be the primary mechanism in which MGMT expression increases in response to TMZ treatment. However, it has also been proposed that MGMT expression is regulated by the

transcription factor hypoxia-inducible factor 1 alpha (Cabrini et al. 2015; Persano et al. 2012). Additionally, MGMT expression may be silenced by specific micro-RNAs that are often altered in glioblastomas (Kreth et al. 2013; Cabrini et al. 2015). It is entirely possible that cancer cells regulate MGMT expression through multiple pathways depending on their mutational patterns.

In addition to TMZ, other alkylating agents are commonly used as chemotherapies including chlorambucil, cyclophosphamide, thiotepa, and busulfan. These alkylating agents produce a wide variety of alkylated DNA adducts. In contrast to the direct reversal target, O⁶mG, many of the other alkylation DNA adducts are repaired by MGP. Knockdown of MPG can sensitize cell lines to alkylation treatment (Paik et al. 2005). Additionally, overexpression of MPG also sensitizes cell lines to alkylation treatment, indicating that abasic sites and strand breaks induced by MPG removal of alkylation DNA lesions are toxic (Fishel et al. 2007; Trivedi et al. 2005). MPG also interacts with P53 and inhibits P53-dependent transcriptional activity. However, under alkylation DNA damage, MPG releases p53 (Song et al. 2012). The interaction between MPG and p53 may explain why both overexpression and knockdown of MPG sensitizes cells to alkylating agents.

Heterogeneity in alkylation DNA repair

In cancer, intratumoral heterogeneity can lead to disease recurrence and treatment resistance. In glioblastoma, evaluation of multiple biopsy samples taken the same tumor can result in different disease classifications and MGMT methylation statuses (Wenger et al. 2019). In fact, MGMT expression can be heterogeneous from the same tumor sample (Parker et al. 2016). The methylation status of the MGMT promoter is often correlated with MGMT expression and activity levels (Uno et al. 2011); however, not all of the evidence promotes that correlation (Rood, Zhang, and Cogen 2004). Additionally, MGMT is highly expressed in the liver, and in mice, the

expression of MGMT is linked to the circadian rhythm and glucocorticoid secretion (Horiguchi et al. 2010), indicating that MGMT activity is also heterogeneous in healthy tissue.

MPG activity has been associated with tumorigenesis in the thymus (N. K. Kim et al. 1998) and over expression of MPG in gliomas is associated with poor prognosis (Liu et al. 2012). MPG activity can be directly regulated by phosphorylation by ATM (Agnihotri et al. 2014). Polymorphisms in MPG have been associated with rheumatoid arthritis progression (S. Y. Chen et al. 2010; Huang et al. 2015); however, the functional consequences of these polymorphisms has yet to be determined. Additionally, systemic treatment with alkylating agents in mice can lead to retinal degeneration in a sex-dependent manner and cerebellar damage independent of sex (Allocca et al. 2017, 2019). These tissue degeneration events are dependent on a number of factors such as MPG, PARP1, and RIP3 kinase, however, the sex differences and cell-type specific cell death have currently not been explained (Allocca et al. 2017, 2019). Together, these are several examples that indicate repair of alkylation DNA damage is heterogenous in diseases and healthy tissues.

Additional DNA repair pathways

In addition to the BER pathway described above, cells have several other DNA repair pathways to repair both endogenous and exogenous DNA damage (Hakem 2008; Sancar et al. 2004). For example, nucleotide excision repair is responsible for repairing bulky lesions. Mismatch repair is responsible for removing mismatched bases incorporated during DNA replication. There are also two pathways to repair double-strand DNA breaks: homologous recombination and non-homologous end joining. The Fanconi anemia pathway repairs inter-strand crosslinks. Each of these pathways include a number of enzymes with controlled regulation and expression patterns and are critical for maintaining a stable genome. Disruption in these DNA repair pathways is common in cancers and neurodegeneration (Hakem 2008).

In addition to tightly controlled DNA repair pathways, cells also have a critical DNA damage response (DDR) pathway. The DDR toes the line between tolerating and repairing DNA damage or initiating cell death (Jackson and Bartek 2009). Many DDR proteins associate with cell cycle regulators like P53 and PARP and if the DNA damage is too great, this can lead to cell death. To avoid cell death, many cancers have mutations in DNA repair proteins or DDR factors in order to rapidly expand and avoid the DNA damage checkpoints (Torgovnick and Schumacher 2015). DNA damage chemotherapies are effective, but some rely on a functional DNA damage response to initiate cell death. Upon resistance acquisition, first line DNA damage chemotherapies are ineffective, sometimes due to an increase in DNA repair capacity or mutations in DDR regulators (Torgovnick and Schumacher 2015).

Overall, cells are poised to withstand and repair any and all sources of DNA damage through an awesomely efficient and complex network of DNA damage sensors, repair factors, and cell signaling pathways. These factors are vital to maintaining cellular homeostasis and allow for relatively error-free inheritance of genetic material.

Methods to measure DNA repair capacity

Measuring genomic DNA damage: Comet assay

One common way of measuring a cell's ability to repair its DNA (or its DNA repair capacity) is by measuring the amount of DNA damage in the cell. The alkaline comet assay (single cell gel electrophoresis) is one of the oldest (Ostling and Johanson 1984) and most common ways of measuring global DNA damage in the cell (Neri et al. 2015) (**Figure 1.4a**). Cells or nuclei are embedded in a thin layer of agarose and then lysed in an alkaline buffer. The alkaline buffer will lyse the cells, disrupt protein interactions, and degrade cellular RNAs. Upon exposure to an electric field, the remaining DNA will migrate. The DNA is stained using fluorescent nucleic acid dyes.

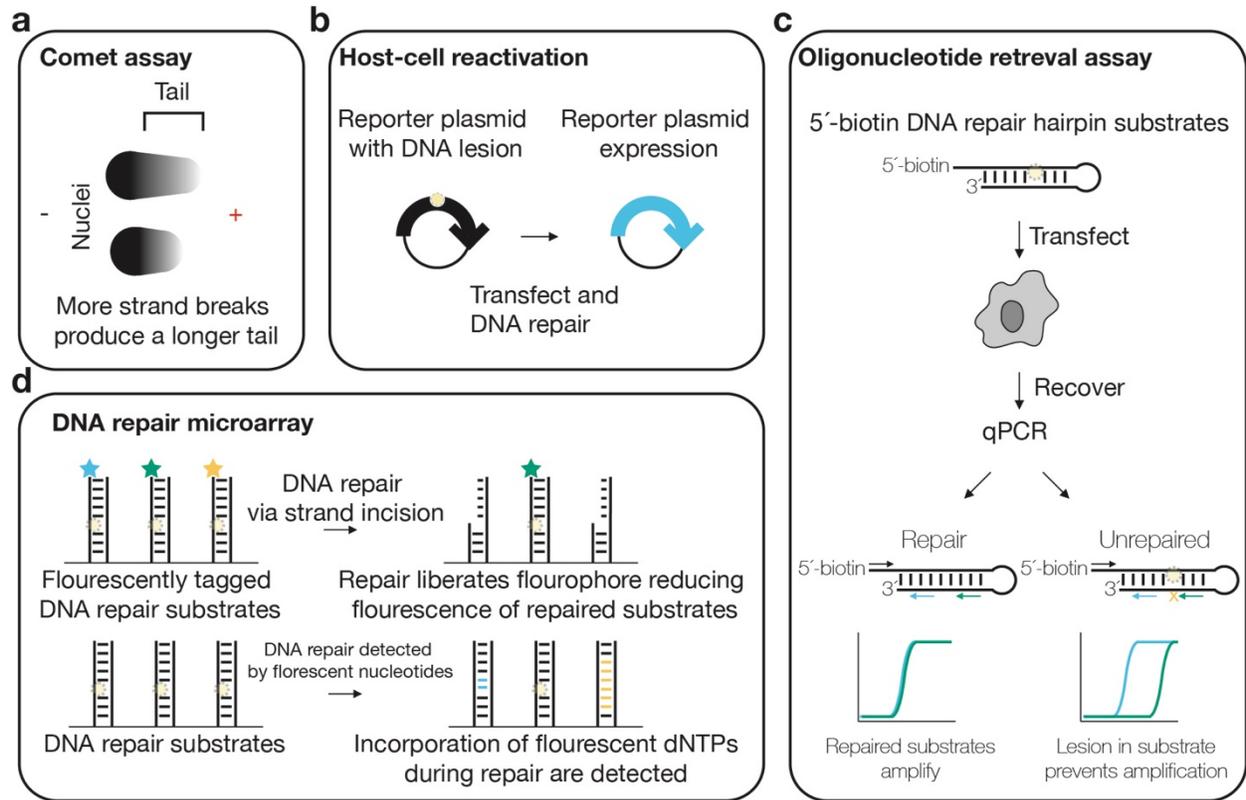


Figure 1.4: DNA repair methods

a. The comet assay measures DNA strand breaks in nuclei. The ‘tail moments’ are greater in samples with DNA damage.

b. Host cell reactivation assays use reporter plasmids with DNA damage to measure the rate of repair in living cells.

c. Oligonucleotide retrieval assays use DNA hairpin substrates to measure DNA repair in cells.

d. DNA repair microarrays use immobilized DNA repair substrates to measure DNA repair in cell-free extracts.

Nuclei with few DNA damage events will remain in a relatively stable position and maintain a circular shape. Single and double-strand DNA breaks or alkaline-labile sites will cause the DNA to migrate further in the agarose creating a comet-like tail to form in the direction of the anode. The intensity of the tail is directly proportional to the amount of DNA damage. The comet assay is sensitive enough to measure hundreds to several thousand strand breaks per cell (Andrew R. Collins 2004). To measure the abundance of specific lesions, following lysis, the embedded cells are treated with DNA glycosylases creating lesion-specific DNA breaks that are again measured by the intensity of tails produced following electrophoresis (Langie, Azqueta, and Collins 2015a).

The comet assay and its variations have been adapted and perfected in order to measure DNA repair by measuring DNA damage. For example, using time course experiments, cells can be treated with DNA damage agents and the comet assay can measure damage over time and thus get a measurement of DNA repair over time in order to comment on DNA repair capacity (Trzeciak, Barnes, and Evans 2008; S. Sauvaigo et al. 2007). The comet assay has been used to measure DNA repair capacity and heterogeneity of DNA repair capacity in cancers and healthy cells (Olive, Banáth, and Durand 1990). Additionally, DNA repair capacity can be measured by using nucleoid DNA from irradiated cells embedded in agarose as the substrate DNA. This substrate DNA can be treated with cell-free extracts supplemented with ATP and nucleotides. Over time, the tails of the substrate DNA disappear as the active enzymes in the extract repair the substrate DNA (A. R. Collins, Fleming, and Gedik 1994).

The comet assay, while crude and simple, has been invaluable in expanding our knowledge of DNA repair capacity. Variations of the comet assay have been used to measure interindividual differences in DNA repair capacity (Slyskova, Lorenzo, et al. 2014) and intratumoral heterogeneity

in DNA repair capacity (Olive, Vikse, and Banath 1996). The comet assay has also found DNA repair capacity can change with age, exposure to sunlight or exposure to cigarette smoke (S. Sauvaigo et al. 2007). Additionally, the comet assay has been used to correlate biomarker expression and DNA repair capacity in tumor biopsies in an effort to move towards precision medicine in cancer therapies (Slyskova, Langie, et al. 2014).

However, the comet assay only provides coarse-grain resolution as to the types of DNA damage events that are being repaired. Given that BER substrates are diverse in structure, consequence, and the enzymes that repair them, the comet assay does not provide insight into which of the overlapping DNA repair enzymes are responsible for repair events, nor does it provide detailed quantification as to the types of DNA damage events that are being repaired quickly or more slowly.

Measuring genomic DNA damage: Mapping

Methods that provide detailed mapping of DNA lesions throughout the genome have been recently developed. For example, Excision-seq can map uracil and pyrimidine dimers throughout the genome (Bryan et al. 2014). Hyde-n-seq (Clausen et al. 2015) and Ribose-seq (Koh et al. 2015) can map ribonucleotides in DNA. However, these methods do not have the resolution to measure these DNA damage events in cells with efficient DNA repair enzymes and they require knockout or knockdown of DNA repair factors in order to measure these damage events. Because of their lack of sensitivity, genome-wide damage measurements have not been used as a way to measure DNA repair capacity. These genome-wide mapping studies have provided information on how DNA damage occurs in cells and where DNA damage hot-spots are located throughout the genome. For example, Bryan, et. al. (2014) found that uracil did not accumulate in early and late replicating regions of the genome. The authors only measured uracil accumulation in the absence

of the uracil DNA glycosylase, indicating that this method is not suitable to measure DNA repair capacity in wildtype cells. Mapping methods are, however, useful in understanding localized DNA repair capacity throughout the genome and the factors that contribute to damage ‘hotspots’.

Measuring DNA repair in cells: Host cell reactivation

Host cell reactivation assays were developed to measure endogenous DNA repair capacity of specific DNA lesions (J. M. Johnson and Latimer 2005). There are several variations to the assay, but the principle remains the same. First, an expression plasmid - usually coding for the expression of reporter or fluorescent protein - is embedded with a single (or multiple) DNA damage event(s). Second the plasmid is transfected into live cells. Third, the expression of plasmid will increase (or decrease) upon repair of the lesion by endogenous DNA repair proteins and the fluorescence is measured over time (J. M. Johnson and Latimer 2005) (**Figure 1.4b**).

Host cell reactivation assays have been used to measure interindividual differences in DNA repair capacity across multiple DNA repair pathways (Chaim et al. 2017a). Additionally, host cell reactivation assays have been able to correlate and predict chemotherapy response based on DNA repair capacity of multiple DNA repair pathways (Nagel et al. 2017).

Host cell reactivation assays have some limitations. First, host cell reactivation assays require cells to tolerate and function following transfection, which make them difficult to use on primary tissue. However, there has been success in measuring DNA repair using primary lymphocytes by host cell reactivation (Mendez et al. 2011). Second, they are limited in the number of DNA repair pathways that they can measure simultaneously. Nagel, *et al.* (2016) was able to co-transfect four expression plasmids into cells in order to measure DNA repair across four pathways, however, this is the upper limit of what can be transfected into a single sample of cells. Finally, HCR assays depend on complete repair of the expression plasmid in order for repair to be

measured; and thus, they lack the resolution to measure repair intermediates. For example, if BER repair is slow, it is difficult to know if glycosylase activity, Ape1 activity, POLB activity, or DNA ligase activity is responsible because all steps need to occur for the readout to be measured. Although by measuring multiple reporters simultaneously, Nagel, *et al.* (2016) was able to determine the importance of Ape1 AP lyase activity over glycosylase activity in glioblastoma cells.

Measuring DNA repair in cells: Oligonucleotide retrieval assay

Another method to measure DNA repair *in vivo* is through oligonucleotide retrieval assays (J.-C. Shen et al. 2014; Golato et al. 2017). In oligonucleotide retrieval assays, DNA repair substrates in the form of biotinylated DNA hairpins are transfected into cells. The cells initiate and complete repair on the DNA hairpin substrates. Then the cells are lysed, and the DNA repair substrates are recovered using streptavidin affinity purification. Repair is quantified through quantitative PCR. The presence of DNA lesions blocks DNA polymerases and alters amplification rates of the modified substrates compared to the repaired substrate (**Figure 1.4c**). Oligonucleotide retrieval assays have been developed for NER substrates (J.-C. Shen et al. 2014) and BER substrates (Golato et al. 2017). Oligonucleotide retrieval assays have shown differential DNA repair rates between cell lines (J.-C. Shen et al. 2014), however, their utility is limited to cells that tolerate transfection and substrates that alter DNA polymerase amplification rates.

Measuring DNA repair in cell extract: DNA repair microarray

DNA repair microarrays use cell-free extracts to measure strand incision events on DNA repair substrates. Plasmid DNA (Millau et al. 2008) or small oligonucleotide substrates with fluorescent tags (Pons et al. 2010a) are immobilized on microarray slides. The slides are then treated with cell free extracts and fluorescence of the array spots is measured. When DNA repair proteins present in the extract excise the damage base and create a strand break, this causes the

fluorophore on the end of the substrate to wash away (Pons et al. 2010a). Alternatively, fluorescently labeled nucleotides are added to the cell extract and when the nucleotides are incorporated into the repaired substrate, the fluorescence increases as repair occurs (Millau et al. 2008). The signal can be measured overtime and is dependent on DNA repair enzyme concentrations (**Figure 1.4d**). The microarray assays have been used to measure interindividual differences in DNA repair capacity across multiple DNA repair pathways. Additionally, they have shown that DNA repair capacity can change with age (Pons et al. 2010a) and how functional DNA repair capacity can be used to understand sensitivity to DNA damage chemotherapies (Forestier, Sarrazy, et al. 2012). DNA repair microarrays have also shown how endogenous and exogenous factors can impact DNA repair capacity (Forestier, Douki, et al. 2012a; Candéias et al. 2010).

DNA repair microarrays require bulk cell extract and thus are unable to measure DNA repair heterogeneity within a single sample. Additionally, DNA repair microarrays cannot measure repair at nucleotide resolution on the substrates. Nucleotide resolution can provide insight on which repair proteins are initiating repair (e.g. RNaseH2 or Top1).

Single cell methods

Single cell methods have rapidly expanded in the last 10 years. Methods that measure aspects of dozens of cells have been replaced with high throughput methods that can measure gene expression (Macosko et al. 2015; G. X. Y. Zheng et al. 2017; Junyue Cao et al. 2017), chromatin accessibility (Cusanovich et al. 2015), and protein abundance (Mimitou et al. 2019) in thousands to millions of cells in one experiment. These single-cell methods employ a number of different techniques to isolate single cells and tag or capture the molecules of interest; and thus, these methods have been widely successful in measuring the abundance of biomolecules in the cell.

Single-cell methods have greatly expanded our understanding of cell heterogeneity, complex tissues, and the development of organs and whole organisms (Stuart and Satija 2019a).

Single-cell mRNA sequencing

Some of the first high-throughput single-cell methods measured gene expression in single cells. Using microfluidic devices, these methods have accelerated the field of single-cell mRNA sequencing. The central idea of microfluidics-based single-cell RNA sequencing methods is that single cells are isolated and lysed in oil-water emulsion droplets (Macosko et al. 2015; G. X. Y. Zheng et al. 2017) (**Figure 1.5a**). Within each droplet, the mRNAs of each cell are uniquely barcoded with a drop- (and cell-) specific barcode. All the mRNAs present within a single droplet get tagged with the same cell barcode and a unique molecular identifier (UMI). Single-cell suspensions and oligo-dT primers that contain unique barcodes are run through different channels of a microfluidics chip, combining and forming droplets in oil. Upon droplet formation, the cells lyse and the polyadenylate 3' ends of the mRNA anneal to the oligo-dT primer and are reverse transcribed (G. X. Y. Zheng et al. 2017). Following reverse transcription, the emulsion is broken, and nucleic acids are isolated. Since each cell was uniquely barcoded in the drop, the remaining mRNA sequencing library preparation can be done in bulk. Since the majority of the library preparation is done in bulk, this method has greatly simplified our ability to study single cells. Similar approaches have been successful, but instead of isolating single cells in oil-water emulsions, cells are sorted into microwells and are uniquely barcoded in each well (Picelli et al. 2014) (**Figure 1.5b**).

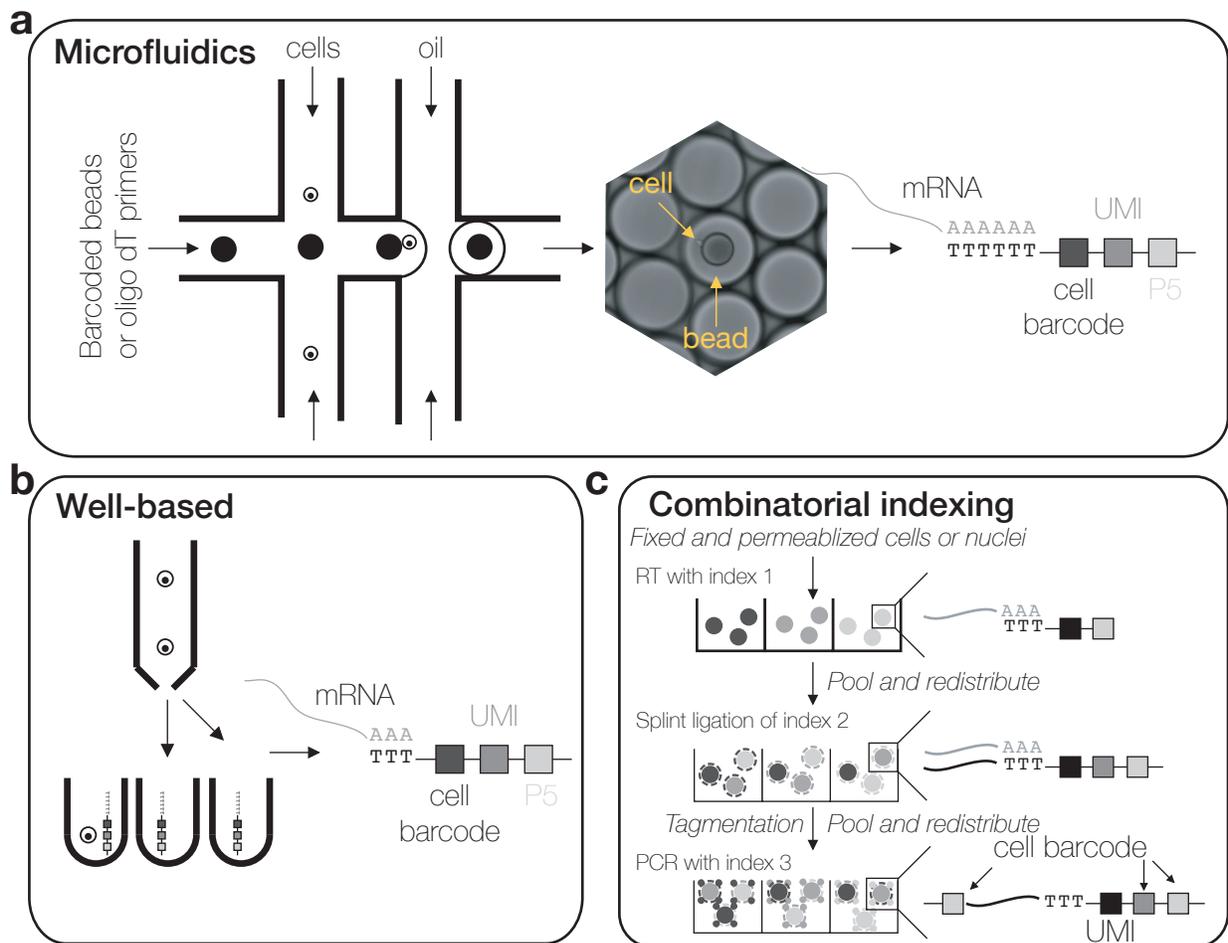


Figure 1.5: Single-cell mRNA sequencing methods

a. Droplet-based single-cell mRNA sequencing methods encapsulate single-cells and barcoded oligo-dT primers in oil-water emulsions. mRNA from single-cell is uniquely barcoded within each drop allowing the researcher to identify and quantify mRNAs from single-cells.

b. Well-based single-cell mRNA sequencing methods isolate single-cells in wells. Within each well, cells are lysed, and mRNAs are uniquely barcoded within each well.

c. Combinatorial indexing single-cell mRNA sequencing methods uniquely barcode mRNA within fixed cells or nuclei using a series of RT, ligation, and PCR reactions in wells that contain dozens of cells. Between rounds of barcoding, cells are pooled and redistributed to ensure each cell receives a unique barcode.

There are two main single-cell mRNA sequencing approaches. One, explained above, involves live, viably frozen or fresh fixed cells where individual cells are isolated and barcoded individually. Another method developed separately by the Shendure and Seelig labs, involves barcoding individual cells through a series of split-pool barcoding reactions (Junyue Cao et al. 2017; Rosenberg et al. 2018) (**Figure 1.5c**). Cells are first fixed and permeabilized and then undergo a series of barcoding reactions. Cells are distributed into 96 or 394-well plates where each well has a uniquely barcoded RT primer. Since cells are fixed, mRNA present within each cell is immobilized and locally reverse transcribed within each fixed cell or nucleus. Next, cells are pooled and redistributed into wells containing a unique barcoded sequence that is ligated onto the 5' end of the cDNA products. The cDNA is barcoded a third and final time during the PCR step where the sample barcode present within the sequencing primer becomes the third and final cell barcode. Through a series of two or three barcoding steps, millions of cells can be uniquely barcoded with very little overlap.

In addition to single-cell mRNA sequencing methods, a number of imaging methods can track mRNAs and proteins at single-cell resolution. Because most imaging techniques rely on *in situ* hybridization of fluorescently tagged probes, there is a limit to how many fluorophores can be detected simultaneously and the number of repeat staining steps each sample can incur before degradation (Moffitt et al. 2016). However multiplexed *in situ* hybridization methods have been developed that significantly increase the number of targets that can be measured (Eng et al. 2019). Through a combination of imaging and sequencing approaches, spatial transcriptomics of tissue slices has approached single-cell resolution (Ståhl et al. 2016; Moffitt et al. 2018). Imaging techniques are inherently single-cell resolution and the throughput of these methods is continuing to increase. As imaging techniques improve and are combined with RNA sequencing, they have

reached the depth of information provided by single-cell RNA sequencing techniques (i.e. thousands of measurements per cell) and are approaching subcellular spatial resolution of imaging techniques (Rodrigues et al. 2019; Eng et al. 2019).

Multimodal single-cell methods

In addition to single modality measurements in single cells, there are some methods that simultaneously measure gene expression and something else. For example, CITE-seq (Stoeckius et al. 2017) measures cell surface markers and mRNA expression simultaneously. To achieve this, antibodies are conjugated with oligonucleotides that contain a 3' poly-A tail, a 5' PCR handle, and an internal antibody-specific sequence tag. Cells are stained with the antibodies prior to a droplet-based poly-A single-cell sequencing protocol. The tagged antibodies bind to cell surface receptors are captured alongside the cellular mRNA and amplified during the single-cell mRNA sequencing library protocol. Similar protocols have been used to pool samples, prior to single-cell protocols, cells are tagged using sample-specific barcodes conjugated to antibodies (Stoeckius et al. 2018) or lipids (McGinnis et al. 2019).

The CITE-seq concept was expanded in a similar ECCITE-seq protocol (Mimitou et al. 2019). ECCITE-seq uses the theory of CITE-seq but expands it to also capture CRISPR guides by including a reverse transcription binding site within the CRISPR guide-RNA transcript. Additional methods have been made to measure mRNA expression in combination with proteins, DNA sequence, chromatin accessibility, and DNA methylation status (Stuart and Satija 2019a).

Cells of a multicellular organism have the same genetic material, but through a number of regulatory networks achieve different cell functions. Chromatin marks and DNA methylation can lead to differential gene expression that can be related to cell function. Additionally, post-transcriptional modifications, transcription factor binding sites, and cis-regulatory sequences can

also affect cell function. With the development of multimodal single cell analysis, we can begin to establish cell type-specific regulatory networks. Because single-cell measurements provide the throughput and resolution to understand cell type-specific characteristics, their development has accelerated our understanding of cell regulatory mechanisms during disease progression and development (Schier 2020).

Computational methods

In parallel to developing molecular biology techniques to measure single-cell characteristics, computational methods have grown to help us understand biology at single-cell resolution. Some computational methods have been developed to deal with the challenges of single-cell biology. For example, there are strong batch effects in single-cell RNA sequencing. This creates a challenge when trying to make conclusions across samples and/or technology. Several methods have been developed to integrate data across datasets that regress out the differential gene expression related to batch effects (Butler et al. 2018a; Stuart et al. 2019a; Tran et al. 2020).

Other computational methods attempt to help us understand the biological consequences single-cell differential gene expression. For example, during development cells differentiate into specific cell types. In order to understand which cells arise from which progenitor cells, computational methods have been developed to predict lineages and create a pseudo-time measurement that can track differentiation of cells through intermediate stages of differentiation (Trapnell et al. 2014; Qiu, Mao, et al. 2017). These methods have been used to track the development of organs, organoids, and even whole organisms (Stuart and Satija 2019a).

Finally, data integration methods have been developed in order to predict phenotypes from single-cell measurements. For example, Stuart, *et al.* (2018) developed a method that can integrate

single-cell or bulk datasets into one to predict cell types or phenotypes. Multimodal single-cell analysis methods that integrate multiple measurements per cell provide a detailed framework to understanding how individual measurements contribute to cell phenotype (Butler et al. 2018a; Stuart et al. 2019a).

Challenges of single-cell methods

While both molecular biology and computational methods have greatly improved the field of single-cell biology, there are still several challenges that the field faces. First, because the raw material of single cells is minimal, the data produced by single-cell methods is inherently sparse. This creates a unique challenge in understanding the true negative and false negative rates of detection. Some data imputation methods have been developed to address the sparsity of the data (Mongia, Sengupta, and Majumdar 2019; Jin et al. 2020); however, these methods may lead to overinterpretation of the data or over correcting by masking true negatives. Second, the variation in single-cell data is difficult to explain. Variation in gene expression or other measurements could be due to biological differences between cells or produced by technical variation intrinsic to the methods. While technical vs biological noise is an important consideration for all molecular biology techniques, single-cell biology has a unique challenge since technical replicates for individual cells is not possible with the current techniques. Lastly, current single-cell methods rely on measuring the abundance of biomolecules in cells and use these measurements to predict phenotypes. The complex nature of cell regulatory pathways makes it difficult to accurately assign cell phenotypes using abundance measurements alone. To this end, we have developed an assay to simultaneously measure mRNA expression and enzyme function in single cells (**Chapter II**).

Summary

Here I have summarized the workings of DNA repair and single-cell methods. DNA repair is important for maintaining cellular homeostasis; however, DNA repair is not static. DNA repair differs between individuals, cell types, cell cycle phases, and can change with age and with disease. Numerous methods have been developed to measure DNA repair enzyme activities in live cells and in cell extracts. These methods have helped us understand the importance of DNA repair in cancer development, treatment, and the heterogeneous nature of DNA repair in tumors and healthy individuals. With the acceleration of the single-cell genomics field, our understanding of cell heterogeneity has grown immensely. However, the single-cell field has been focused on measuring the abundance of molecules in cells and using that data to predict cell phenotypes. Cell phenotypes can be affected by many different modifications to mRNA and proteins that are not measured using current single-cell technology. To help bridge the gap between the current DNA repair fields and single-cell biology fields, I have developed a method to measure DNA repair at single-cell resolution. My findings are reported in the following chapters. Enjoy. ☺

CHAPTER II

SIMULTANEOUS MEASUREMENT OF BIOCHEMICAL PHENOTYPES AND GENE EXPRESSION IN SINGLE CELLS¹

Abstract

Methods to measure heterogeneity among cells are rapidly transforming our understanding of biology but are currently limited to molecular abundance measurements. We developed an approach to simultaneously measure biochemical activities and mRNA abundance in single cells to understand the heterogeneity of DNA repair activities across thousands of human lymphocytes, identifying known and novel cell-type-specific DNA repair phenotypes.

Introduction

New methods to study heterogeneity at cellular resolution measure differences in gene expression (Macosko et al. 2015; Junyue Cao et al. 2017, 2019; G. X. Y. Zheng et al. 2017), chromatin accessibility (Cusanovich et al. 2015), and protein levels (Mimitou et al. 2019) across thousands to millions of cells to understand developmental trajectories of tissues, tumors, and whole organisms. But these methods only measure static levels of DNA, RNA, and proteins, limiting our ability to extract dynamic information from individual cells.

We developed a functional assay as a new modality for single-cell experiments. Our key innovation is that, instead of measuring the abundance of molecules—i.e., levels of DNA, RNA, or protein—from single cells and predicting functional states, we directly measure enzymatic activities present in single cells by analyzing the conversion of substrates to intermediates and products in single-cell extracts within a high-throughput DNA sequencing experiment. Our approach is compatible with existing platforms that measure gene expression at single-cell

resolution and can measure many different enzymatic activities simultaneously, querying different biochemical activities by combining unique substrates.

We measured DNA repair activities in single cells because the enzymatic substrate (i.e., a DNA lesion to be repaired by cellular enzymes) yields a product that can be directly analyzed by DNA sequencing. DNA damage is repaired by multiple different and often redundant pathways including base excision repair, nucleotide excision repair, mismatch repair, and direct reversal (Bauer, Corbett, and Doetsch 2015a). Current methods to study DNA repair in cells and cell extracts use synthetic DNA substrates to measure repair activities (J.-C. Shen et al. 2014; Nagel et al. 2014), but these approaches do not scale to multiple measurements (i.e., gene expression and biochemical activities) from the same cell, and their reliance on substrate transfection precludes facile application to primary cells.

We were able to show that we could measure base excision repair and ribonucleotide excision repair at single-cell resolution. The repair signals we measured were stable across time and substrate concentration. We applied our method to primary human lymphocytes and measured known and novel differences in DNA repair across cell types. Our method is the first of its kind in measuring gene expression and DNA repair activities in single-cells and could be broadly applied to many cell types and primary samples.

Materials and methods

DNA repair substrates for single cell experiments

Oligonucleotides were purchased from IDT (**Table 2.1**). Substrates contain a 5' and 3' C3 spacer to prevent exonuclease degradation and reverse transcriptase extension of the substrates. Hairpins were gel purified prior to use in single cell experiments. Briefly, 2-5 nmoles of hairpins were loaded in denaturing buffer (47.5% formamide, 0.05% Orange G) on 8% 19:1 acrylamide

(BioRad) TBE-Urea gels (7 M urea, 0.1 M Tris base, 0.1 M boric acid, 2 mM EDTA). Hairpins were visualized with UV shadowing on a TLC Silica gel 60 F₂₅₄ plate (Millipore), cut from the gel, crushed in a 1.5 ml Eppendorf tube and eluted in 400 µl 0.3 M sodium acetate overnight at 37 °C shaking at 400 RPM. Acrylamide was removed using 0.45 µm cellulose acetate filters (Costar). Hairpins were then purified via ethanol precipitation and resuspended in water. The concentration of purified hairpins was determined via absorbance at 260 nm on a Nanodrop 2000 (Thermo Scientific).

Preparation of single cell suspensions

Single cell suspensions from cell lines were prepared according to 10x Genomics guidelines. Briefly, cells were quickly washed with 0.25% trypsin (ThermoFisher) and then incubated in 0.25% trypsin for 5 minutes at 37 °C. Trypsin digestion was quenched by the addition of cell culture medium. Cells were isolated by centrifugation at 150 xG for 3 minutes (these same conditions were used for all cell washes). For cell mixing experiments, approximately 10⁶ cells from each knockout cell line (UNG^{KO} or RNASEH2C^{KO}) were filtered through a 30 µm strainer and mixed in the same tube. Cells were washed twice with cold PBS containing 0.04% BSA. Cells were resuspended in 500 µl PBS with 0.04% BSA and filtered through a Flowmi™ Tip Strainer. Cells were stained with trypan blue and counted on a hemocytometer. Cell concentration ranged from 400 - 1000 cells per µl and viability was between 80-95%.

Fresh peripheral blood mononuclear cells (PBMC) were isolated from whole blood donated by healthy human donors according to University of Colorado IRB guidelines in sodium heparin tubes. Approximately 5-10 ml of whole blood was diluted with PBS to a total volume of 35 ml. Diluted whole blood was layered over 10 ml Ficoll-Paque PLUS (GE) and centrifuged at 740 xG for 20 minutes with no deceleration. Cells located above the Ficoll layer were removed and washed

twice with PBS. Cells were counted and approximately 2 million cells were washed an additional two times with PBS plus 0.04% BSA. Cells were resuspended in 500 μ l PBS plus 0.04% BSA and run through a Flowmi™ Tip Strainer. Cells were counted on a hemocytometer: cell concentration ranged from 400-1000 cells per μ l and viability was between 80-95%.

Single cell repair measurements using the 10x Genomics platform

The most current version of this protocol is available at: <https://dx.doi.org/10.17504/protocols.io.uhyet7w>.

Cells were loaded onto the 10x Genomics single cell 3' expression kit V2 according to the manufacturer's instructions (CG 000075 Rev C) with the following exceptions:

1. When preparing the single cell master mix, 5 μ l was subtracted from the appropriate volume of nuclease-free water. After the nuclease-free water was added to the master mix and prior to the addition of the single cell suspension, 5 μ l of mixed DNA repair substrates were added (see **Table 2.2** for substrate concentrations for each experiment).
2. The GEM-RT incubation was changed to the following:
 1. Lid temperature: 53 °C
 2. 37 °C for 60 minutes (unless otherwise noted in experiment **Figure 2.6**)
 3. 53 °C for 45 minutes
 4. 4 °C Hold and proceed directly to GEM-RT cleanup
3. After GEM-RT cleanup, DNA repair substrates and products were separated from mRNA prior to cDNA amplification. 0.6x volume of AmpureXP was added to the eluted RT products (21 μ l AmpureXP to 35 μ l RT product) and incubated for 5 minutes at room temperature. The sample was placed on a magnetic strip (High on 10x Magnetic Separator) until the liquid was clear. The supernatant was transferred to a new tube since it contained

the DNA repair substrates and products. The beads containing the RT products were washed twice with 150 μ l of 80% ethanol, then dried for 2 minutes at room temperature, and eluted in 35.5 μ l of Elution Solution 1 according to 10x SPRIselect cleanup protocols. This fraction was used to prepare the mRNA expression library according to manufacturer's instructions. The supernatant was cleaned up by added 1.8x of the original volume of AmpureXP (42 μ l), mixed, and then incubated for 5 minutes at room temperature. The sample was placed on a magnetic strip until the liquid was clear. The supernatant was discarded, and the beads were washed twice with 150 μ l 80% ethanol. The beads were dried at room temperature for 2 minutes, then eluted in 20 μ l water. This fraction was used to prepare the DNA repair libraries. Note: DNA repair substrates *may* be visible on Tapestation or Bioanalyzer prior to or following size separation, however, this was not measured in these experiments.

Preparation of DNA repair libraries from single cells

The DNA repair libraries were prepared with the following steps:

1. *End repair*: 20 μ l of the purified DNA repair libraries were added to an end repair reaction with a total volume of 30 μ l (NEBNext End repair Module E6050) and incubated for 30 minutes at 20° C.
2. *Clean up by precipitation*: 120 μ l of 0.3 M sodium acetate and 400 μ l 100% ethanol were added to the end repair reaction (step 1). The reaction was allowed to precipitate at -20 °C for at least 30 minutes. Samples were centrifuged at 10,000xg for 10 minutes and the supernatant was removed, and the pellet was washed with 500 μ l of 70% ethanol and centrifuged at >10000 xG for 10 minutes. The supernatant was removed ,and the pellet was

dried for 2 minutes at room temperature. The pellet was resuspended in 20 μ l nuclease-free water.

3. *A-tailing*: 15 μ l of the end repaired DNA repair library was added to an A-tailing reaction with a total volume of 20 μ l (1x Blue Buffer (Enzymatics), 400 μ M dATP, 5 units Klenow 3'-5' exo- (Enzymatics)) for 30 minutes at 37 $^{\circ}$ C. The A-tailing reaction was cleaned up using precipitation as in step 2.
4. *Adapter ligation*: 13 μ l of the A-tailing reaction was added to an Illumina Y adapter ligation reaction with a total volume of 20 μ l (66 mM Tris-HCl, 10 mM MgCl₂, 1 mM DTT, 1 mM ATP, 7.5% PEG 6000, pH 7.6, 0.3 μ M annealed Y adapters, 600 units Rapid T4 DNA Ligase (Enzymatics)) and incubated at 25 $^{\circ}$ C for 30 minutes. The ligation reaction was purified using 1.8 x volume of Agencourt AMPure XP (Beckman Coulter) beads as described by the manufacturer. The reaction was eluted in 20 μ l nuclease-free water.
5. *Illumina TruSeq PCR*: 13 μ l of the purified ligation reaction was added to a PCR reaction with a total volume of 50 μ l (1x Phusion HF buffer (NEB), 200 μ M dNTPs, 0.6 μ M ILMN PCR primers (F and R), 2 units Phusion High Fidelity DNA polymerase). 14-20 cycles of PCR were done with 98 $^{\circ}$ C melting temperature for 15 seconds, 65 $^{\circ}$ C annealing temperature for 15 seconds, and 72 $^{\circ}$ C extension temperature for 15 seconds.
6. *PCR cleanup and sequencing*: The DNA repair library was purified using 1.2x volume Agencourt AMPure XP (Beckman Coulter) beads as described by the manufacturer. The DNA repair library was quantified using the Qubit HS dsDNA fluorometric quantitation kit (Thermo Scientific). 1 μ l of the DNA repair library was analyzed on the Agilent D1000 TapeStation. The DNA repair library was ~230-250 base pairs. The DNA repair library was paired end sequenced on a NovaSeq 6000 system with 2x150 base pair read lengths at the

University of Colorado Anschutz Medical Campus Genomics and Microarray core. Each library was sequenced with at least 10 million reads per sample.

Single cell data processing

Data processing scripts are available at <https://github.com/hesselberthlab/sc-haircut>. Briefly, FASTQ files from the 10x mRNA libraries were processed using the cellranger count pipeline (v3.0.2). Reads were aligned to the GRCh38 reference. For the repair libraries, the cell barcodes and UMIs were extracted from R1 using `umi_tools` (Smith, Heger, and Sudbery 2017a). All of the known 10x cell barcodes were provided as the whitelist. R2 was trimmed to remove the 3' polyA sequence and the 5' template switching sequence. Then R2 was aligned to a hairpin reference fasta file using `bowtie2` (v2.3.2)(Langmead and Salzberg 2012a), no reverse complement alignment was allowed to ensure sequences aligned in the correct orientation to the reference. The chromosome (same as substrate name) and 5' alignment position were concatenated and added to the bam file in the XT flag. UMIs were grouped and appended to the BAM files as a tag using `umi_tools group`. UMIs were counted per cell per hairpin position using `umi_tools count`. The table output was converted into a sparse matrix and filtered by matching cell barcodes found in the cellranger filtered feature matrix output using functions in the `scrunchy` R package (<https://github.com/hesselberthlab/scrunchy>).

Seurat

Downstream analysis of RNA and repair data was performed using the Seurat R package (v3.0.0)(Stuart et al. 2019a). Raw, filtered counts for repair was added to the same Seurat object as gene expression. Gene expression counts and repair counts were log normalized (`LogNormalize`) where feature counts for each cell are divided by the total counts for that cell and multiplied by a scaling factor (10^4) and then natural-log transformed. PBMC samples were filtered

for number of genes per cell >150-200 and <2000-2500 and for percent mitochondrial reads <15%-25%. Gene expression data was scaled and centered (ScaleData). 5000 variable features (FindVariableFeatures) were used for PCA calculation (RunPCA) and the first 10-20 principal components were used to find clusters (FindNeighbors, FindClusters) and calculate uniform manifold approximation and projection (UMAP) (RunUMAP). Cell types were identified using the Seurat functions FindTransferAnchors and TransferData(Stuart et al. 2019a). Reference PBMC data was downloaded from Seurat vignette and used as reference for PBMC cell types (https://satijalab.org/seurat/v3.1/pbmc3k_tutorial.html; <https://support.10xgenomics.com/single-cell-gene-expression/datasets/1.1.0/pbmc3k>). Cells were filtered to exclude platelets unless otherwise noted (**Figure 2.11**). Significant differences and fold changes in repair activities and gene expression between cell types were calculated using Wilcoxon Rank Sum test (FindMarkers, FindAllMarkers) for all pairwise combinations (**Table 2.3** and **Table 2.4**). PBMC replicates were merged and integrated using Seurat functions (FindIntegrationAnchors and IntegrateData) and then analyzed the same as individual replicates). In the cell mixing experiment (**Figure 2.1**), cell types were determined by repair activities. Knockout cells were identified if counts at the repair site (position 44 for ribonucleotide and position 45 for uracil) for one repair activity was > 5% of the maximum for the repair activity and the other was < 5% of the maximum. If both repair activity counts were >5% of the maximum, that cell was considered a doublet and if both repair activity counts were <5% of the maximum that cell was not classified. Filtered single cell gene expression matrices from previously published data(G. X. Y. Zheng et al. 2017) for **Figure 2.10** were downloaded from 10x Genomics (https://support.10xgenomics.com/single-cell-gene-expression/datasets/3.0.2/5k_pbmc_v3_nextgem, https://support.10xgenomics.com/single-cell-gene-expression/datasets/3.1.0/5k_pbmc_protein_v3, https://support.10xgenomics.com/single-cell-gene-expression/datasets/3.1.0/5k_pbmc_protein_v3,

cell-gene-expression/datasets/3.0.0/pbmc_10k_v3, <https://support.10xgenomics.com/single-cell-gene-expression/datasets/2.1.0/pbmc4k>) and analyzed the same as above.

Genome coverage

To calculate genome coverage for cell mixing experiment (**Figure 2.5**), BAM files produced by cellranger were split into cell type (as assigned above and in **Figure 2.1**) specific BAM files by cell barcodes using samtools view (v1.9)(H. Li et al. 2009a). Bulk genome coverage was calculated for UNG^{KO} or RNASEH2C^{KO} cells using bedtools genomecov (v2.26.0)(Quinlan and Hall 2010a). Coverage was visualized with the UCSC Genome Browser(Kent et al. 2002).

Cell type classification with repair data

To determine whether DNA repair activities are useful in classifying PBMC cell types, we used the PBMC replicates 1 and 2 (**Figure 2.8** and **Figure 2.9**, left). True cell types were determined using reference PBMC data from 10x Genomics as described in Seurat analysis section above. All other cell type classifications were compared to these reference cell types. Next, cell types were determined by renaming the defined Seurat clusters as the majority cell type present within each cluster (**Figure 2.11b**). Additionally, mRNA (**Figure 2.11c**) and/or repair data (**Figure 2.11d,e**) from PBMC replicate 1 was used as the reference input for assigning cell types to PBMC replicate 2 using Seurat's FindTransferAnchors and TransferData functions. Cell types were randomly reassigned using R (sample) (**Figure 2.11f**). True positive, true negative, false positive, and false negative numbers were calculated for each cell type for each classification method. These numbers were then used to calculate true positive rate and false positive rate for each cell type for each classification method (**Figure 2.11g**).

Bulk RNA-seq from tissues analysis

Data was downloaded from the Human Protein Atlas (https://www.proteinatlas.org/download/rna_tissue_consensus.tsv.zip) for the RNA consensus tissue gene data set. These data are gene transcript expression levels from 74 tissues. The normalized expression values in the data are the maximum expression value from 3 different sources. These data were visualized using the R package ggplot2 (**Figure 2.10**).

Empty drops vs cells analysis

To determine biological vs background activity in single cells, we calculated hairpin coverage in empty drops (**Figure 2.3**). Empty drops were determined by filtering out cell-associated barcodes from the repair matrix. The resulting repair matrix contains many barcodes that are associated with only a single UMI, so the matrix was filtered by descending UMI counts to the same number as cell-associated barcodes. This repair matrix from empty drops was used as the input to calculate empty drop signal across the hairpin by calculating the mean count across all drops at each hairpin position.

Haircut signal detection sensitivity

To determine the lower limit of haircut signal detection suitable for classifying cells as either UNG^{KO} or $RNASEH2C^{KO}$ read alignments were randomly downsampled using samtools view (v1.9)(H. Li et al. 2009a). The downsampled BAM files were then processed using the haircut single cell processing pipeline to produce haircut signal matrices. Cells were classified as either UNG^{KO} , $RNASEH2C^{KO}$, doublets, or low signal using the same cutoffs described in the Seurat analysis methods.

Estimated recovery of substrates

To determine the proportion of substrates recovered from each droplet (**Figure 2.7**), the total number of hairpins added to each drop was estimated using the following assumptions:

1. When the oil-water emulsion droplet is formed within the 10x Genomics Chromium chip, the volume of the master mix is doubled, thus reducing the concentration of hairpins to $\frac{1}{2}$ the concentration of hairpins in the master mix.
2. Diameter of each drop: $\sim 80 \mu\text{m}$
3. Volume of each drop: $\sim 270 \text{ pL}$

The number of substrates in each drop was then calculated and ranged from 40,000 to 800,000 hairpin molecules per drop depending on the hairpin concentration in the master mix. To determine the proportion of hairpins recovered per drop, the total number of aligned reads per hairpin per drop was divided by the approximate number of hairpins per drop.

Direct reversal substrate and 5' biotin cleanup

To measure repair of direct reversal substrates we included O⁶methyl-G substrate in PBMC experiments (**Figure 2.3** and **Figure 2.9**). Prior to end repair, the repair fraction was digested with PstI (NEB, 20 U in 1x Cutsmart buffer) at 37 °C for 60 minutes. The reaction was cleaned up by precipitation, followed by the above steps starting at end repair. To remove background signal on the 5' end of the substrates, we included a uracil substrate with a 5' biotin in PBMC experiments (**Figure 2.3** and **Figure 2.9**). To remove uncleaved substrate, prior to end repair, the repair fraction was incubated with Dynabeads™ M-270 Streptavidin (5 μg , Thermo) for 5 minutes at room temperature. Following incubation, the beads were discarded and the supernatant was cleaned with precipitation. The remainder of the protocol proceeded starting from end repair.

Oligonucleotides for repair libraries

Other oligonucleotides used in the library preparation can be found in **Table 2.1**. To anneal Y adapters, 100 μM Y adapter 1 and 2 were mixed in equimolar concentration in 10 mM Tris-HCl pH 7.5, 50 mM NaCl and heated to 95 $^{\circ}\text{C}$ and cooled to 4 $^{\circ}\text{C}$ over 5 minutes. Annealed adapters were diluted to 10 μM final concentration in cold 10 mM Tris-HCl pH 7.5, 50 mM NaCl. Annealed adapters were stored at -20 $^{\circ}\text{C}$ until use.

Cell lines and cell culture

Hap1 UNG^{KO} (HZGHC001531c012) and RNASEH2C^{KO} (HZGHC004633c003) cells were purchased from Horizon Discovery. Cell lines were cultured in IMDM (Gibco, purchased from ThermoFisher) supplemented with 10% FBS (ThermoFisher) and Penicillin-Streptomycin (ThermoFisher) at 37 $^{\circ}\text{C}$ with 5% CO₂.

RT-qPCR

Total RNA from cells was isolated using TRIzol reagent (ThermoFisher) according to manufacturer's instructions. Total RNA (5 μg) was treated with TURBO DNase (2 U) (ThermoFisher) according to manufacturer's instructions. Following DNase treatment, 1 μg of total RNA was reverse transcribed using Superscript II (200 U, ThermoFisher) and random hexamers primers (0.5 μM , ThermoFisher) to make cDNA. The cDNA was then used for quantitative PCR (qPCR) using Sso Advanced Universal SYBR Green Supermix (Bio Rad) and cycled on a Bio Rad C1000 384-well thermal cycler and plate reader. qPCR experiments were done in technical triplicate and biological duplicates.

Table 2.1: Oligonucleotides for single cell DNA repair measurements

Oligo	Sequence	Source
A:U - 1	/5SpC3/GTCGTGATGCATGCCTGTATGTGACACAAGTAATTGTGTCACAUCACAGGCATG CATCACGACAAAAAAAAAAAAAAAAAAAAAAAAA/3SpC3/	IDT
C:riboG - 1	/5SpC3/ACTCGAGTCACACTCGTACTGATGCATGAGTAATCATGCATCrGTACGAGTGT GACTCGAGTAAAAAAAAAAAAAAAAAAAAAAAAA/3SpC3/	IDT
G:U	/5SpC3/TGAATTCGAGAGTCGTTCCGGCGATATAACGTAAGTTATATCGCUGAACGACTC TCGAATTCAAAAAAAAAAAAAAAAAAAAAAAAA/3SpC3/	IDT
G:Abasic	/5SpC3/ACGTACGTTAGCATAACTGTAATCTTAATGTAATTAAGATTA/idSp/AGTTATG CTAACGTACGTAAAAAAAAAAAAAAAAAAAAAAAAA/3SpC3/	IDT
C:I	/5SpC3/GAGCGCTACTCAGATGACTTCGAGTGATTGTAATACTCACTCGAIGTCATCTGA GTAGCGCTCAAAAAAAAAAAAAAAAAAAAAAAAAA/3SpC3/	IDT
T:I	/5SpC3/AGTGCACGCTCTATGTATCGAAGAGTTGTGTAACAACCTTTTCIATACATAGAG CGTGCACTAAAAAAAAAAAAAAAAAAAAAAAAA/3SpC3/	IDT
Normal	/5SpC3/CGTAGCCTTCAGTATCTTCTACCCATCGTAAGATGGGTAGAAGATAGCTGA AGGCTAGCGAAAAAAAAAAAAAAAAAAAAAAAAA/3SpC3/	IDT
A:U - 2	/5SpC3/GCTTGCCTTGTGATCACAAGTATGTCAGGTAACGACATACTUGTATCGAC AAGGCAAGCAAAAAAAAAAAAAAAAAAAAAAAAAA/3SpC3/	IDT
A:U - 3	/5SpC3/GCTGGCCTTGGCACTAGGAACCTACCGCGGTAACGCGGTAAGTUCCTAGTGCA AAGGCCAGCAAAAAAAAAAAAAAAAAAAAAAAAAA/3SpC3/	IDT
A:U - 4	/5SpC3/TGCCAACGGTGGAGTACGAGGTAAGAAGCGTAAGCTTCTTACCUCGTACTCCA CCGTTGGCAAAAAAAAAAAAAAAAAAAAAAAAAA/3SpC3/	IDT
A:U - 5	/5SpC3/ATGGTTACGTTGGACATAGCGATCGTGGCGTAAGCAGCATCGCUATGTCCCAC GTGAACATAAAAAAAAAAAAAAAAAAAAAAAAAA/3SpC3/	IDT
C:riboG - 2	/5SpC3/TCCGACGGCAAGAGTCCTTCCAATTACCGTAAGGTAATTGGArGAGGACTCT TGCCGTGGAAAAAAAAAAAAAAAAAAAAAAAAA/3SpC3/	IDT
C:riboG - 3	/5SpC3/TCAATTGTTGGCAGAGGCCAATTAGTGTGTAAGACTAATTrGGCCCTGTGCC AACAAATTGAAAAAAAAAAAAAAAAAAAAAAAAA/3SpC3/	IDT
C:riboG - 4	/5SpC3/TCGGACCAAGTTATGGGCGCGAATTTCCGTAAGGAAATTCGGrGGCCATAA CTTGGTCCGAAAAAAAAAAAAAAAAAAAAAAAAA/3SpC3/	IDT
C:riboG - 5	/5SpC3/CTCAGACGAACGTTGCTACGGACCGTATGTAATAACGGGTCCrGTAGCAACG TTCGTCTGAGAAAAAAAAAAAAAAAAAAAAAAAAA/3SpC3/	IDT
T:ethenoA	/5SpC3/AAGGCCTGATGACGCATATCTGAGTGTGGTAACAGCACTCAG[Etheno- dA]TATGCGTCATCAGGCCTTAAAAAAAAAAAAAAAAAAAAAAAAA/3SpC3/	Phosphoroamidites from Glen Research
C:O6mG	/5SpC3/GTGTACAACATACGACTGCAGTAGAGTGCCTAAGCACTCTACT[O6mG]CAGTC GTATGTTGTACACAAAAAAAAAAAAAAAAAAAAAAAAA/3SpC3/	Phosphoroamidites from Glen Research
A:hmU	/5SpC3/AATCGATGAGCTAGAGACGTCGAATTGCAGTAATGCAATTCGA[hmU]GTCTCT AGTCCATCGATTAAAAAAAAAAAAAAAAAAAAAAAAAA/3SpC3/	Phosphoroamidites from Glen Research
Uracil-biotin	/5Biosg/GCTTGCCTTGTGATCACAAGTATGTCAGGTAACGACATACTUGTATCGAC AAGGCAAGCAAAAAAAAAAAAAAAAAAAAAAAAAA/3SpC3/	IDT
Y adapter 1	/5Phos/GATCGGAAGAGCACACGCTCT	IDT
Y adapter 2	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	IDT
PCR F (Indexed P7 primer)	CAAGCAGAAGACGGCATAACGAGATNNNNNNNGTACTGGAGTTCAGACGTGTGCTC TTCCGA*T*C*T	IDT
PCR R (Indexed P5 primer)	AATGATACGGCGACACCGAGATCTACACNNNNNNNTCTTCCCTACACGACGCTCT TCCGA*T*C*T	IDT

* represents a phosphonothioate linkage

Table 2.2: DNA repair substrate experimental conditions

Experiment	Substrate concentration (each in master mix) (nM)	Number of Substrates	Substrate concentration (Total in maseter mix) (nM)	Substrates	Figure reference
Barnyard Dilution	0.5 - 25	15	111	A:U (x5), C:riboG (X5), G:Abasic, G:U, C:I, T:I, Normal	Figure 2.1, 2.3, 2.4, 2.5
PBMC Dilution	0.5 - 25	15	111	A:U (x5), C:riboG (X5), G:Abasic, G:U, C:I, T:I, Normal	Figure 2.6, 2.7
PBMC 1	10	7	70	A:U, C:riboG, G:Abasic, G:U, C:I, T:I, Normal	Figure 2.8, 2.9, 2.10, 2.11
PBMC 2	0.9	11	10	A:U, C:riboG, G:Abasic, G:U, C:I, T:I, Normal, T:ethenoA, C:O6mG, A:hmU, A:U- 5'biotin	Figure 2.8, 2.11
PBMC 3	0.9	11	10	A:U, C:riboG, G:Abasic, G:U, C:I, T:I, Normal, T:ethenoA, C:O6mG, A:hmU, A:U- 5'biotin	Figure 2.8
Barnyard	10	2	20	A:U, C:riboG	Data not shown

Table 2.3: Differential gene expression for base excision repair genes

gene	cluster	p_val	avg_logFC	pct.1	pct.2	p_val_adj	sample
POLE4	T	2.60E-15	-0.359	0.052	0.135	8.71E-11	pbmc1
PARP1	B	5.84E-14	0.726	0.255	0.09	1.96E-09	pbmc1
POLE4	Mono	1.13E-40	0.662	0.247	0.052	3.78E-36	pbmc1
APEX1	DC	8.49E-14	0.716	0.333	0.059	2.85E-09	pbmc1
RNASEH2C	DC	2.46E-05	0.447	0.25	0.076	8.24E-01	pbmc1
HMGB1	NK	5.26E-21	0.388	0.712	0.744	1.76E-16	pbmc2
POLE4	Mono	1.27E-73	0.438	0.459	0.12	4.25E-69	pbmc2
APEX1	DC	9.00E-19	0.520	0.517	0.145	3.02E-14	pbmc2
RNASEH2B	DC	5.91E-09	0.267	0.371	0.133	1.98E-04	pbmc2
POLD4	B	3.74E-08	0.481	0.247	0.116	1.25E-03	pbmc2
MBD4	B	5.64E-08	0.434	0.229	0.103	1.89E-03	pbmc2
PARP1	B	1.48E-05	0.353	0.307	0.183	4.97E-01	pbmc2
POLE4	T	1.16E-20	-0.275	0.107	0.214	3.88E-16	pbmc3
HMGB1	NK	1.10E-23	0.383	0.711	0.733	3.68E-19	pbmc3
POLE4	Mono	2.23E-76	0.487	0.389	0.104	7.48E-72	pbmc3
APEX1	DC	2.27E-13	0.440	0.471	0.134	7.62E-09	pbmc3
POLE3	DC	1.96E-08	0.379	0.426	0.152	6.56E-04	pbmc3
MBD4	B	2.17E-13	0.524	0.226	0.083	7.27E-09	pbmc3
PARP1	B	1.88E-10	0.422	0.346	0.176	6.32E-06	pbmc3
POLD4	B	6.01E-08	0.436	0.26	0.136	2.02E-03	pbmc3
RNASEH2B	B	9.14E-04	0.310	0.192	0.122	1.00E+00	pbmc3
POLE4	T	1.10E-64	-0.270	0.091	0.211	3.70E-60	all_cells
PARP1	B	5.26E-24	0.528	0.305	0.152	1.76E-19	all_cells
MBD4	B	2.70E-20	0.425	0.176	0.072	9.05E-16	all_cells
POLD4	B	7.56E-13	0.385	0.197	0.105	2.53E-08	all_cells
RNASEH2B	B	5.52E-05	0.261	0.156	0.106	1.00E+00	all_cells
HMGB1	NK	1.13E-46	0.406	0.656	0.611	3.78E-42	all_cells
POLE3	NK	1.35E-03	0.268	0.142	0.125	1.00E+00	all_cells
POLE4	Mono	5.51E-210	0.509	0.384	0.092	1.85E-205	all_cells
APEX1	DC	1.16E-45	0.540	0.459	0.114	3.89E-41	all_cells

Significant differences and fold changes in gene expression between cell types and all other cells were calculated using Wilcoxon Rank Sum test (using the FindAllMarkers function from Seurat v3.0.0 (Stuart et al. 2019b)) for all cell types. The result was filtered for genes within the KEGG base excision repair pathway (hsa03410). Statistics were calculated for each replicate individually (sample: pbmc1, pbmc2, pbmc3) and for all samples integrated (sample: all cells) (**Figure 2.8** and **Figure 2.9**).

Table 2.4. Differential repair calculated for biological repair positions

gene	p_val	avg_logFC	pct.1	pct.2	p_val_adj	celltype1	celltype2
Abasic-45	1.46E-165	9.94E-01	1	0.997	6.20E-163	T	Mono
Abasic-45	1.06E-73	5.41E-01	1	1	4.53E-71	T	NK
Abasic-45	2.89E-71	9.26E-01	1	0.997	1.23E-68	B	Mono
Abasic-45	3.92E-38	4.73E-01	1	1	1.67E-35	B	NK
Abasic-45	6.55E-36	4.53E-01	1	0.997	2.79E-33	NK	Mono
Abasic-45	2.51E-23	8.85E-01	1	1	1.07E-20	T	DC
Abasic-45	3.19E-22	8.16E-01	1	1	1.36E-19	B	DC
Abasic-45	3.02E-07	3.44E-01	1	1	1.29E-04	NK	DC
Abasic-46	7.65E-108	-6.04E-01	0.998	1	3.26E-105	T	Mono
Abasic-46	1.17E-41	-4.84E-01	1	1	4.97E-39	NK	Mono
Abasic-46	2.34E-25	-9.40E-01	0.998	0.958	9.96E-23	T	DC
Abasic-46	7.44E-23	-3.41E-01	0.998	1	3.17E-20	T	B
Abasic-46	3.41E-21	-8.19E-01	1	0.958	1.45E-18	NK	DC
Abasic-46	1.97E-16	-5.99E-01	1	0.958	8.41E-14	B	DC
Abasic-46	1.87E-13	-2.63E-01	1	1	7.96E-11	B	Mono
Abasic-46	1.55E-10	-3.35E-01	1	0.958	6.60E-08	Mono	DC
GU-45	3.17E-34	4.31E-01	0.984	0.959	1.35E-31	T	Mono
GU-45	1.53E-28	5.89E-01	0.995	0.959	6.51E-26	B	Mono
GU-45	5.45E-26	5.76E-01	0.995	0.969	2.32E-23	B	NK
GU-45	6.41E-24	4.18E-01	0.984	0.969	2.73E-21	T	NK
riboG-44	8.06E-52	1.03E+00	1	0.897	3.43E-49	B	NK
riboG-44	2.99E-37	-5.64E-01	0.957	1	1.27E-34	T	B
riboG-44	2.19E-27	6.32E-01	1	0.827	9.35E-25	B	Mono
riboG-44	2.11E-24	4.64E-01	0.957	0.897	8.98E-22	T	NK
riboG-44	5.83E-22	-1.28E+00	0.897	0.958	2.48E-19	NK	DC
riboG-44	6.68E-19	-8.14E-01	0.957	0.958	2.84E-16	T	DC
riboG-44	2.58E-16	-8.82E-01	0.827	0.958	1.10E-13	Mono	DC
riboG-44	2.06E-05	-3.96E-01	0.897	0.827	8.77E-03	NK	Mono
riboG-44	6.90E-05	-2.50E-01	1	0.958	2.94E-02	B	DC
Uracil-45	6.40E-144	1.51E+00	0.979	0.704	2.72E-141	T	Mono
Uracil-45	8.78E-68	1.59E+00	0.995	0.704	3.74E-65	B	Mono
Uracil-45	2.19E-40	6.68E-01	0.979	0.945	9.32E-38	T	NK
Uracil-45	3.23E-36	8.39E-01	0.945	0.704	1.38E-33	NK	Mono
Uracil-45	1.17E-33	7.54E-01	0.995	0.945	4.99E-31	B	NK
Uracil-45	6.94E-16	9.53E-01	0.995	0.917	2.96E-13	B	DC
Uracil-45	3.97E-12	8.67E-01	0.979	0.917	1.69E-09	T	DC

Uracil-45	2.97E-08	-6.40E-01	0.704	0.917	1.26E-05	Mono	DC
Abasic-45	6.27E-183	5.91E-01	1	0.999	2.67E-180	T	all_other_cells
Abasic-45	2.83E-155	-9.37E-01	0.997	1	1.21E-152	Mono	all_other_cells
Abasic-45	8.37E-42	-4.42E-01	1	1	3.57E-39	NK	all_other_cells
Abasic-45	3.30E-16	-7.61E-01	1	1	1.41E-13	DC	all_other_cells
Abasic-46	6.91E-101	-4.43E-01	0.998	0.998	2.94E-98	T	all_other_cells
Abasic-46	1.55E-95	5.42E-01	1	0.998	6.59E-93	Mono	all_other_cells
Abasic-46	4.35E-23	8.14E-01	0.958	0.999	1.85E-20	DC	all_other_cells
GU-45	1.63E-31	-4.06E-01	0.959	0.983	6.94E-29	Mono	all_other_cells
GU-45	1.98E-26	2.58E-01	0.984	0.97	8.43E-24	T	all_other_cells
GU-45	2.36E-19	-3.82E-01	0.969	0.981	1.01E-16	NK	all_other_cells
riboG-44	4.18E-39	5.88E-01	1	0.935	1.78E-36	B	all_other_cells
riboG-44	1.08E-26	-5.21E-01	0.897	0.943	4.58E-24	NK	all_other_cells
riboG-44	2.22E-18	8.12E-01	0.958	0.939	9.46E-16	DC	all_other_cells
Uracil-45	6.71E-139	-1.45E+00	0.704	0.976	2.86E-136	Mono	all_other_cells
Uracil-45	4.41E-108	6.92E-01	0.979	0.851	1.88E-105	T	all_other_cells
Uracil-45	2.56E-22	-5.58E-01	0.945	0.943	1.09E-19	NK	all_other_cells
Uracil-45	7.38E-12	2.50E-01	0.995	0.941	3.14E-09	B	all_other_cells
Uracil-45	3.94E-07	-7.26E-01	0.917	0.944	1.68E-04	DC	all_other_cells

Significant differences and fold changes in repair activities between cell types and all other cells and all pairwise comparisons were calculated using Wilcoxon Rank Sum test (FindAllMarkers, FindMarkers, Seurat v3.0.0 (Stuart et al. 2019b)) for all cell types. The result was filtered for adjusted $P < 0.05$ and repair positions (**Figure 2.8**). Column information is as follows (from Seurat documentation for FindMarkers):

gene: concatenation of repair substrate and position.

p_val: p value of Wilcoxon Rank Sum test

avg_logFC: The log fold-change of average repair activity between group and all other cells. Positive values indicate that the repair activity is greater in the first group.

pct.1: Percentage of cells where the repair activity was detected in the group.

pct.2: Percentage of all other cells where the repair activity was detected.

p_val_adj: Adjusted p-value, based on Bonferroni correction using all hairpin positions in the dataset

Figure 2.1: Development and validation of a single-cell assay for measuring DNA repair capacity

a. Schematic of DNA repair substrates used in single-cell Haircut. Strand incision generates a new 5'-end whose position is captured by cDNA synthesis with barcoded oligonucleotides.

b. Overview of single-cell Haircut. After droplet formation, cell lysis creates a ~50 pL reaction wherein enzymes contributed by the cell catalyze substrate turnover (i.e., strand incision for specific DNA repair substrates). Repair products and mRNAs are converted to cDNA with barcoded oligo-dT primers and separated by size to enable separate library preparations. The cDNA for each fraction is analyzed to identify the cell barcode and either mRNA abundance or the amount of enzymatic activity (i.e., number of strand incisions).

c. Polyadenylated hairpins containing a single uracil or ribonucleotide (25 nM each) were added to a single-cell suspension of a mixture of Hap1 cells containing null alleles of either UNG or RNASEH2C prior to capture in a 10x Genomics 3' Gene Expression experiment. Sequences from the DNA repair fraction in **(b)** were grouped based on their cell barcodes, and the level of strand incision for uracil and ribonucleotide substrates was used to classify cells as either UNG^{ko} (green) or RNASEH2C^{ko} (blue) based on strand incision activity (UMI counts at position 44 for ribonucleotide, position 45 for uracil) greater than 5% of the maximum for all cells. Cells that fall on the x- or y-axis are single RNASEH2C^{ko} and UNG^{ko} cells, respectively; doublets are in yellow; and cells with low signal (<5% of the max for both activities) are in grey.

d. Aggregate counts of strand incision events are plotted against hairpin position for cells classified in **(c)** on the U:A substrate (top) and rG:C substrate (bottom). The vertical dashed line notes the position of the uracil and ribonucleotide (position 44 in both cases). UNG^{ko} cells fail to incise uracil-containing hairpins (green line in top panel) and RNASEH2C^{ko} cells fail to incise ribonucleotide-containing hairpins (blue line in bottom panel). The predominant signal at position 45 (top) reflects UNG conversion of the uracil (position 44) to an abasic site, followed by removal of the abasic site by Ape-1 or Pol β. The predominant signal at position 44 (bottom) reflects 5'-incision of the ribonucleotide at position 44 followed by copying of the rG template by reverse transcriptase in the droplet. Additional processing of incised repair intermediates yields signals at positions 3' of the lesion (signals at positions 46-50 in the uracil substrate (top), and positions 45-50 in the ribonucleotide substrate (bottom)).

e. Variable mRNA expression from cells classified in **(c)** was used to calculate a UMAP projection (identical coordinates in all 4 panels). Uracil repair activity (natural logarithm of strand incision counts (position 45 for uracil substrate and position 44 for ribonucleotide substrate; **d**) divided by total counts for that cell multiplied by a scaling factor of 10⁴; top left) and ribonucleotide repair activity (top right) were superimposed in a grey-to-red scale and delineate two major cell types in the experiment; 90% of cells in each class have a scored activity. Levels of UNG and RNASEH2C mRNA (natural logarithm of mRNA counts divided by total counts for that cell multiplied by a scaling factor of 10⁴) are plotted on the bottom panels are not sufficient to classify cell types. Stabilization of the null-mutation-containing RNASEH2C mRNA yields uniform RNASEH2C mRNA detection across both cell types.

Results and discussion

We included synthetic DNA hairpins with polyadenylate tails and DNA lesions at defined positions (**Figure 2.1a**) in a single-cell mRNA sequencing experiment and developed a protocol to capture incised DNA repair intermediates and products from single-cell extracts by library preparation and sequencing (**Figure 2.1b** and **Figure 2.2**). Because our method employs massively parallel measurement of strand incision on DNA hairpins, we named it “Haircut”.

Cell mixing experiments confirmed we could measure DNA repair activities at cellular resolution with single-cell Haircut. We added DNA repair substrates with a uracil (U:A base-pair) or ribonucleotide (rG:C) to a single-cell suspension of haploid human UNG^{KO} and $RNASEH2C^{KO}$ knockout cells immediately prior to emulsion formation in a droplet-based single-cell mRNA sequencing experiment (10x Genomics 3' Gene Expression; **Figure 2.1b**). After sequential incubations at 37 °C and 53 °C to first promote endogenous enzymatic activity and then reverse transcription, the emulsion was broken and cDNA molecules synthesized from mRNA and repair substrate templates (the “repair fraction”) were separated by size. The mRNA fraction was subjected to the standard protocol to measure gene expression for single cells, whereas the repair fraction (containing DNA substrates, intermediates, and products) was captured in a modified protocol (**Figure 2.2**). Analysis of the mRNA fraction by DNA sequencing yielded the mRNA identity, a cell-specific barcode, and a unique molecular identifier (UMI (Islam et al. 2014)). Similarly, DNA sequencing of the repair fraction yielded the cell barcode, UMI, and a 5' position derived from cDNA synthesis on either full-length hairpins or incised repair intermediates and products.

We captured thousands of single cells with expected DNA repair defects: $RNASEH2C^{KO}$ cells could incise uracil but not a ribonucleotide repair substrate, and vice versa for UNG^{KO} cells,

with a few droplets containing more than one cell from each genotype and therefore both uracil and ribonucleotide repair activities (**Figure 2.1c** and **d**). DNA repair associated incision and processing activities were measured at expected positions based on known repair pathways (**Figure 2.1d**, right) and were only present in cell-associated droplets (as determined from mRNA signals; **Figure 2.3**). Some signals, discussed more below and in **Chapter III**, were robustly captured in empty drops and were excluded from analysis. We calculated repair activity for each cell as the natural logarithm of strand incision counts at the primary repair site (position 45 for the uracil substrate and position 44 for the ribonucleotide substrate) divided by total counts for that cell multiplied by a scaling factor of 10^4 . We also calculated two-dimensional UMAP projections based on variable mRNA expression and colored individual cells by enzymatic activity (UNG or RNASEH2) and mRNA abundance (**Figure 2.1e**). DNA repair activity was robustly detected for each cluster (**Figure 2.1e**, top row) and was sufficient to assign 75% of cells to the correct cell type using 1,500 reads-per-cell (**Figure 2.4**), similar to read depths required for cell type classification using gene expression (Heimberg et al. 2016). While UNG (229 cells) and RNASEH2C (1,075 cells) mRNAs were identified in these cells, they were not variably expressed across cell clusters (**Figure 2.4**). Moreover, our analysis of RNASEH2C mRNA levels in RNASEH2C^{KO} cells found that whereas the mutation yields cells that lack RNASEH2 activity (**Figure 2.1e**, top right), it does not cause mRNA decay, with similar mRNA isoform abundance detected in both cell types (**Figure 2.5**). Altogether, this experiment illustrates the unique and orthogonal information provided by single-cell biochemical assays, which is especially useful in situations where mRNA expression is not predictive or may even be misleading of functional status.

Oligo dT primer in single-cell experiment captures DNA repair substrates

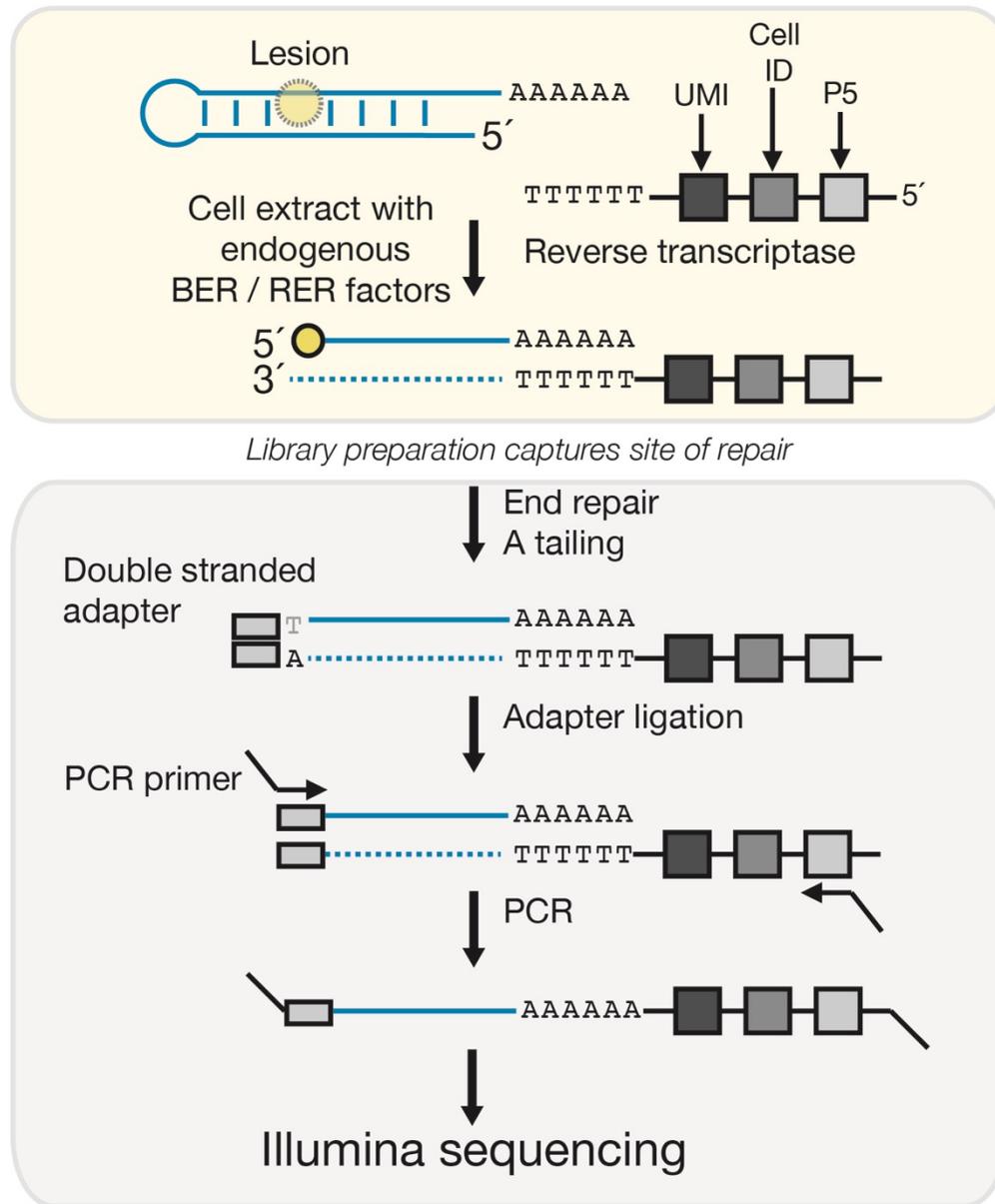


Figure 2.2: Schematic of single-cell Haircut library preparation

DNA repair substrates are added to a 10x Genomics Chromium Single Cell 3' v2 kit. Within each droplet, DNA repair substrates are exposed to cell extract containing endogenous active DNA repair enzymes. Additionally, each drop contains a reverse transcriptase and an oligo-dT reverse transcription primer. Cell-derived DNA repair enzymes initiate DNA repair on the substrates creating a strand incision, oligo-dT primer and reverse transcriptase present in the drop capture the DNA repair intermediates along with cellular mRNAs. After the emulsion is broken, DNA repair substrates are separated from mRNAs by size. The 5' site of strand incision is captured through end repair, A-tailing, ligation of a TruSeq adapter, followed by PCR. This library is compatible with next generation sequencing.

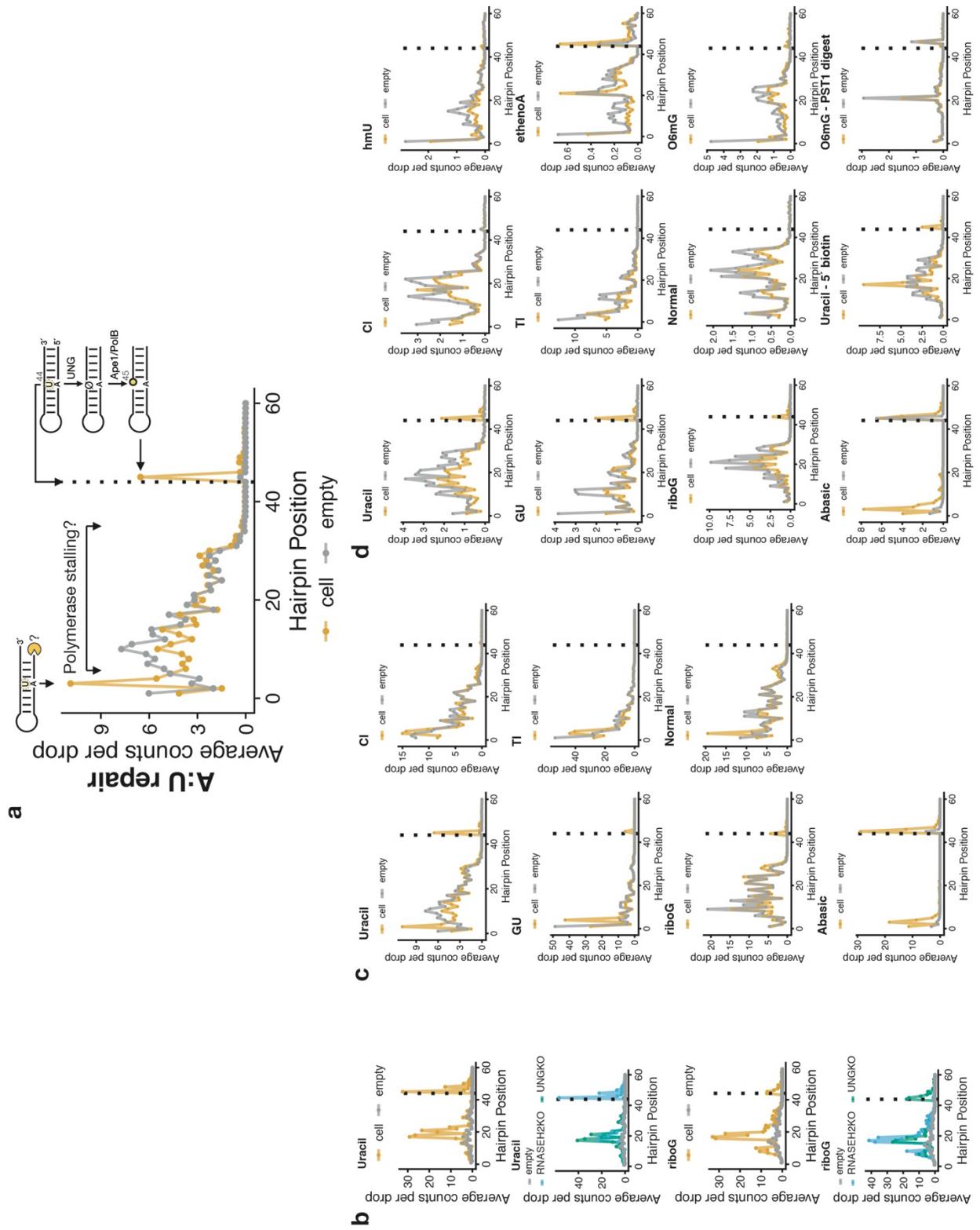


Figure 2.3: Determining biological DNA repair activity from single-cell Haircut signals

a. Counts per cell (mean) (orange) by position were compared to counts per empty drop (mean) to determine which positions contained biological activity. Empty drops were determined by filtering out cell-associated barcodes from the unfiltered repair matrix. The resulting repair matrix contains many barcodes that are associated with only a single UMI, so the matrix was filtered by descending UMI counts to the same number of cell-associated barcodes. This repair matrix was used as the input to calculate empty drop signal across the hairpin by calculating the sum across all drops at each hairpin position. Positions exhibiting signal above empty drop background and associated with a known DNA repair position (e.g., position 45 for U:A) were considered biologically relevant. Some signals above background at the 5' end of the substrate could be due to cellular exonuclease activity and were not considered in the analysis.

b. Coverage across all cells (orange) or across cell types (blue, green) and empty drops (grey) for barnyard experiment (**Figure 2.1**). While signals at the 5' end of the substrate in cells are higher than signals in empty drops, only signals at or adjacent to the site of the lesion (dotted line) are consistent with known DNA repair pathways (**Figure 2.1**) and are dependent on the presence of DNA repair factors.

c. Coverage across all cells (orange) and empty drops (grey) for PBMC experiment (**Figure 2.8**). C:I and T:I substrates did not show biological activity above background.

d. Coverage across all cells (orange) and empty drops (grey) for PBMC experiments (**Figure 2.9**). The Uracil-5' biotin substrate contained a 5'-biotin. During the library preparation, uncleaved substrate was removed using streptavidin beads. We only saw a modest reduction in signal on the 5' end of the substrate in cell-containing and empty drops. The hmU substrate did not have signal above empty drop background. The ethenoA substrate did not have signal above background due to a high level of background signal in empty drops, which could be due to the bulky ethenoA lesion causing the RT to stop. To measure direct reversal substrates, we included an O6mG containing hairpin. Following substrate isolation from the mRNA fraction, substrates were digested with PstI to measure the removal of O6mG (Encell and Loeb 1999a); however, we measured digestion in empty drops as well as drops with cells indicating the method was not specific for droplets containing cells.

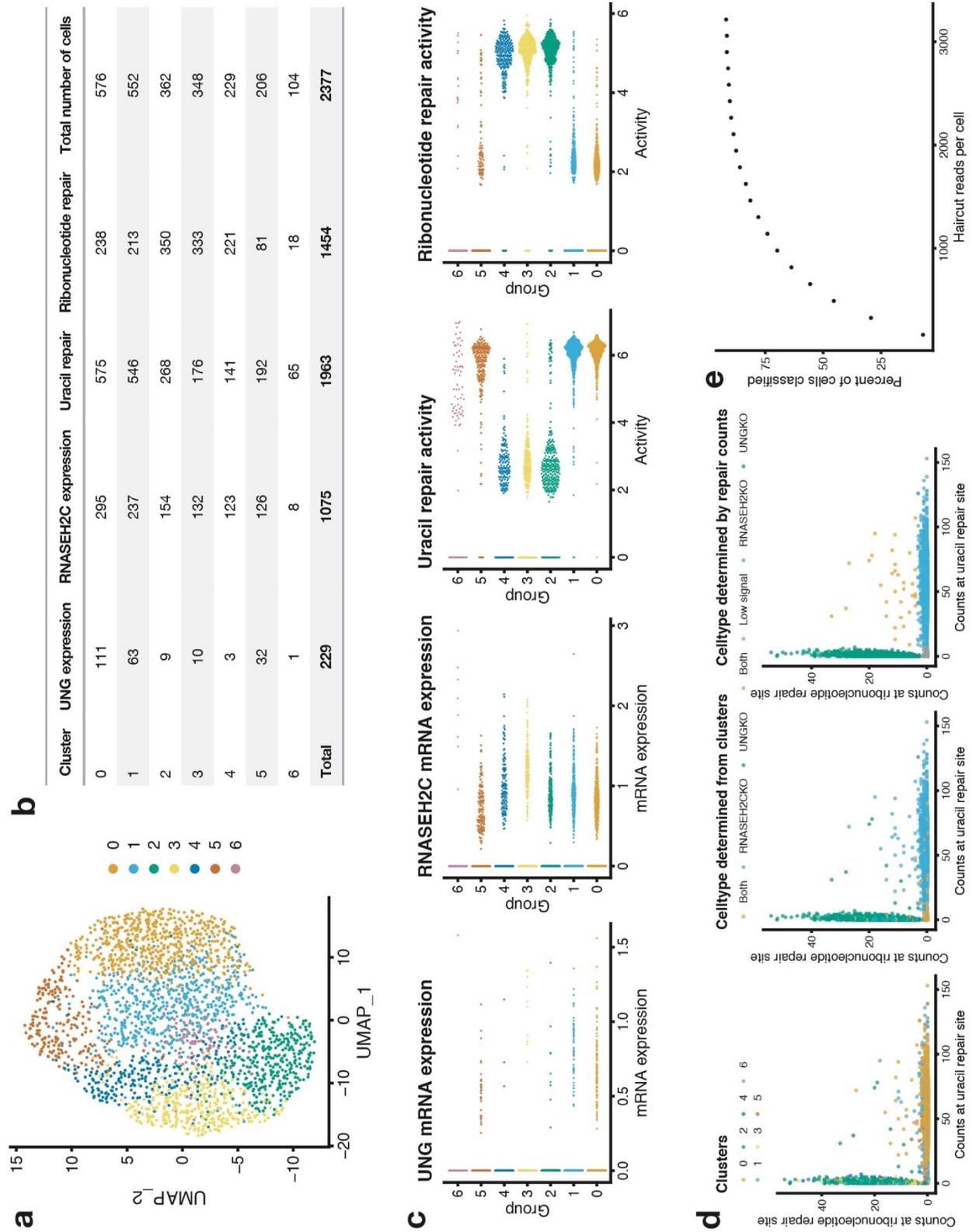


Figure 2.4: DNA repair measurements determine cell types in a cell mixing experiment

a. UMAP plot of cell mixing experiment (**Figure 2.1**). mRNA expression was used to calculate UMAP projections and cluster cells. Cells were clustered using an unsupervised shared nearest neighbors method in Seurat (FindNeighbors, FindClusters). Cells were colored by the resulting cluster numbers.

b. A table of cells expressing UNG mRNA, RNASEH2C mRNA, uracil repair activity, or ribonucleotide repair activity in each cluster. UNG mRNA was detected in < 10% of cells while RNASEH2C mRNA was detected in ~45% of cells. One or both repair activities were measured in >90% of cells.

c. Beeswarm plots of UNG mRNA expression, RNASEH2C mRNA expression, uracil repair activity, or ribonucleotide repair activity by cluster. To determine if UNG and RNASEH2C mRNA expression could be used to assign cell types, we made beeswarm plots of UNG expression and RNASEH2C expression and attempted to assign cell types based on expression levels by cluster (left 2 plots). Due to few cells with UNG or RNASEH2C mRNA measurements we could not assign cell types independent of repair measurements. To assign cell types by repair activity, we made beeswarm plots of repair activities by clusters (right 2 plots). Clusters with high uracil repair activity and low ribonucleotide repair activity were assigned as RNASEH2C cells (clusters 0, 1, and 5). Clusters with high ribonucleotide repair activity and low uracil repair activity were assigned as UNG (clusters 2, 3, and 4). Cluster 6 was assigned as having both repair activities.

d. Counts at ribonucleotide repair site and uracil repair site were plotted and colored by cluster (left), cell type as determined above (**c**) (middle), and cell types as determined in **Figure 2.1** (right). Cell types as determined in **Figure 2.1** show the greatest segregation in these plots.

e. BAM files from the cell mixing experiment in **Figure 2.1** were down sampled to different sequencing depths to determine how many reads per cell are required for cell type classification. At each threshold, cell types were classified as in **Figure 2.1** and the percent of cells that can be classified as either UNG^{KO} or RNASEH2C^{KO} is plotted as a function of sequencing depth. Approximately 75% of the 2,377 cells are correctly classified at 1,500 reads per cell.

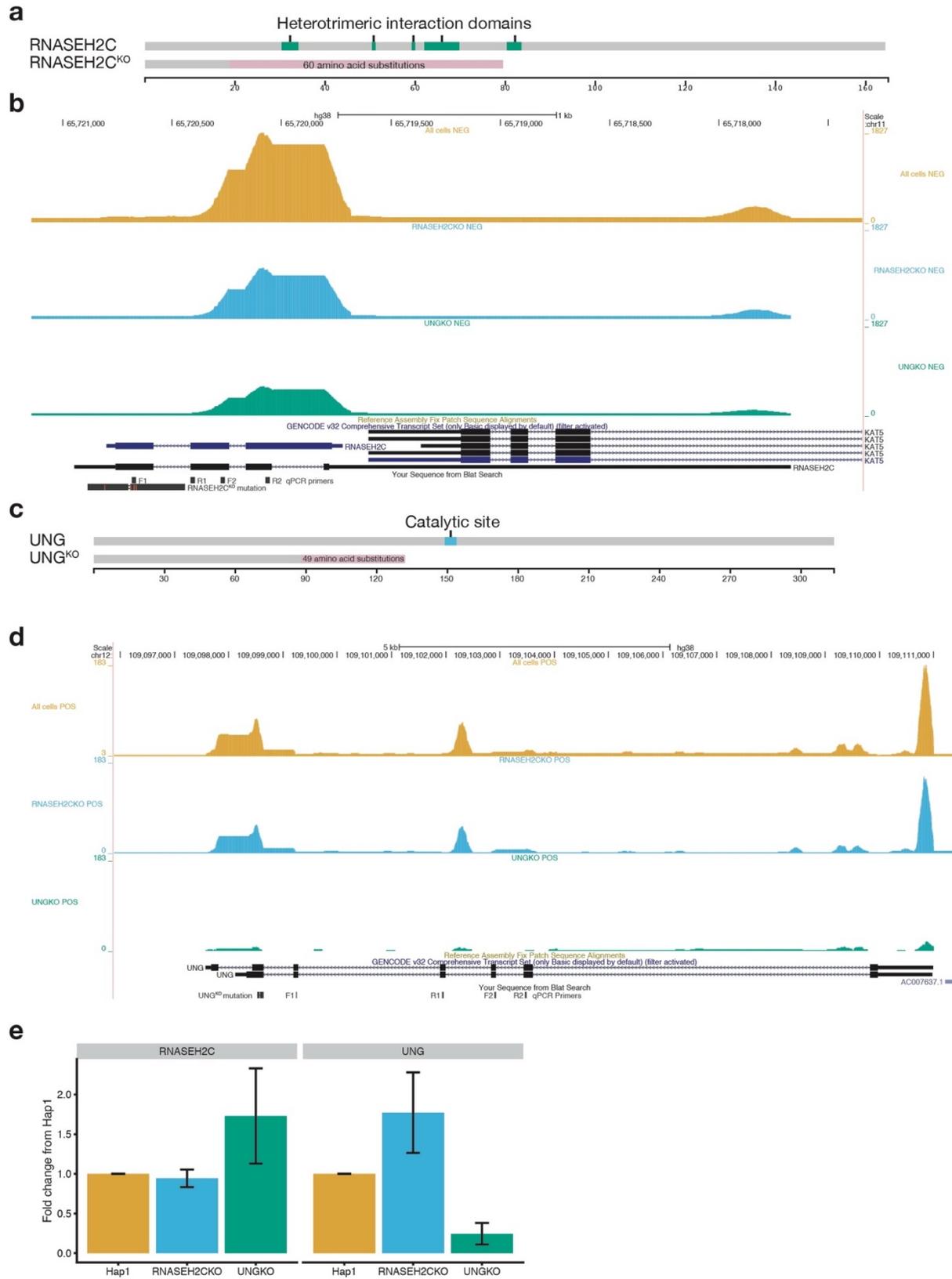


Figure 2.5: Single-cell mRNA expression of RNASEH2C and UNG in Hap1 knockout cells

a. Protein domains for RNASEH2C (top) and mutant RNASEH2C (bottom) from Horizon knockout line. A 10 base pair deletion in the RNASEH2C^{KO} Hap1 line causes a frameshift at amino acid 19 that leads to 60 amino acid substitutions and a stop codon at position 79, disrupting the heterotrimeric interaction regions of RNASEH2C (Reijns et al. 2011).

b. mRNA coverage of RNASEH2C from single-cell RNA sequencing data (**Figure 2.1**). Coverage from all cells from negative strand is in orange. Cell types were identified by repair of uracil-containing or ribonucleotide-containing substrates (**Figure 2.1**). Alignment files were then separated by cell barcodes associated with cell types. Bulk mRNA coverage for RNASEH2C for RNASEH2C^{KO} cells (blue) and UNG^{KO} cells (green) show similar coverage for the RNASEH2C gene independent of cell type. Coverage for the RNASEH2C gene is located near the 3' end of the 2 mRNA isoforms (gene diagrams on bottom). Horizon RNASEH2C^{KO} sequencing results provided from Horizon Discovery displayed on bottom (RNASEH2C^{KO} mutation). qPCR primers are also displayed on bottom (F1/2, R1/2).

c. Protein domains for UNG (top) and mutant UNG (bottom) from Horizon knockout line. An 11 base pair deletion leads to a frameshift at amino acid 88, 49 amino acid substitutions, and a stop codon at amino acid 138 - likely leading to a truncated protein lacking the catalytic domain.

d. mRNA coverage of UNG from single-cell RNA sequencing data. Coverage from all cells from positive strand is in orange. Cells were identified by repair of uracil-containing or ribonucleotide-containing substrates. Alignment files were then separated by cell type. mRNA coverage for RNASEH2C^{KO} cells (blue) and UNG^{KO} cells (green) show reduced UNG mRNA coverage in UNG^{KO} cells. Gene diagram is on bottom. Horizon UNG^{KO} sequencing results provided from Horizon Discovery is displayed on bottom (UNG^{KO} mutation). qPCR primers are also displayed on bottom (F1/2, R1/2).

e. Quantitative PCR results for RNASEH2C and UNG mRNA confirm single-cell RNA sequencing results. RNASEH2C^{KO} cells express RNASEH2C mRNA. UNG^{KO} cells do not express UNG mRNA at high levels. Error bars represent errors across biological duplicates and 2 sets of primers indicated in **b** and **d**.

DNA repair activity measurements in single-cell extracts from human peripheral blood mononuclear cells (PBMCs) with separately barcoded uracil and ribonucleotide repair substrates spanning a 50-fold range in concentration showed that measured DNA repair signals change as a function of substrate concentration and time (**Figure 2.6**). Moreover, the proportion of DNA repair substrates recovered in the assay was independent of substrate, concentration, and incubation time (**Figure 2.7**). However, we did find that the proportion of substrates that were repaired increased as the concentration of substrates increased (**Figure 2.7**). Differences in DNA repair among cell types persisted independent of substrate concentration and time, enabling us to choose a single substrate concentration and time point (10 nM substrates at 60 min) for further experiments (**Figure 2.6**).

We next used single-cell Haircut to measure the biochemical heterogeneity of DNA repair in PBMCs using five DNA substrates (unmodified, and containing U:A, U:G, rG:C, and abasic:G lesions - added at 10 nM each; **Figure 2.8** and **Figure 2.9**). We were unable to measure DNA repair activity on several other base excision repair substrates (I:C, I:T, EthenoA:T, hydroxymethyl-U:A) and one direct reversal substrate (O⁶mG:C) (**Figure 2.3**), either due to the sensitivity of the assay (I:C, I:T, or hydroxymethylU:A) or the assay specificity (EthenoA:T and O⁶mG:C). We used single-cell mRNA expression to classify cells based on expression of common cell-type-specific markers (e.g., IL7R for CD4⁺ T cells, CD14 for monocytes; **Figure 2.8a** and **Figure 2.9**) and then used these classifications to assign cell-type-specific DNA repair activities (**Figure 2.8b-f** and **Figure 2.9**).

We found little signal on the unmodified DNA substrate, confirming it is not a repair substrate. In contrast, incision and processing activities measured on uracil (on U:A and U:G substrates), ribonucleotide, and abasic repair substrates matched expected positions based on

known repair pathways (left and middle panels in **Figure 2.8c-f** and **Figure 2.9**) and were only present in cell-associated droplets (as determined from mRNA signals; **Figure 2.3**). These data revealed unique signatures of DNA repair activities in specific cell types. Monocytes and dendritic cells had low incision activity on the U:A substrate (**Figure 2.8**) resonating with the low level of uracil base excision in monocytes (Briegert and Kaina 2007) and confirming that myeloid lineages have unique uracil repair phenotypes (Hansen et al. 2016) consistent with lower UNG mRNA expression in these cell populations (**Figure 2.10**). However, these differences in uracil repair were not apparent for the U:G substrate, presumably due to redundant activities of SMUG1 (Nilsen et al. 2000) and MBD4 (Hendrich et al. 1999) in incising U:G-containing substrates. Dendritic cells demonstrated a unique repair phenotype, with increased levels of DNA substrate processing, measured as increased signals downstream of the position of the synthetic lesion. To explain these differences in DNA repair phenotypes, we examined cell type-specific mRNA expression and found that expression of APEX1, encoding the abasic endonuclease Ape-1, was consistently and uniquely elevated in dendritic cells (**Table 2.3**), possibly explaining the increased processing of DNA repair intermediates on U:A, U:G, and abasic substrates in dendritic cells.

B cells and dendritic cells also had higher levels of ribonucleotide repair activity than other cell types (**Figure 2.8** and **Figure 2.9**), and the increase in ribonucleotide repair in B cells and dendritic cells was corroborated by elevated expression of some RNASEH2 subunits in our (**Table 2.3**) and previous single-cell mRNA sequencing data sets (**Figure 2.10**) (G. X. Y. Zheng et al. 2017). Increased RNASEH2 activity in B cells may aid processing of R-loops that form during class switching (Yu et al. 2003).

Finally, we focused on cell-type-specific differences in repair of a DNA hairpin containing a synthetic abasic site. These substrates undergo two unique events in droplets: on intact substrates,

reverse transcription halts at the abasic site, yielding extension products that map one base downstream of the abasic site (**Figure 2.8** and **Figure 2.9**). Alternatively, incision and removal of the abasic site by Pol β and Fen1 (A. Klungland and Lindahl 1997) yields repair intermediates with 5'-ends that map further downstream of the abasic site. Differences in signals from the abasic substrate specifically in monocytes and dendritic cells again indicate that they are more proficient at processing abasic lesions, evidenced by an increase in levels of intermediates 2 nt downstream of the abasic site (position 46; **Figure 2.8f**), likely due to elevated Ape-1 expression. The unique DNA repair phenotype provided some power for cell type classification (**Figure 2.11**). As additional activities are multiplexed with DNA repair, we expect the power of single-cell biochemical measurements for cell type classification will increase.

Our approach to measure heterogeneity of single-cell biochemical phenotypes can be expanded to measure other types of DNA repair activities (e.g., nucleotide excision repair and mismatch repair) and adapted to measure other enzyme classes using substrate-DNA conjugates (Jetson and Krusemark 2016), enabling simultaneous measurement of many biochemical activities (e.g., kinases, phosphatases, and proteases) with gene expression at single-cell resolution.

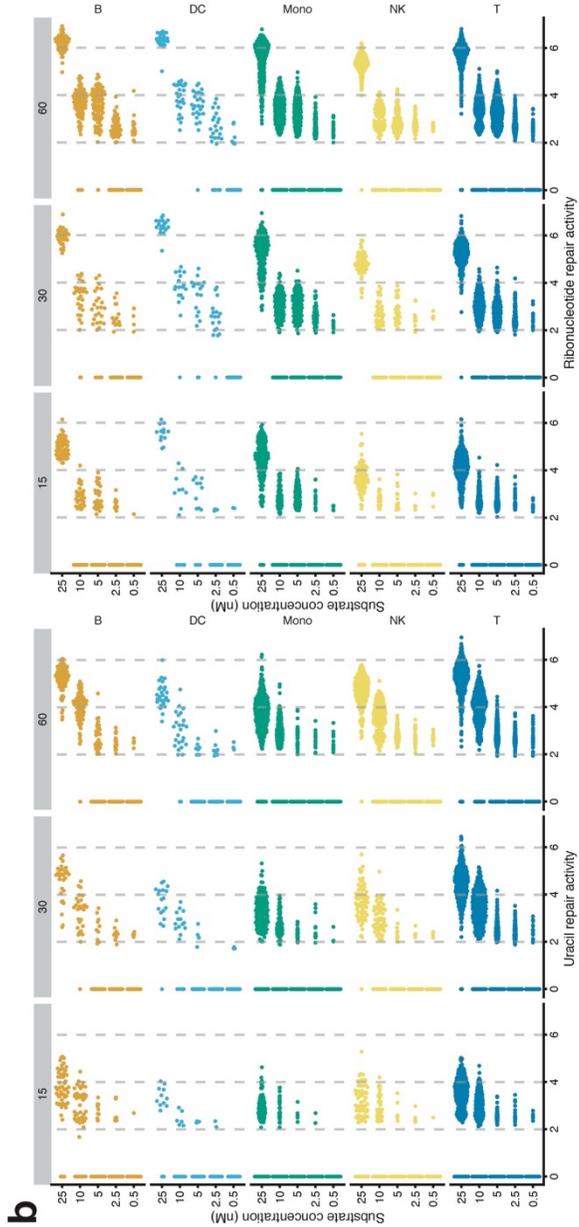
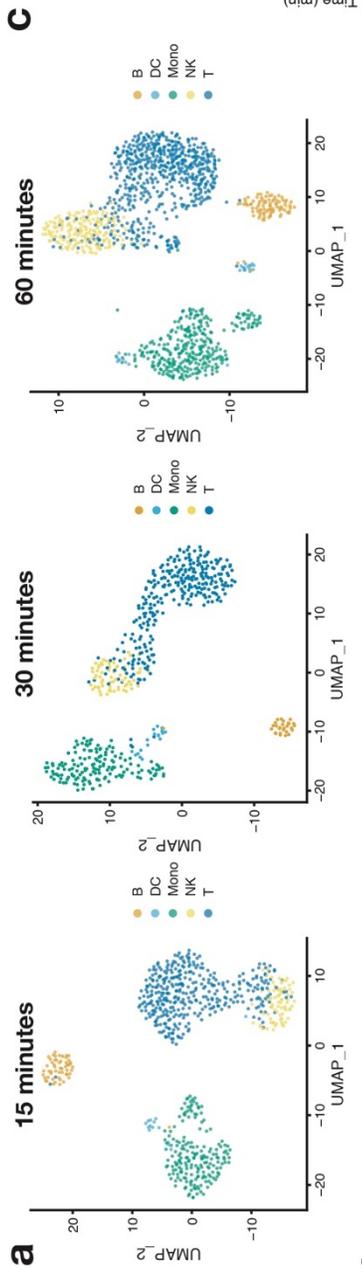
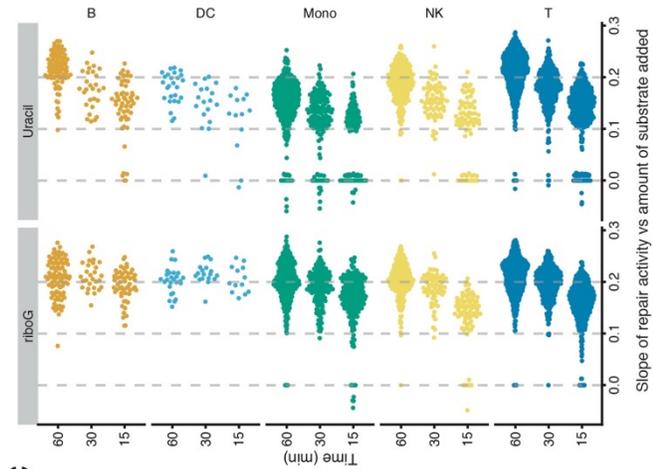


Figure 2.6: Measuring DNA repair in single cells across multiple concentrations and time points

a. DNA repair activities were measured in healthy human PBMCs. Uracil (U:A) and ribonucleotide (rG:C) substrates were mixed and added in a range of concentrations (0.5, 2.5, 5, 10, and 25 nM each). After the emulsion was created, the sample was separated into 3 tubes and incubated for 15, 30, or 60 min at 37 °C prior to reverse transcription at 53 °C. 800-1,500 cells were captured at each timepoint. Cell types were identified using gene expression markers using Seurat and visualized on UMAP projections.

b. Repair activities measured increased as substrate concentration increased. Repair activity increased over time from 15 min to 60 minutes. Differences in repair between cell types were consistent across substrate concentrations and most remained significant ($P < 0.05$).

c. A linear model (log normalized count at repair site (as defined in **Figure 2.1**) ~ substrate concentration) was fitted for each cell for each repair activity. The slope of each linear model is a measurement of repair activity. Repair activity increases as time increases. Additionally, repair activity trends measured using a single concentration and time point are consistent with activity measurements made across substrate concentration (e.g., monocytes have lower uracil repair compared to B cells and T cells, significant P values).

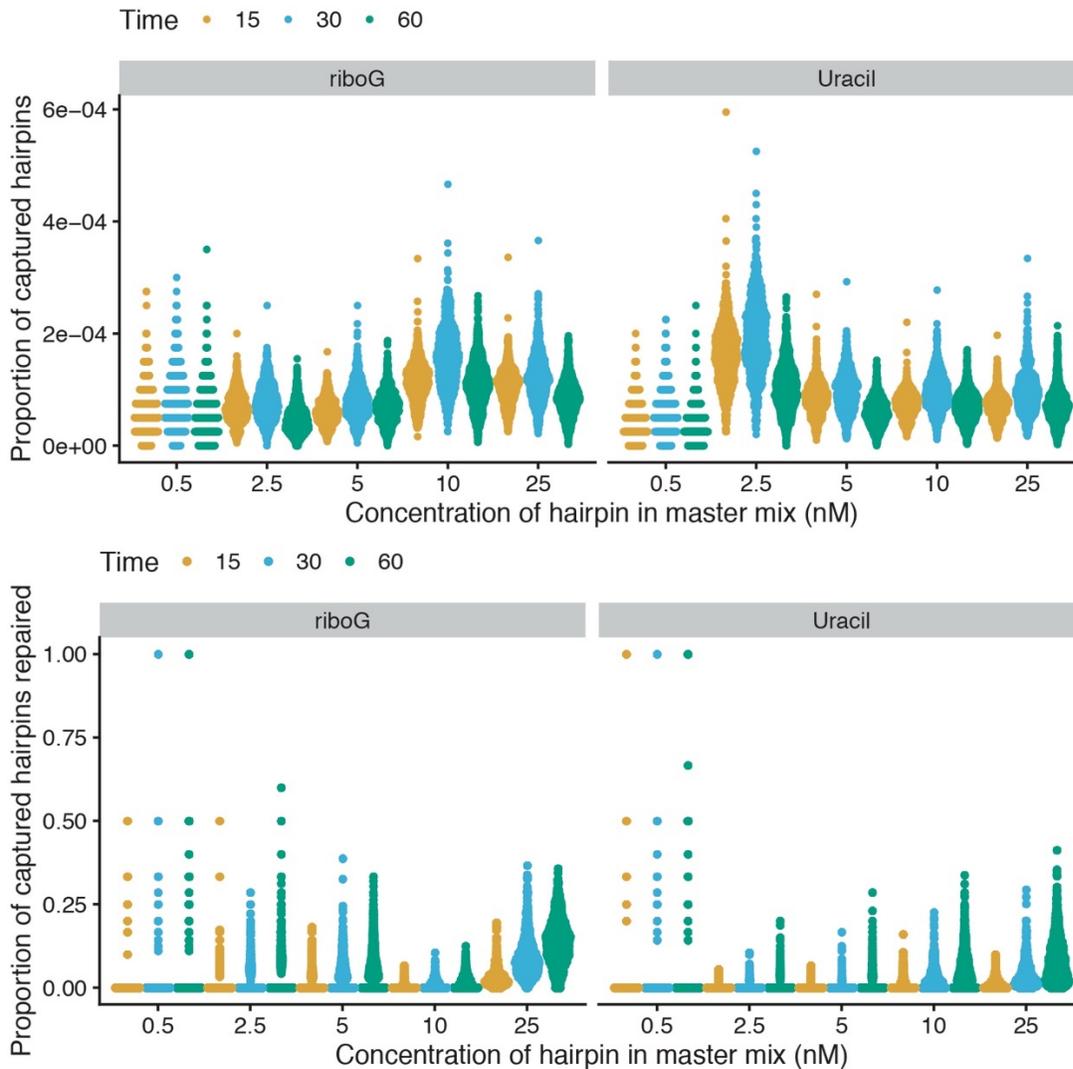


Figure 2.7: Proportion of DNA substrates captured per cell.

Uracil and ribonucleotide containing DNA repair substrates mixed at 5 concentrations (0.5, 2.5, 5, 10, and 25 nM) and were added to the cells/master mix in a PBMC experiment. After the emulsion was created, the sample was separated into 3 tubes and incubated for 15, 30, or 60 min at 37 °C prior to reverse transcription at 53 °C. 800-1,500 cells were captured at each timepoint (**Figure 2.6**). The number of hairpin molecules in each drop was estimated by assuming each drop was ~300 pL in volume and the hairpin concentration in the droplet is half that of the concentration in the master mix. The number of repaired hairpins was calculated by dividing the number of UMIs at repair positions (defined in **Figure 2.1**) by the number of UMIs for each hairpin. We captured approximately 0.01-0.025% of the hairpins that were added to the assay (with ~10 million total library reads) (top). The proportion of hairpins recovered was independent of the concentration with the exception of the lowest concentration of 0.5 nM and the uracil hairpin at 2.5 nM. The proportion of hairpins recovered was also independent of the 37 °C incubation time. The proportion of hairpins repaired was not independent of incubation time or concentration. With the exception of the lowest concentrations of 0.5 nM and 2.5 nM, the proportion of hairpins repaired was greater at longer time points and higher substrate concentrations (bottom).

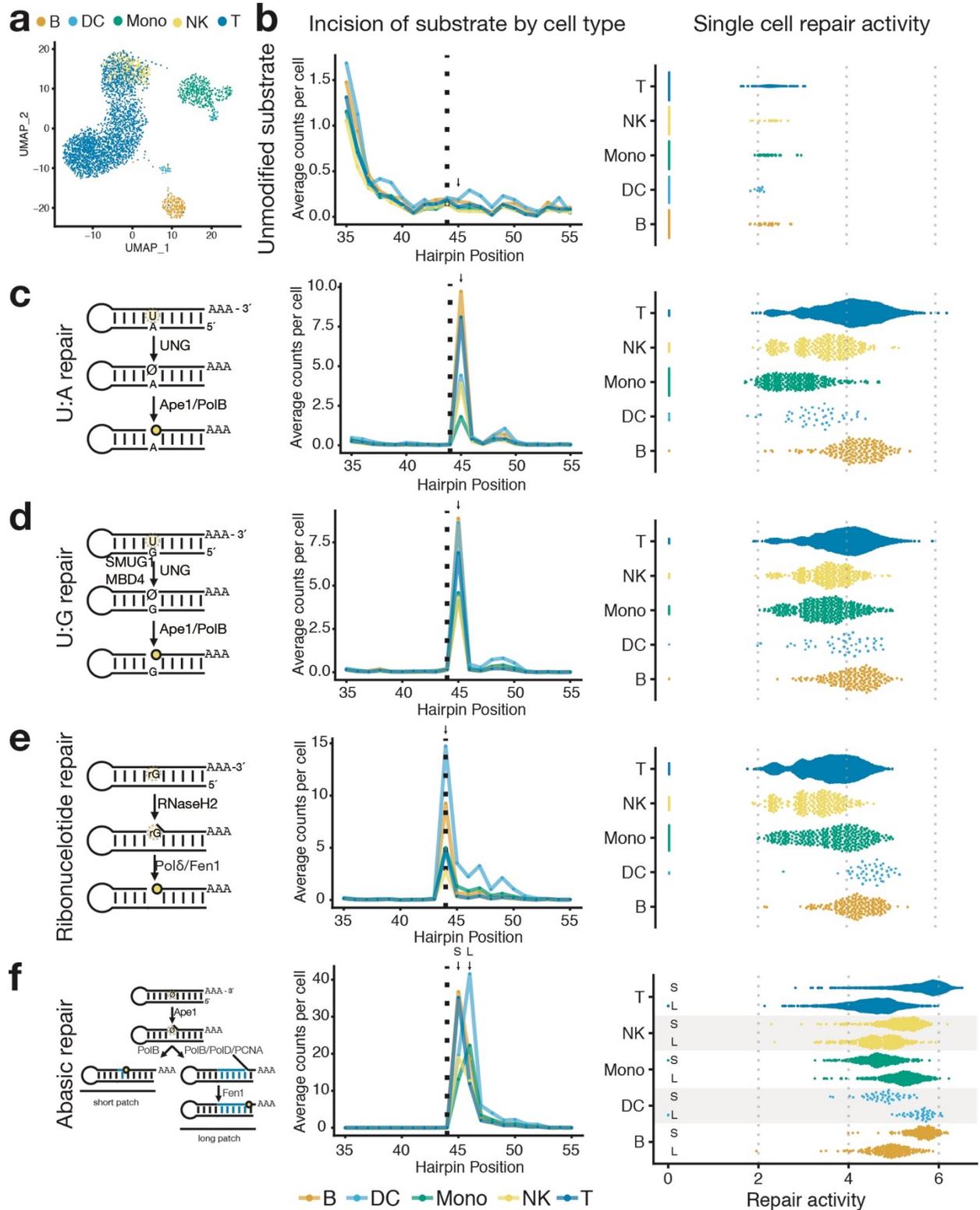


Figure 2.8: Analysis of DNA repair heterogeneity in human lymphocytes

a. Two-dimensional UMAP projection of variable gene expression across 3,298 human PBMCs captured in a 10x Genomics 3' Gene Expression experiment. Major cell types were determined by marker gene expression.

b. Single-cell DNA repair activity of an unmodified DNA substrate. The dashed vertical line indicates the position of lesions in other DNA repair substrates (left). Unmodified DNA substrates yield few measured incisions in single-cell Haircut with an average count per cell per position of < 1 . Very few cells have measured repair activity (natural logarithm of strand incision counts (at position 45; arrow, left) divided by total counts for that cell multiplied by a scaling factor of 10^4), indicating they are not substrates for cellular DNA repair activities.

c. Repair of a substrates containing a uracil:adenine (U:A) base-pair initiates with UNG-mediated removal of the uracil nucleobase followed by processing of the abasic site by Ape1 and Pol β (left). Cell-type-specific counts of incision and processing (mean) are plotted against the position of the hairpin and colored as in (a) (middle). Single-cell repair activities (natural logarithm of counts at the incision site (arrow) divided by total counts for that cell multiplied by a scaling factor of 10^4) are plotted for each cell from each cell type (right). Monocytes and dendritic cells have reduced uracil incision relative to other cell types ($P < 10^{-140}$ monocytes to T cells, and $P < 10^{-8}$ dendritic cells to T cells; Wilcoxon signed-rank test; **Table 2.4**).

d. Repair of a substrates containing a uracil:guanine base-pair initiates with UNG, SMUG, or MBD4-mediated removal of the uracil nucleobase followed by processing of the abasic site by Ape-1 and Pol β (left). Cell-type-specific incisions are plotted as in (c) with a predominant incision site one base downstream of the uracil (arrow, position 45), similar to the U:A substrate (a) (middle). The higher uracil repair activity (as defined in (c)) on the U:G relative to U:A substrates for monocytes and dendritic cells likely reflects recognition of the U:G substrate by SMUG and MBD4.

e. Repair of a substrates containing a riboG:C base-pair initiates with RNASEH2-mediated incision 5' of the ribonucleotide followed by processing by Pol δ and Fen1 (left). Cell-type-specific counts of incision and processing (mean) are plotted against the position of the hairpin and colored as in (a) (middle), with the predominant signal at the ribonucleotide, reflecting incision by RNASEH2 and cDNA synthesis using the ribonucleotide template by reverse transcriptase in the droplet (arrow, position 44). B cells and dendritic cells have higher levels of ribonucleotide repair activity (as defined in (c)) than other cell types ($P < 10^{-33}$ B cells to T cells; $P < 10^{-15}$ dendritic cells to T cells; Wilcoxon signed-rank test, **Table 2.4**).

f. Repair of substrates containing a abasic:guanine base-pair initiates with Ape-1-mediated incision followed by processing of the single-base gap by either Pol β (short-patch repair) or Pol δ/β and Fen1 (long-patch repair) (left). Cell-type-specific incisions are plotted as in (c) with a predominant incision sites one or more bases downstream of the abasic, depending on the cell type (middle). For each cell type, the levels of short-patch (top; signals 1 nt downstream of lesion, S arrow, position 45) and long-patch (bottom; signal 2 nt downstream lesion, L arrow, position 46) repair activities (as defined in (c)) are plotted (right). Monocytes and dendritic cells have lower levels of short-patch repair relative to other cell types ($P < 10^{-162}$ monocytes to T cells; $P < 10^{-19}$ dendritic cells to T cells; Wilcoxon signed-rank test, **Table 2.4**).

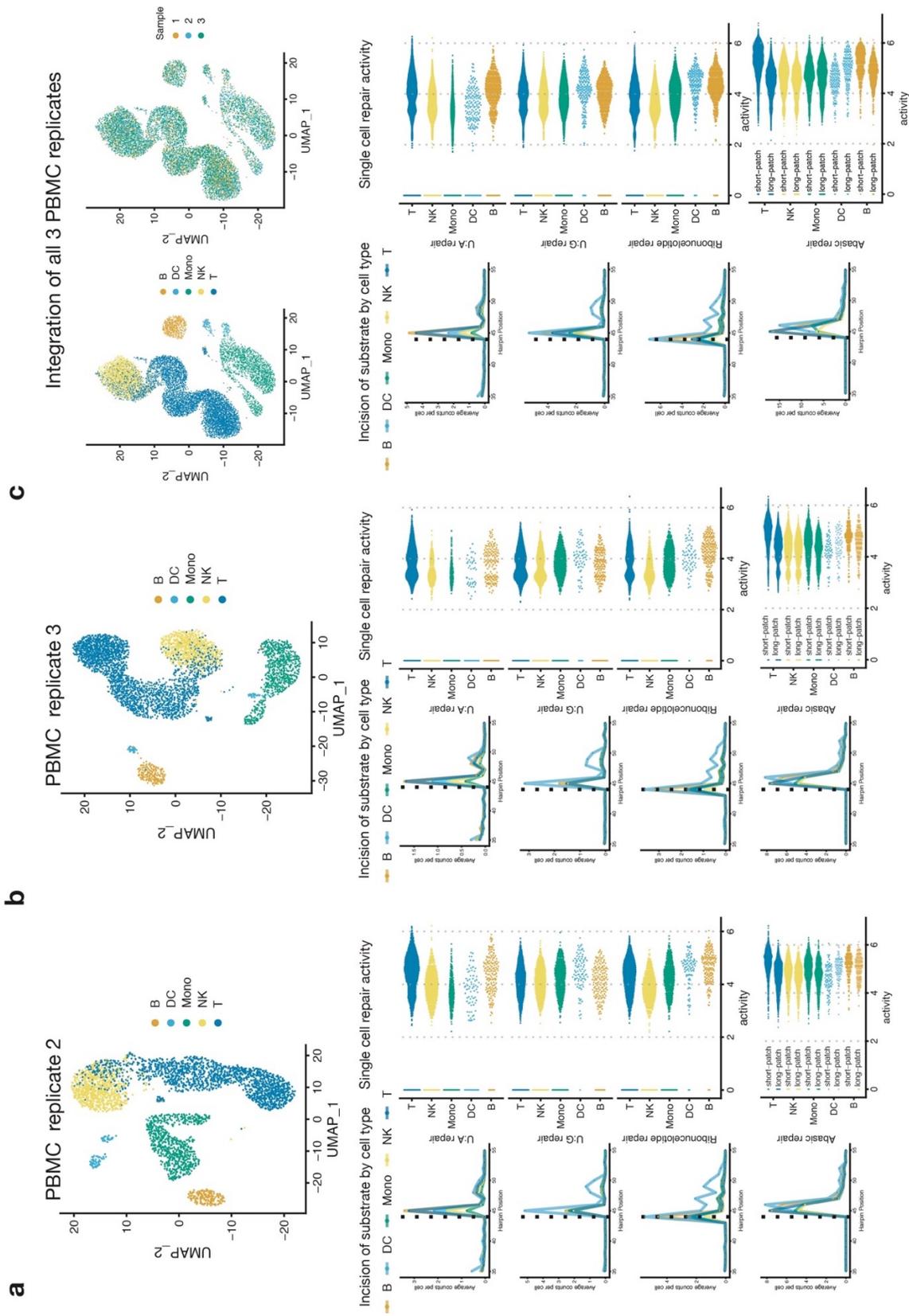
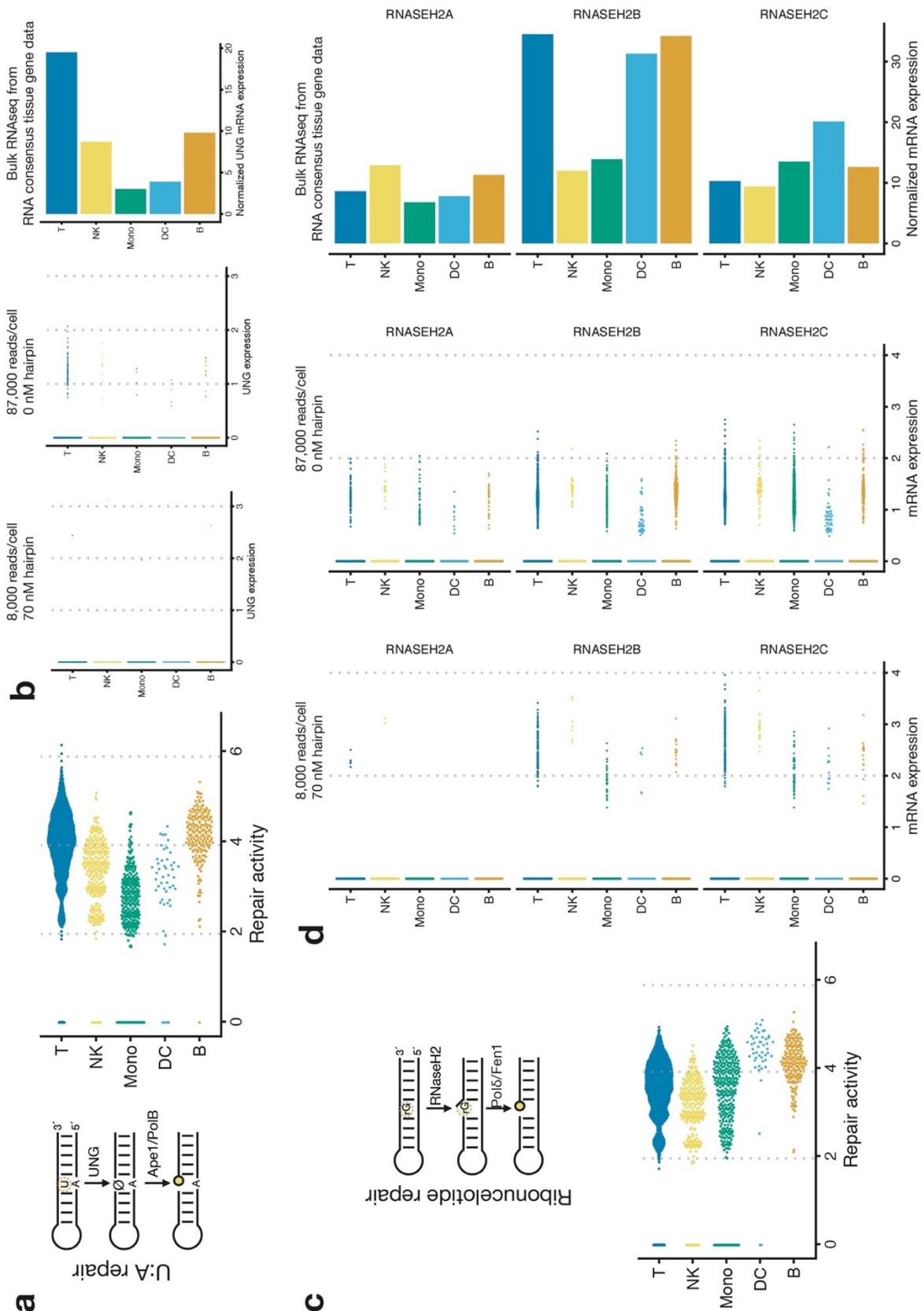


Figure 2.9: Biological replication of DNA repair phenotypes in human PBMCs

a-b. PBMCs were isolated from a single healthy human donor in two batches. 3' single cell gene expression and DNA repair activities were measured for each batch. mRNA expression was used to cluster cells and classify cell types. UMAP plot of cell types (top). Cell-type-specific counts of incision and processing (mean) are plotted against the position of the hairpin (left). Trends for single cell repair of U:A, U:G, riboG:C, and abasic:G substrates were similar across all three replicates (**Figure 2.8** and **a-b**).

c. Single cell gene expression and repair measurements for 3 replicates (**Figure 2.8** and **a-b**) were integrated into one Seurat object. mRNA expression was used to cluster cells. Cell types were identified separately for each sample. UMAP plot of cell types and sample number (top) indicate that cells cluster with cell types rather than with samples. Cell-type-specific counts of incision and processing (mean) are plotted against the position of the hairpin (left). Trends for single cell repair of U:A, U:G, riboG:C, and abasic:G substrates were similar to individual replicates (**Figure 2.8** and **a-b**).



a. Repair of a substrate containing a uracil:adenine (U:A) base-pair initiates with UNG-mediated removal of the uracil nucleobase followed by processing of the abasic site by Ape-1 and Pol β (left). Single-cell repair activities (natural logarithm of counts at the incision site divided by total counts for that cell multiplied by a scaling factor of 10^4) are plotted for each cell from each cell type (right) (Also in **Figure 2.8**).

b. UNG mRNA in single cells from our data (left) and data from 10x Genomics (middle) is not robustly detected in single cells at sequencing depths of 8,000 or 87,000 reads per cell. UNG expression from bulk RNA sequencing experiments (right) correlates with repair activities measured in single cells (**a**).

c. Repair of a substrates containing a riboG:C base-pair initiates with RNASEH2-mediated incision 5' of the ribonucleotide followed by processing by Pol δ and Fen1 (left). Single-cell repair activities (natural logarithm of counts at the incision site divided by total counts for that cell multiplied by a scaling factor of 10^4) are plotted for each cell from each cell type (right) (Also in **Figure 2.8**).

d. Single cell measurements of RNASEH2 subunit mRNA from our data (left) and data from 10x Genomics (middle) are not robustly detected at sequencing depths of 8,000 or 87,000 reads per cell. RNASEH2 subunit expression from bulk RNA sequencing experiments (right) is difficult to correlate to repair activities measured in single cells (**a**). The catalytic subunit RNASEH2A does not greatly vary between cell types. While the structural subunit RNASEH2B roughly correlates to RNASEH2 repair activity measured in single cells, this phenotype would be difficult to predict given that each subunit is required for enzymatic function.

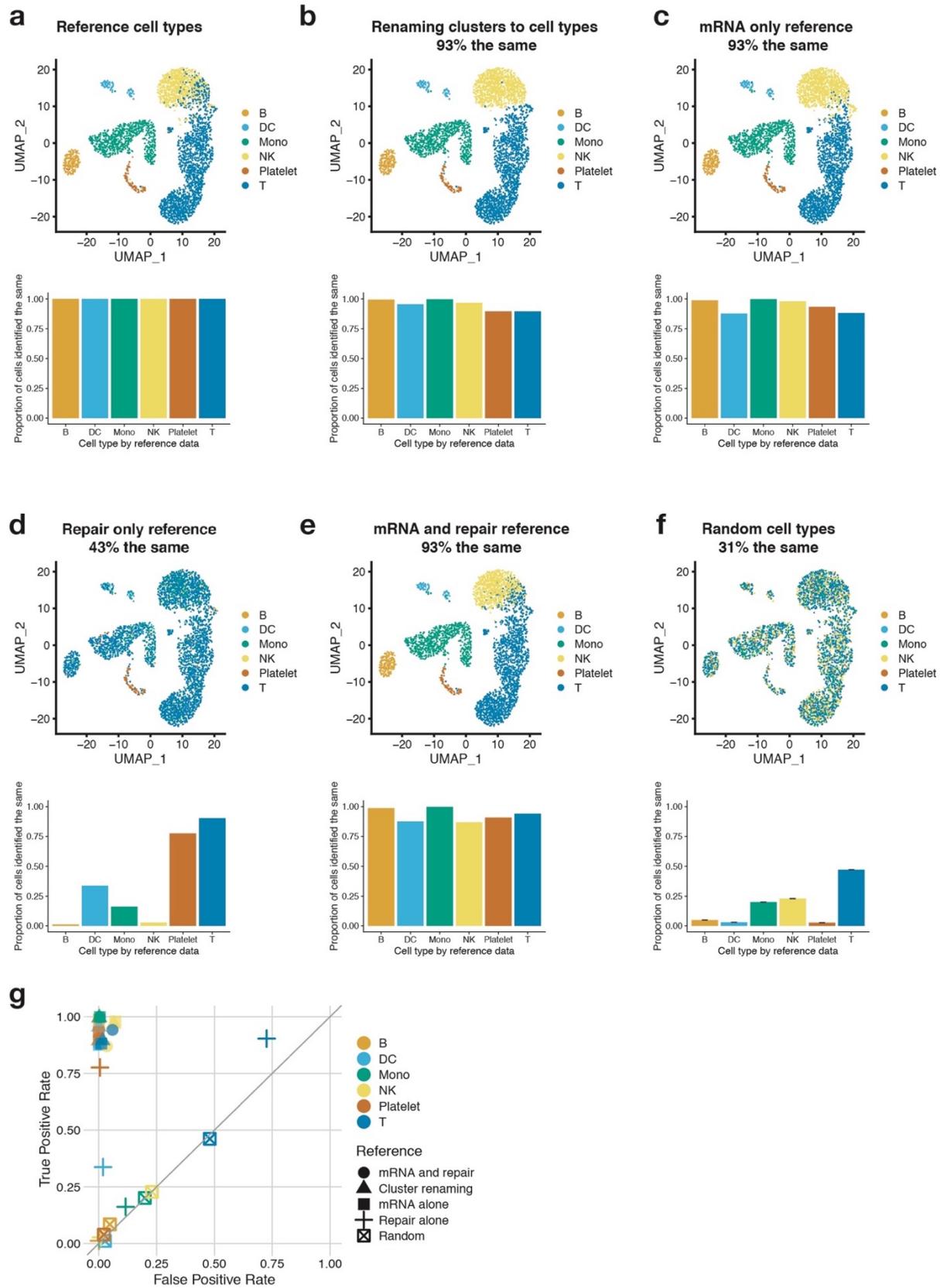


Figure 2.11: Cell-type classification using DNA repair measurements

a. PBMC UMAP and cell classifications for PBMC replicate 2 (**Figure 2.9**). Cells were classified using reference data from 10x Genomics and Seurat v3.0.0 FindTransferAnchors and TransferData functions. These cell classifications are used as the true cell type for other classification methods, however, these classifications may not be 100% accurate. The percentage of each cell type that is classified the same as this reference data (bottom) is 100% since it is compared to itself.

b. Seurat clusters (0-10) were renamed for the majority cell type marker from **a** in each cluster. UMAP plot of renamed cells (top). Using this classification method 93% of cells are classified the same as **a**. > 89% of each cell type were classified the same (bottom).

c. mRNA data in **Figure 2.8** was used as the reference for classifying cell types in PBMC replicate 2 (**Figure 2.9**) using Seurat as before (**a**). Using this reference, 93% of the cells were classified the same. > 87% of each cell type were classified the same (bottom).

d. DNA repair measurements alone from PBMCs (**Figure 2.8**) was used as the reference for classifying cell types PBMC2 replicate. Using only DNA repair for reference data, only 43% of cells were classified the same as **a**. Platelets lack DNA repair activities in Haircut (data not shown), which could contribute to their identification using DNA repair measurements alone as they are distinctly different from all other cells. > 90% of T cells were also identified using repair alone, however, T cells make up ~40% of all cells in the data and when classified using repair data alone, ~80% of all cells are classified as T cells, so by chance alone, T cells are more likely to be classified correctly. Dendritic cells make up only 2.6% of all cells in the data, however, 33% of them are classified correctly using repair activity data alone. DC have several unique repair signatures (**Figure 2.8** and **Figure 2.9**) that are likely to contribute to their classification using DNA repair activities alone.

e. Count matrices for mRNA expression and DNA repair were combined and used as a reference (from **Figure 2.8** data) to classify PBMC replicate 2 cells. Using this reference, 93% of the cells were classified the same as in **a**. > 86% of each cell type were classified the same (bottom).

f. Cell type labels from PBMC replicate 2 as defined in **Figure 2.9** were randomly reassigned to cell ids. Only 31% of cells were classified the same. More abundant cell types were more likely to be correctly classified by chance (bottom). Error bars represent the 95% confidence interval for 1000 independent samplings.

g. True positive rate ($\# \text{ of true positives} / (\# \text{ of true positives} + \# \text{ of false negatives})$) and false positive rate ($\# \text{ of false positives} / (\# \text{ of false positives} + \# \text{ of true negatives})$) for each cell type classified by each method. Cells classified using mRNA expression with or without repair had a high true positive rate and low false positive rate. Most cell types defined by repair alone had relatively equal true and false positive rates, with the exception of platelets that had a high true positive and low false positive rate and DC that had a relatively low true positive rate but a very low false positive rate indicating DNA repair measurements for these cell types may be helpful for classification.

CHAPTER III

SUMMARY AND FUTURE DIRECTIONS

This dissertation describes my main contribution to the single-cell field: measuring biochemical activities in single cells. I developed a single-cell assay that measures DNA repair enzyme activities simultaneously with single-cell gene expression called Haircut. Single-cell enzymatic measurements provide unique and orthogonal information to gene expression and may prove useful in characterizing single-cell phenotypes independent of gene expression. In **Chapter II**, I applied Haircut in primary human immune cells and found known and unknown differences in DNA repair between immune cell types. I also used Haircut to measure DNA repair activities and gene expression in primary human immune cells from a preliminary cohort of individuals with T21 (**Appendix B**). In **Appendix A** I described the steppingstones towards making a single-cell DNA repair assay, and in doing so, I developed a multiplexed DNA repair assay in bulk cell-free extracts. In this final chapter, first I will explore improvements and potential modifications to Haircut. Second, I will explore new methods to expand the single-cell biochemistry toolkit by measuring other enzymatic activities (e.g. protease, kinases, phosphatases, etc.). Finally, I will discuss the applications of single-cell biochemistry in the fields of precision medicine, developmental biology, and pharmaceutical sciences.

Improvements to Haircut

Expand the substrate repertoire for Haircut

In its current form, Haircut is able to measure DNA repair on three main substrates: uracil, ribonucleotides, and abasic sites. These substrates represent three of the most common DNA damage events that occur in cells; and are the most robust repair activities. However, there are eleven known human DNA glycosylases, and currently Haircut can measure the activity of one to

four glycosylases at most. While the substrate requirements for DNA glycosylases fit within the constraints of the current Haircut design, there are two main challenges we need to overcome to measure their activity: specificity and sensitivity.

Increasing Haircut specificity

Currently, Haircut does not have the specificity to measure the repair of some bulky lesions, like thymine glycol and ethenoA. Polymerases cannot use thymine glycol or ethenoA as a template and therefore these lesions produce truncated extension products that look identical to a strand incision events catalyzed by BER enzymes in Haircut data (**Figure 3.1a**, left and middle). From data I collected in bulk and in single cells, these lesions create signals in Haircut in the absence of repair enzymes (**Figure A.3a** and **Figure 2.3**).

One solution to this challenge is to add a translesion synthesis step to the library preparation. *In vitro* bypass of the lesions to remove the background noise created by stalled polymerases would allow us to measure only strand incision events catalyzed by DNA repair enzymes. There are several translesion DNA polymerases that are commercially available. Following single-cell encapsulation and reverse transcription and prior to end repair (**Figure 2.2**), we could add a translesion synthesis and extension step so that polymerase-blocking lesions are bypassed producing full-length substrates that are carried through the remaining library preparation steps (**Figure 3.1a**, right). Substrates that have been repaired by cellular DNA repair enzymes would remain truncated allowing us to accurately measure these repair events.

Another solution is purely computational. Haircut collects signals from cells and from empty drops in every single-cell experiment, so we have a large collection of negative control data that represents non-enzymatic signals on Haircut substrates (**Figure 2.3**). There are several methods available that remove background noise from single-cell mRNA data (Fleming, Marioni,

and Babadi 2019; Young and Behjati 2018). These methods are somewhat useful in removing background signals from our data; but since they are optimized for mRNA data, their applicability to DNA repair data may be limited. Haircut data does not suffer from the same dropout rates as single-cell mRNA data, and thus different methods to remove background signal should be developed. I already use a simplistic, non-quantitative method for determining enzymatic signals in our data by comparing the average counts per position in drops with cells and the same number of empty drops (**Figure 2.3**). However, a more probabilistic method should be developed in order to systematically quantify cellular signal over signals that originate from empty drops. A simple method to do this is to normalize signal at each position to empty drops and any signal that is less than one (i.e. higher in empty drops), would be considered background noise (**Figure 3.1b**). Another method is to examine the signal distributions from empty drops and cells at each position and determine if the signal in cells is significantly different from the signal in empty drops. I have tried a variety of these simplistic methods for determining biological signals with some success; however, removing technical noise via alterations in the molecular biology steps would produce higher confidence results.

Increasing Haircut sensitivity

Currently, Haircut lacks the sensitivity to measure repair of some DNA repair substrates. For example, the inosine containing substrates do not generate any detectable signal at expected repair positions in single cells (**Figure 2.3**). Presumably this is because any glycosylase activity on the inosine substrates is below the sensitivity of the assay. DNA repair microarray studies (described in further detail in **Chapter I**) have shown that inosine and 8-oxoG substrates are repaired on the order of 10 times less than uracil and abasic contain substrates in cell-free extracts (Pons et al. 2010a). On a single-cell scale, we measure 10-100 repaired molecules per cell for the

most robust repair activities (i.e. uracil, ribonucleotides, and abasic sites). If repair events are reduced by a factor of 10, then we can expect to measure the repair event 1-10 times per cell. When we down sampled our data to reflect 1-10 repair events per cell (~150 reads per cell), we were unable to classify cells in a knockout experiment; indicating that repair events captured in low abundance are not sufficient to produce enzyme-specific signals (**Figure 2.4**). To capture low abundance repair events, we need to increase the sensitivity of the assay.

One simple way to increase Haircut's sensitivity is to sequence the samples at a higher depth. We sequenced ~3500 reads per cell and >85% of those reads align to hairpins substrates, which is sufficient to quantify abundant repair events. If we sequenced at a higher depth, it is possible that we would also be able to quantify low abundant repair events. However, most of the substrates included in the assay are not repaired (i.e., converted to products). Sequencing at a higher depth may result in sequencing more unrepaired substrates. One way to combat this, is to include more substrates in the assay. I estimate that 0.01-0.05% of the substrates added to the assay are sequenced and of those sequenced 5-25% are products of repair for uracil and ribonucleotide substrates (**Figure 2.7**). When I increase the substrate concentration, the proportion of substrates that were sequenced remained constant. However, the proportion of sequenced molecules that were products of repair increase as substrate concentration increased (**Figure 2.7**). As such, it is possible that including more substrates for low abundant repair events would increase our ability to measure those enzymatic activities.

Another strategy to increase the sensitivity of Haircut is to decrease the background signal. Most of the signal that is sequenced in Haircut libraries is present in empty drops or at positions that are not specific to enzyme activity (**Figure 2.3**). If background signal is removed, then the majority of sequencing results would come from repair events. One way I tried to remove

background was to include a 5'-biotin on the substrates. The idea is that strand incision events would liberate the 5'-end of the substrates and we could remove unrepaired substrates and liberated 5'-ends with a streptavidin pull down. However, this method did not reduce background noise (**Figure 2.3**) possibly because the proportion of streptavidin beads to biotin in the sample was not appropriate to fully remove the unrepaired substrates or because the signal was generated from the newly synthesized second strand/reverse transcription reaction that occurs in the drop.

Most sequencing libraries, by nature, are double stranded. In Haircut, the single-stranded substrate is copied during the reverse transcription step of the library preparation and only the newly synthesized strand is PCR amplified and compatible with sequencing (**Figure 2.2**). It is possible that the streptavidin cleanup that I used only removed unrepaired substrates, but the newly synthesized strand created after reverse transcription remained and was carried through the library preparation. One way around this challenge is to only PCR amplify the substrates themselves, rather than the newly synthesized strand. To do this, the assay needs to be redesigned in five ways (**Figure 3.2**).

- 1.** Remove the 3'-C3 spacer on the substrates. The 3'-C3 spacer acts as a blocking group on the end of the substrate preventing it from degradation and preventing the substrates from being extended during reverse transcription. However, if we remove this blocking group, then the substrates themselves can be extended to incorporate the cell barcode onto the substrate molecule.
- 2.** Prevent the reverse transcription primer from extending and copying the substrates. The reverse transcription primer present in the drop needs to be prevented from copying the full-length substrate. To accomplish this, we need to include a reversible, polymerase halting lesion in the substrates that is not recognized or repaired by cellular machinery. One option, although expensive, is to include a pyrimidine dimer in a single-strand region of the substrate. The pyrimidine dimer would block polymerases and since its present in single-strand DNA, it cannot

be removed by cellular NER machinery. **3.** Include a 5'-biotin on the substrates. The unrepaired substrates can be removed by a streptavidin pulldown and the repaired products can proceed through the library preparation. **4.** Remove the polymerase-blocking lesion present in the substrates. Following single-cell barcoding of the substrates in the drop, the polymerase blocking group on the substrate needs to be removed. This could be done by translesion synthesis or in the case of a single-strand pyrimidine dimer, by a photolyase (Selby and Sancar 2006). **5.** Add second strand synthesis to library preparation. Finally, the barcoded and purified repair products can be amplified using a standard library preparation: second strand, A-tailing, and PCR (**Figure 3.1c**). These steps may improve Haircut's sensitivity by preventing PCR amplification of full-length substrates or truncated RT products.

Measuring direct reversal

In contrast to BER, direct reversal DNA repair does not result in a strand incision event. The alkylation DNA lesion, O6mG, is removed by the single-use methyltransferase, MGMT (described in detail in **Chapter I**). MGMT activity can be measured using a restriction enzyme digest that is inhibited by O6mG, and I show that MGMT activity can be measured in bulk extracts on a Haircut substrate (**Figure A.2g**). However, I was unable to measure MGMT activity in single cells (**Chapter II, Figure 2.3**). To measure direct reversal in single cells, two main issues need to be addressed. **1.** I need to confirm that Pst1 restriction does not cut the O6mG-containing substrate. While other studies have used Pst1 restriction to measure removal of O6mG (Encell and Loeb 1999b), my readout is through PCR and sequencing. A few cleavage events on the O6mG substrate can be amplified and sequenced resulting in non-specific 'repair' events. **2.** I need to have high-quality O6mG substrates. In bulk, the O6mG substrate was commercially synthesized and I measured MGMT-specific activity using this substrate (**Figure A.2g**). For the single-cell

experiments, I synthesized the O6mG substrate myself. While I followed the manufacturer's instructions for synthesizing the oligonucleotide, the final product did not undergo rigorous quality control. The O6mG phosphoramidite requires special handling during the deprotection step of synthesis, including a five-day incubation period in the dark. The commercially produced O6mG substrate failed quality control several times before I received the final product, indicating that its synthesis is not straight forward. Using a high quality O6mG substrate may resolve the non-specific PstI cleavage events in single cells.

Other DNA repair pathway substrates

To fully assess DNA repair in single cells, I would like to measure all of the DNA repair pathways. Base excision repair, ribonucleotide excision repair, and direct reversal substrates are simple to design and manufacture since they are relatively small oligonucleotides of only 20 base pairs. The substrates for nucleotide excision repair (NER) need to be much larger (~200 base pairs for cell-free extract activity). The substrates for mismatch repair (MMR) need to be larger yet (~1kb) and include a nick or ribonucleotide to direct the repair machinery to the correct strand. The substrates for double-strand break (DSB) repair are challenging since this repair pathway joins two separate molecules into one. In its current design, Haircut may be able to measure NER on large substrates, especially if the extracts are supplemented with ATP and the lesions are either bypassed or removed prior to library preparation to prevent non-specific truncated products (**Figure A.3b**). For MMR and DSB repair, Haircut needs to be changed in one significant way: substrates need to be transfected into cells and then recovered to measure DNA repair.

The substrates for MMR and DSB repair are quite large, so using them with the current Haircut design is not possible since they cannot be separated from the mRNA by size. Host cell reactivation assays use reporter plasmids to measure MMR and the two pathways of DSB repair:

non-homologous end joining and homologous recombination (Nagel et al. 2014). A similar method could be used in the context of a single-cell mRNA assay where the reporter mRNA is captured alongside cellular mRNA. The presence of the mRNA and/or point mutations within the mRNA could be used to quantify repair on a single-cell level. The main drawbacks to this method are that it is difficult to do on primary samples since it requires transient transfection of the reporter plasmids; and DNA repair cannot be measured in all cells and all substrates since transfection is heterogenous and not all cells will contain the reporter plasmid.

Measuring DNA repair in millions of cells

The iteration of Haircut I describe in this dissertation uses the 10x Genomics platform to measure gene expression and DNA repair activities. The 10x Genomics system and other droplet-based single-cell methods are very efficient at studying a few thousand cells (~2K - 8K). There are now methods that use the droplet-based systems in new ways to measure gene expression from >100K cells (Datlinger et al. 2019). To expand Haircut to measure >100,000 cells, I could alter the single-cell combinatorial fluidic indexing (Datlinger et al. 2019) or other combinatorial indexing protocols (Junyue Cao et al. 2017; Rosenberg et al. 2018). Rather than use cell-free extracts to measure DNA repair enzyme activities, small hairpin substrates could be transfected into live cells. Repair of the substrates would take place within the cell over a short time course (0.5-2 hours). Cells could then be fixed and permeabilized and hairpins and mRNA could be captured using similar combinatorial indexing protocols that have already been described (Datlinger et al. 2019; Junyue Cao et al. 2017; Rosenberg et al. 2018) (**Figure 3.3**).

One motivation behind expanding the cell numbers of Haircut is to identify new DNA repair regulators in a CRISPR screen. Several protocols have been developed to sequence CRISPR guide RNA (gRNA) in single-cell sequencing (Mimitou et al. 2019) and they are compatible with

the current version of Haircut and the proposed combinatorial indexing version of Haircut. The approach for the screen is fairly simple. **1.** Transfect cells with a library of CRISPR guides, including guides that target known DNA repair regulators. **2.** Once cells have incorporated CRISPR guides, they can be transfected with DNA repair substrates for the appropriate amount of time. **3.** Capture single-cell CRISPR gRNA and DNA repair substrates (**Figure 3.4**). **4.** Sequence the samples and identify the known DNA repair effectors and confirm that DNA repair measurements reflect changes in known DNA repair proteins. This method would produce a robust dataset where known and unknown regulators of DNA repair could be identified.

This approach only measures CRISPR gRNA and DNA repair activities, not mRNA abundance. Sequencing CRISPR gRNAs and DNA repair substrates allows us to significantly reduce sequencing cost, while still having high coverage of the CRISPR gRNAs and a quantitative functional readout for DNA repair. To illustrate, 10x Genomics recommends users sequence samples at 50,000 reads per cell. If we want 100X coverage (meaning that each cell is likely to be transfected with a single gRNA and each gRNA is in approximately 100 cells) of a targeted CRISPR library for the base excision repair pathway (KEGG: hsa03410) that contains 33 genes and we include four gRNA per gene, then we would need to sequence 13,200 cells. At 50,000 reads per cell, we would need to get 660 million reads that cost \$50 per 10 million reads for a grand total of \$3,300 of sequencing cost. However, if we were to only sequence CRISPR guides and DNA repair activities via Haircut, we only need 2,000 reads per cell to quantify DNA repair activities (**Figure 2.4**), which totals to \$150 in sequencing cost. Since this screen is designed to identify effectors of DNA repair, the functional readout that Haircut provides is arguably a cheaper and more accurate readout than global gene expression alone.

Improving Haircut specificity

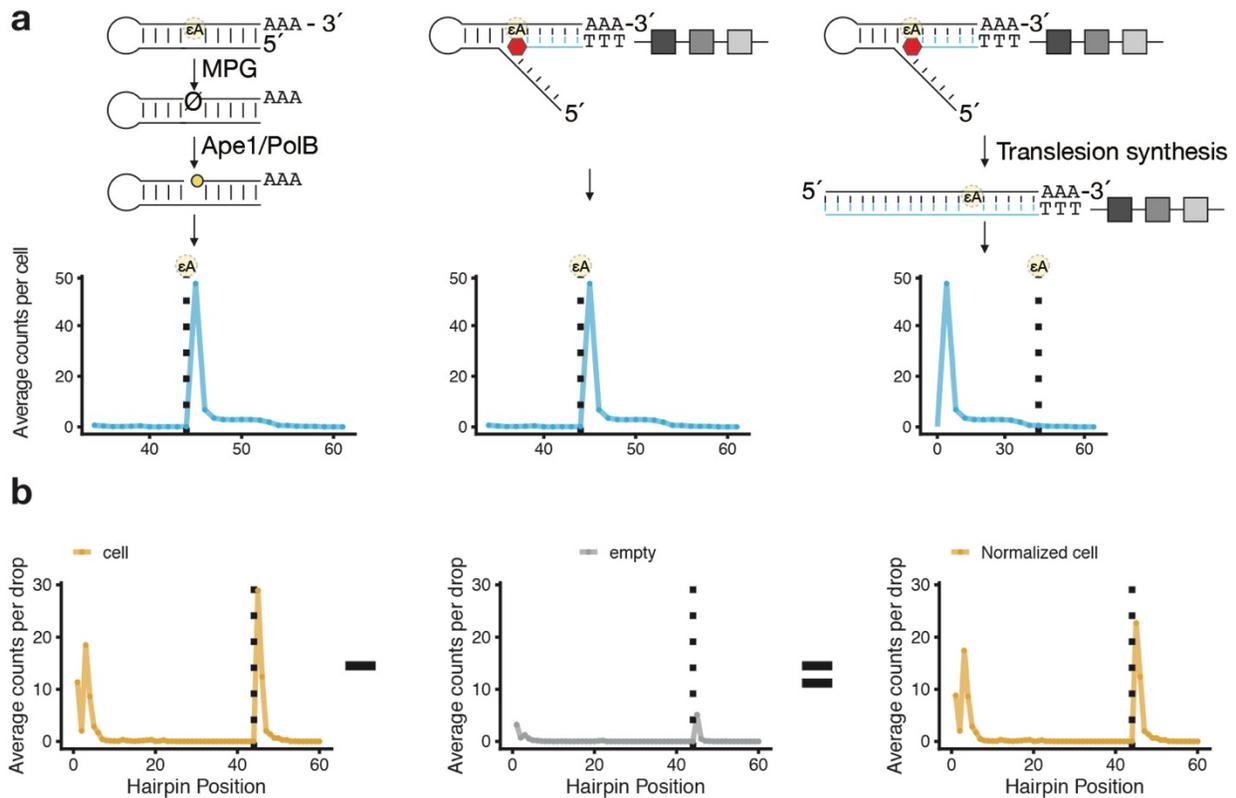


Figure 3.1: Improving Haircut specificity through translesion synthesis and computationally

a. Haircut signals are dependent on strand incision events (left); however, if a bulky lesion prevents second strand synthesis or reverse transcription (RT), it can lead to a truncated second strand product (middle). These truncated products look identical to Haircut signals but are not dependent on enzymatic activity. Including a translesion synthesis step in the library preparation would reduce nonspecific signal by bypassing bulky lesions and synthesizing full-length second strand products (right).

b. Haircut specificity could be improved by removing non-specific signals computationally. Each single-cell experiment has thousands of negative control droplets that do not contain cells. We could subtract this background signal (middle) from signals in drops with cells (left) to measure only cell-associated repair events (right).

Improving Haircut sensitivity

Oligo dT primer in single-cell experiment captures DNA repair substrates

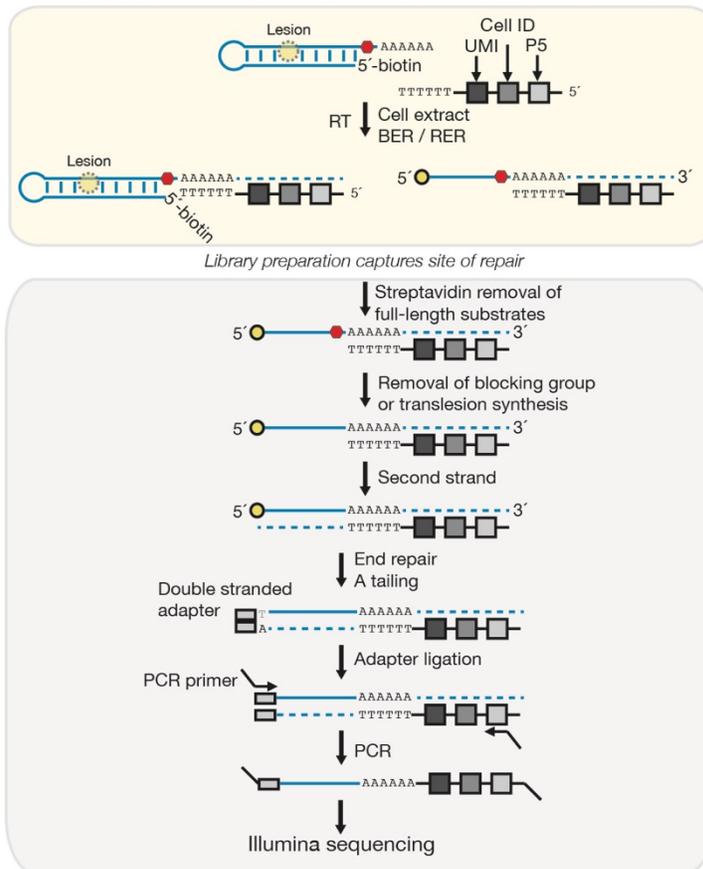


Figure 3.2: Improving Haircut sensitivity by removing unrepaired substrates

To improve Haircut sensitivity, we could sequence only strand incision events rather than sequencing both unrepaired substrates and incised products. To achieve this in the 10x Genomics system, we have to prevent the RT primer from extending onto the hairpins by including a polymerase-blocking lesion on the substrates. The reverse transcriptase extends the 3' end of the DNA repair substrate on the RT primer template to incorporate the cell barcode and UMI onto the substrate (top). Next, we remove the unrepaired substrate and proceed with the library preparation with the repaired substrates only. To complete a standard library preparation (bottom), the blocking lesion needs to be removed or bypassed (middle). This method could decrease the background signal of the assay as an effort to improve the assay sensitivity.

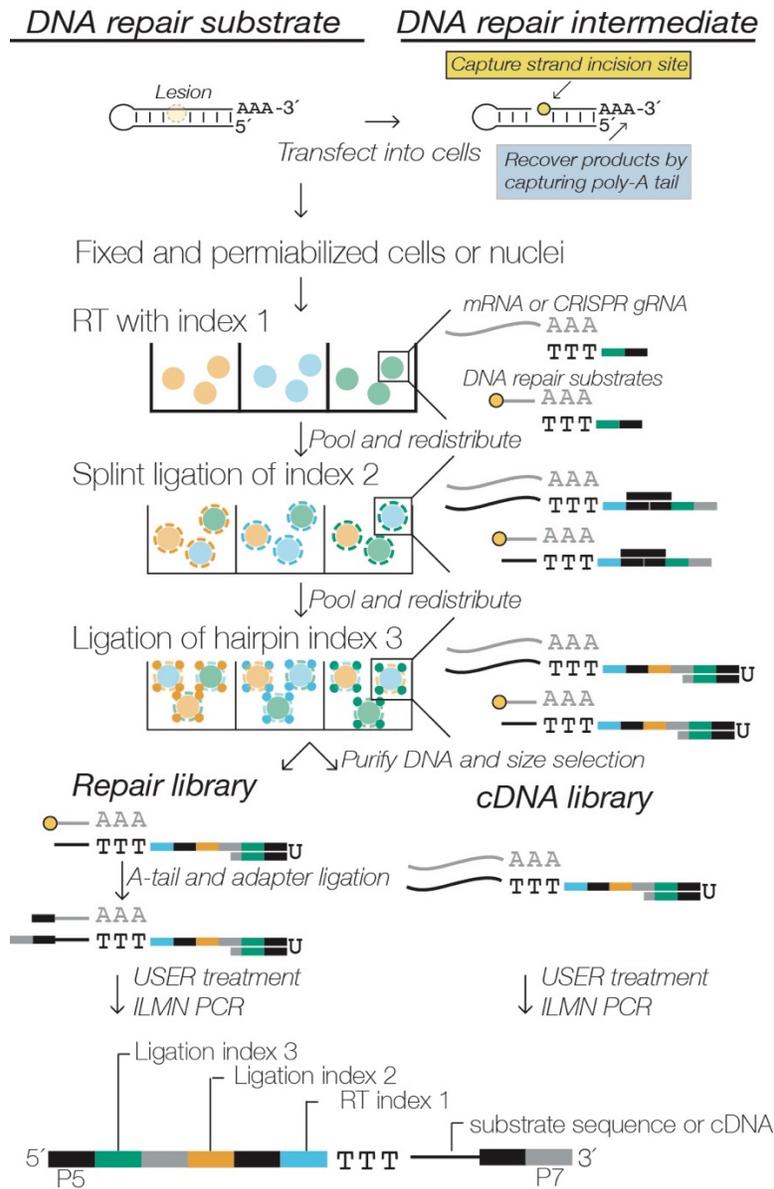


Figure 3.3: Combinatorial indexing protocol to measure DNA repair and RNA

To measure DNA repair in currently available combinatorial indexing protocols, DNA repair substrates need to be transfected into live cells for a short time course (top). Cells are then fixed and permeabilized. DNA repair substrates and RNA (either mRNA or CRISPR gRNA) in each cell are tagged via a series of barcoding reactions. The approach allows for inexpensive barcoding of millions of cells.

Expanding the single-cell biochemistry toolkit

One of the true innovations of Haircut is that we show that it is possible to semi-quantitatively measure enzymatic activity at single-cell resolution. DNA repair activities are simple to measure since the substrates (DNA hairpins) can be read out by DNA sequencing. However, measuring other enzymatic activities introduces unique technical challenges.

Measuring the activity of RNA modifying enzymes

Many cellular RNAs are heavily modified (Frye et al. 2018); however we do not fully understand how RNA modifications are regulated. RNAs are differentially modified during development, in cancers, and throughout the circadian rhythm (Zhao, Roundtree, and He 2017). High throughput sequencing methods have been developed to map RNA modifications throughout the transcriptome (Molinie and Giallourakis 2017; Riemondy et al. 2018) and single-cell resolution of RNA editing events indicate that RNA editing is differentially regulated spatially and temporally during development (Lundin et al. 2020). RNA modifications could potentially be measured in single-cell RNA sequencing experiments given that there is sufficient read depth and the modifications are within a few hundred bases of the 3' or 5' end of the mRNA. However, given the sparsity of current single-cell mRNA sequencing techniques, finding modifications events on low-abundant mRNA or other cellular RNAs is unlikely.

Developing RNA substrates to measure RNA modifying enzyme activities is relatively simple. Substrates for RNA editing enzymes and m6A-RNA methyltransferases are small RNA oligonucleotides (Y. Wang, Park, and Beal 2018; F. Li et al. 2016). If designed to be compatible with the 10x Genomics platform, the substrates can be captured alongside mRNA and then separated by size, similar to Haircut and other multimodal single-cell methods (Amanda L. Richer et al. 2020a; Mimitou et al. 2019). The challenge is to separate or identify the modified RNA

products from the unmodified RNA substrates. For m6A modifications, the methylated RNA substrates can be affinity purified using m6A antibodies (Molinie and Giallourakis 2017) or the substrates can be digested using an m6A-sensitive RNA endoribonuclease that recognizes ACA motifs (Z. Zhang et al. 2019) (**Figure 3.4a**). Either method allows us to distinguish between modified and unmodified substrates. RNA editing substrates are reverse transcribed and the editing event would produce a point mutation and can be identified in the sequencing data (**Figure 3.4a**). Either of these methods could be developed to measure RNA modifying enzyme activities in heterogeneous cell populations.

Measuring protease activity

Intracellular and extracellular proteases play an important role in maintaining cellular homeostasis, during development, and in immunity. Caspases are tightly controlled enzymes that regulate apoptosis, necrosis, and inflammation (McIlwain, Berger, and Mak 2013). Other intracellular proteases, like cathepsins, help regulate protein turnover and when overexpressed in cancer can lead to unregulated cell migration (Mohamed and Sloane 2006). Granzymes regulate cytotoxic lymphocyte function. Extracellular proteases like matrix metalloproteinases are upregulated in almost every type of cancer and they can regulate cell migration by remodeling the extracellular matrix (Egeblad and Werb 2002). The diverse and important functions of proteases make them an ideal target to measure at single-cell resolution, especially in heterogeneous tumors.

Unlike the nucleic acid substrates for DNA repair and RNA modifying enzymes, proteases require a peptide substrate that cannot be directly readout via sequencing. To measure protease activity, we designed peptide-DNA conjugates that can be generally applied to many peptide substrates (**Figure 3.4b**). The key challenge to overcome by using peptide-DNA conjugates is to reliably separate the products from the unreacted substrates. For proteases, peptide substrates with

a C-terminal biotin and an N-terminal azide are conjugated to a 5'-alkyne barcoded DNA oligonucleotide via copper-assisted azide-alkyne cycloaddition. The barcoded DNA oligonucleotide is captured and tagged with a cell barcode in a single-cell mRNA experiment. Unreacted peptide substrates are removed via a streptavidin pulldown. The flow through contains only the cleaved products of protease activity. Many proteases have at least some sequence specificity, which allows us to include many different peptide substrates that are only cleaved by specific cellular proteases.

One consideration to measuring protease activity (and other peptide-DNA conjugates), is the stability of the peptide construct in cell lysate. In measuring protease activity, we want to ensure we are measuring specific protease activity, not general protein degradation via relatively non-specific cellular proteases. Peptide substrates used to measure kinase activity are relatively unstable in cell lysate, although some of the instability can be mitigated by modifications to the peptide (Proctor et al. 2012; S. Yang et al. 2013). The N-terminal modifications on our substrates may provide some protection since some secondary structures on the N-terminal increase the stability of kinase substrates in cell lysate (S. Yang et al. 2013). Many proteases have some sequence specificity and ensuring that we design peptide substrates that do not cross react with other proteases is essential. However, given that our substrates are stable and specific, measuring protease activity in single cells is feasible with few modifications to the current assay design.

Measuring E3 ubiquitin ligase activity

Regulation of protein stability and targeted degradation is an important process in maintaining cellular homeostasis. The ubiquitin proteasome system is a crucial regulator of controlled protein degradation in cells. Dysregulation of this system is a hallmark of cancer and an important regulatory process during development (D. Wang et al. 2017). Small peptide

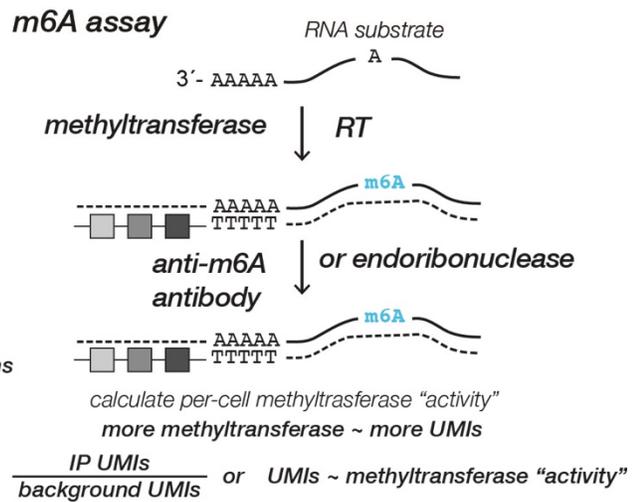
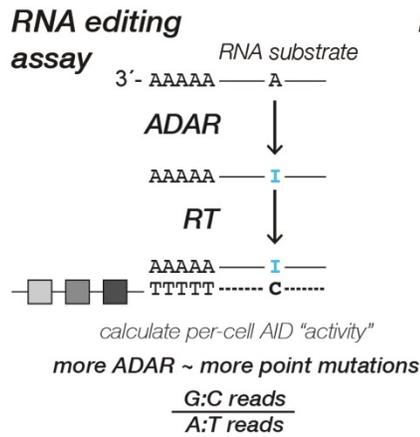
substrates have been used successfully to measure E3 ubiquitin ligase activity in cell extract and in single cells (Melvin et al. 2013; Houston et al. 2017). Similar substrates can be designed as peptide-DNA conjugates (as described above). To isolate ubiquitinated products from unreacted substrates, we can use one of several commercially available anti-ubiquitin antibodies (**Figure 3.4b**).

Measuring signal transduction pathways

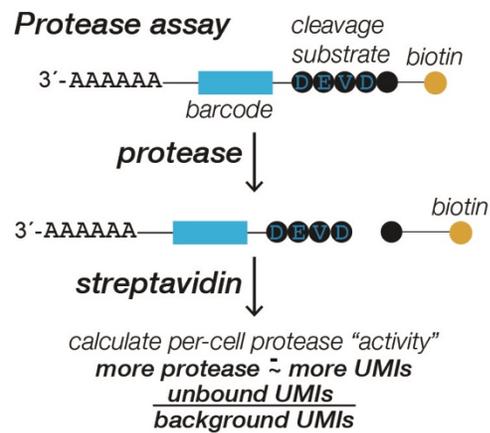
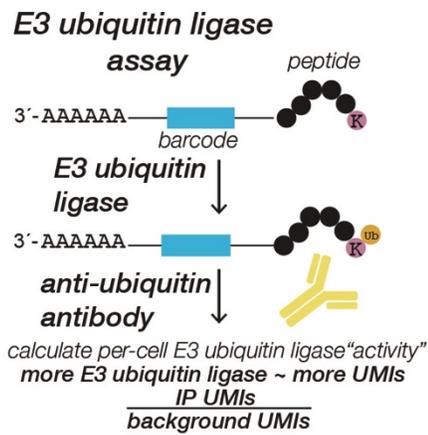
Intracellular signaling pathways achieved through protein kinase and phosphatase activity are essential for cell viability and are altered in essentially all cancers. These pathways allow cells to respond to extracellular signals, control cell cycle and division, and alter cell metabolism. There is immense heterogeneity in protein kinases. Protein kinases can be differentially expressed and their activities can be ‘leaky’ meaning that they do not respond to extracellular signals in a uniform fashion (Voliotis et al. 2014; R. Zhang et al. 2017).

To measure protein kinases and phosphatases in single cells, the substrates are peptide-DNA conjugates. Protein phosphatase substrates contain a synthetically available phosphorylated amino acid and protein kinase substrates are unmodified. The amino acid sequence is designed to target the activity of specific kinases and phosphatases of interest. To separate products of kinase or phosphatase activity from the substrates, we can affinity purify the phosphorylated peptide using anti-phospho antibodies specific to the substrates of interest (**Figure 3.4c**). There are several anti-phosphotyrosine, anti-phosphoserine, and anti-phosphothreonine antibodies that are commercially available, allowing us to design an array of kinase or phosphatase substrates. For each activity, the amount of phosphorylated and unphosphorylated peptide can be measured giving us a ratio of substrate to product for each activity, which may be important to quantify these activities.

a RNA modifying enzymes



b Protein turnover



c Signal transduction

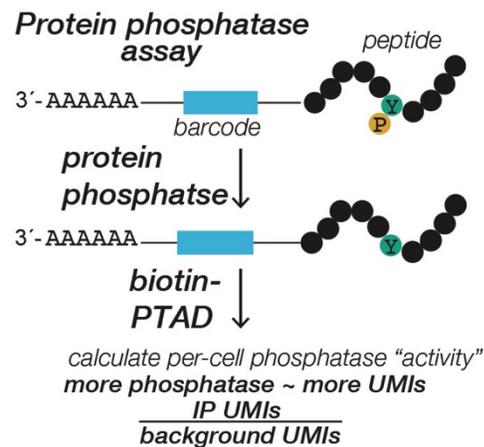
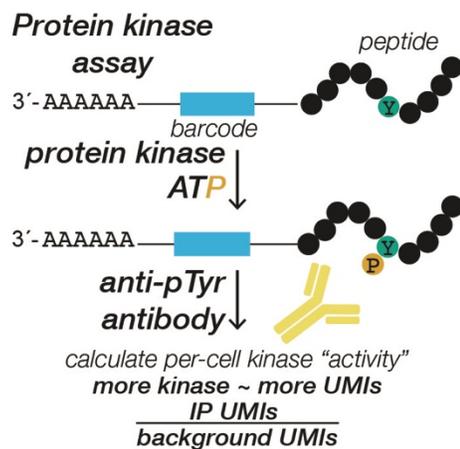


Figure 3.4: Expanding the single-cell biochemistry toolkit

a. We can measure RNA modifying enzymes using RNA substrates that contain a polyadenylate or another defined 3' end. RNA editing activity of ADAR can be measured by quantifying point mutations caused by the presence of inosine (left). The activity of methyltransferases that add the m6A modification to RNAs can be quantified by affinity purifying m6A-containing substrates with an anti-m6A antibody. m6A modified substrates can also be isolated by digesting non-methylated RNA substrates with a specific endoribonuclease that is inhibited by the presence of m6A (right).

b. The activity of enzymes involved in protein turnover can be measured using peptide-DNA conjugates. The DNA portion of the conjugates contains a polyadenylate or other defined 3' end and a substrate-specific barcode. E3 ubiquitin ligase activity can be measured by affinity purifying ubiquitinated substrates with an anti-ubiquitin antibody (left). Protease activity can be measured by removing intact biotin-containing substrates by streptavidin pulldown (right).

c. The activity of enzymes involved in signal transduction can be measured using peptide-DNA conjugates. Protein kinase activity can be quantified by isolated phosphorylated substrates using anti-pTyr, anti-pSer, or anti-pThr antibodies (left). Protein phosphatase activity can be quantified capturing non-phosphorylated substrates either by affinity purification by antibodies recognizing non phosphorylated substrates or through biotin-conjugated PTAD that will covalently bind to non-phosphorylated tyrosine residues (right).

Applications for single-cell biochemistry

If the outlined single-cell methods to measure enzymatic activities across several classes of enzymes are reduced to practice, we would have an expansive toolkit to measure biochemical phenotypes in single cells. Measuring the heterogeneity of enzymatic function has applications that span several fields of study and could greatly expand our current understanding of these complex cell systems. I have outlined several applications below that are of interest to me.

Single-cell biochemistry in precision medicine

Heterogeneous diseases, like cancer, are notoriously difficult to treat. Not only are cancers derived from heterogeneous mutations from person to person, but cells with dozens of different mutational patterns and phenotypes can exist within a single tumor. The heterogeneous nature of cancer is one of the reasons why it is difficult to treat. While cancer is arguably the most studied complex human disease, it is estimated that 90% of drugs are ineffective for 50% of patients with complex diseases (Shalek and Benson 2017). There is a huge need for methods to enable data-driven treatments on an individual basis, also known as personalized medicine.

There is no shortage of data that indicate how the mutational landscape of cancers should dictate the course of treatment (Amanda L. Richer et al. 2015). However, many of the current clinical methods used to determine treatments only measure targeted mutational patterns of known oncogenic alterations. Single genetic changes can have significant effects on global gene expression and cell function, and gene expression studies in cancer cells have found pathway-level functional differences present within individual tumors. This intratumoral heterogeneity is thought to be a major driver in resistance to chemotherapies (Shalek and Benson 2017). The use of single-cell methods can illuminate how a mixture of malignant and the surrounding non-malignant cells respond to conventional chemotherapies (K.-T. Kim et al. 2016; Shalek and Benson 2017).

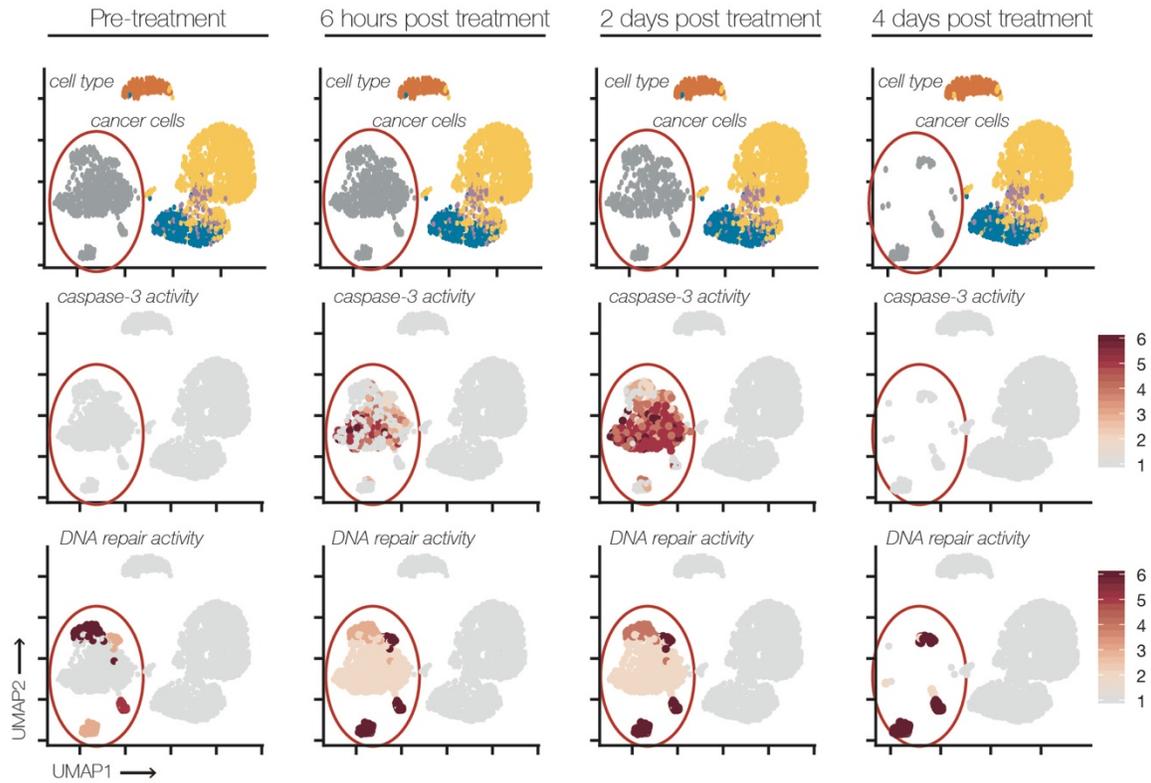
Single-cell biochemical assays could be used to directly probe the functional response of individual cancer cells to chemotherapies. For example, a novel combination of the BCL-2 inhibitor, venetoclax, and the DNA damaging agent, azacitidine, shows preclinical efficacy in treating acute myeloid leukemia (AML) (Pollyea et al. 2018). To test the efficacy of this drug combination using our assays, we can measure DNA repair capacities using Haircut and measure apoptosis through the caspase-3 protease activity. One prediction is that cells with high DNA repair activity may have low caspase-3 activity, showing that high DNA repair activity can result in cell survival (**Figure 3.5a**). Another example is the combined use of a WEE1 kinase inhibitor and a platinum-based DNA damage agent, cisplatin, in the treatment of lung cancer (A. L. Richer et al. 2017). This combination therapy is effective in transgenic mouse models, in part because WEE1 inhibition inhibits activation of the G2/M checkpoint causing DNA damage induced cell death. In this example, we can measure caspase-3 activity, DNA repair activity, WEE1 kinase activity, and the activity of other G2/M protein kinases and phosphatases that may regulate cell survival. This strategy enables us to evaluate the therapy from the main competing regulatory mechanisms that may determine its efficacy in heterogeneous samples. The G2/M checkpoint is a tightly regulated pathway that is often dysregulated in cancers, and by using a panel of kinase and phosphatase substrates, we can potentially measure the activity of many enzymes throughout the pathway to determine if WEE1 inhibition is effective when combined with DNA damage. Through comprehensive measurements of the G2/M pathway, DNA repair activities, and apoptosis, we can thoroughly measure the efficacy of this therapy at single-cell resolution.

Single-cell methods have also been used to evaluate treatments and heterogeneity of other complex human diseases. For example, rheumatoid arthritis is an autoimmune disease that ultimately destroys the joints. T cells, macrophages, and synovial fibroblasts in the joints play a

pivotal role in rheumatoid arthritis disease progression (Huber et al. 2006). Single-cell biochemistry assays show heterogenous phosphatase and protease activity in fibroblasts from rheumatoid arthritis patients that may help explain why a third of patients are unresponsive to some rheumatoid arthritis treatments (Mainz et al. 2016). Additionally single-cell analysis of primary samples from patients with lung fibrosis indicate that there are heterogenous populations of cells that may contribute to the disease and may help dictate treatments for lung fibrosis (Reyfman et al. 2019). Our single-cell assay may be useful in determining successful treatments in these populations as well.

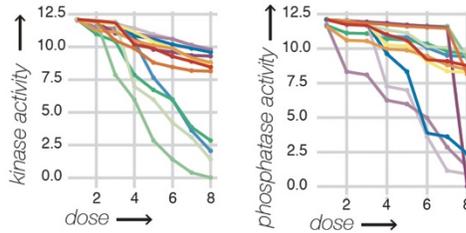
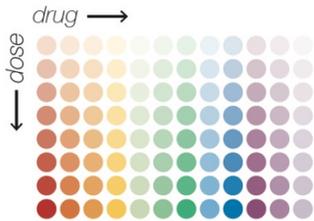
Currently, single-cell methods have great potential in pre-clinical studies to understand the heterogeneity of complex diseases and their treatments. However, single-cell methods are in a relatively early stage of optimization and development. If single-cell methods are to transition into a clinical setting, there are several considerations that must be met (Shalek and Benson 2017). Namely, the process of collecting, processing, evaluating, and analyzing data from clinical samples must be standardized into a reproducible workflow (Shalek and Benson 2017). Arguably, using biochemical data is simpler than distilling gene expression networks into targetable phenotypes, which may simplify the application of single-cell biochemistry in a clinical setting. Before single-cell methods can be standardized in a clinical setting, there needs to be a well characterized database of normal tissues and cell types. Single-cell methods continue to uncover new cell types (Macosko et al. 2015), indicating that we have more still to uncover about the natural variability in human tissues. Biochemical methods will only add to our understanding of normal and diseased cell types to contribute to the field of personalized medicine.

a Precision medicine

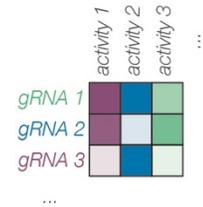


b Pharmaceutical sciences

drug screens

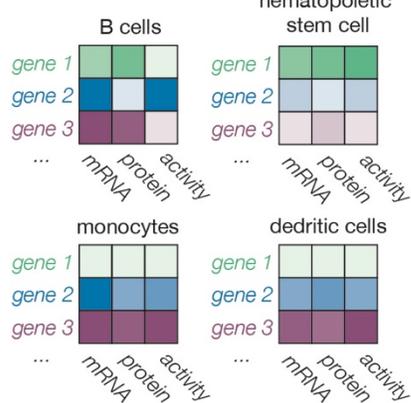


CRISPR screens

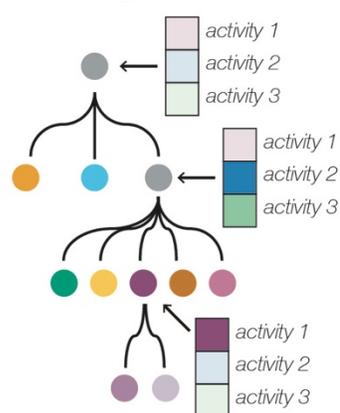


c Developmental biology

cell atlas



cell lineages



cell states

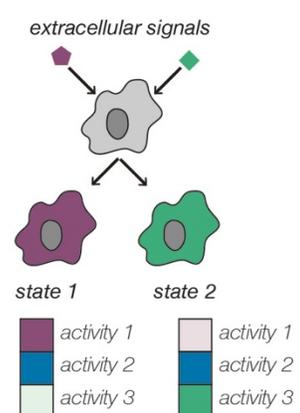


Figure 3.5: Single-cell biochemistry has the potential to contribute to the fields of precision medicine, pharmaceutical sciences, and developmental biology

a. Hypothetical data of how single-cell biochemistry could be used to predict or understand tumor response to chemotherapies. Pre-treatment phenotypes for cancer cells could show some tumor cells have high DNA repair capacity (left). Immediately post-treatment measurements could indicate some cells rapidly undergo apoptosis, measured by caspase-3 activity, and some cells increase DNA repair capacity in response to treatment (middle). Several days following treatment, some tumor cells persist and have high DNA repair capacities (right). It may be possible to use the pre-treatment measurements to dictate or predict whether cells will persist during treatment.

b. Single-cell biochemistry could be used in combination with sample barcoding single-cell methods. Cells could be treated with a panel of drugs across a number of concentrations and time points (left). Each treatment could be barcoded separately, and the response could be measured by single-cell biochemistry. Since the samples are barcoded, cell responses could be measured for each treatment simultaneously and drugs could be screened by enzyme activity for efficacy (middle). Single-cell biochemistry could also be used as a readout for CRISPR knockout screens. A pooled CRISPR knockout library could be transfected into cells and through single-cell methods the CRISPR gRNA and enzyme activities could be measured to identify regulators enzymatic activities (right).

c. Single-cell biochemistry could contribute to the field of developmental biology by adding to cell atlas databases (left), measuring enzymatic activity at key lineage decision points (middle), or by characterizing cell states based on extracellular signals (right).

Single-cell biochemistry in pharmaceutical sciences

Developing compounds that effectively target cellular proteins often requires high throughput readouts for measuring compound effectiveness. As single-cell technology improves, our ability to develop high throughput screens also increases. Sample barcoding methods currently enable us to pool hundreds of samples in a single experiment (Shin et al. 2019; McGinnis et al. 2019; Stoeckius et al. 2018; Srivatsan et al. 2020). These treat, barcode, and pool methods have already been used to screen compounds and evaluate their mechanisms and efficacy through gene expression changes (Shin et al. 2019; Srivatsan et al. 2020). One study even found that some drugs produced a heterogeneous response, while other drugs with similar targets did not (Srivatsan et al. 2020), indicating single-cell analysis provides a high-throughput and detailed readout for drug screening. One drawback to these methods is the sequencing cost (summarized in detail above). While some methods can prepare single-cell libraries at less than a penny per cell, the sequencing costs that enable thorough characterization of compounds is still quite high and distilling global gene expression into a response phenotype is computationally intensive.

Our single-cell biochemistry assays can simplify drug screening by directly measuring enzymatic activity. There are two main ways that single-cell biochemistry can improve drug development and discovery (**Figure 3.5b**).

First, drug screens can be done in a direct, high throughput, and inexpensive experiment. If testing a panel of kinase inhibitors, samples can be barcoded using one of several single-cell sample barcoding methods (Shin et al. 2019; Stoeckius et al. 2018; McGinnis et al. 2019; Srivatsan et al. 2020) and treated with a panel of inhibitors spanning different time points and concentrations. Then, we can measure kinase activities independently or simultaneously with gene expression in single cells. Individually barcoding each drug, dose, and time point enables us to demultiplex the

data to assess drug activities using single-cell kinase activity measurements. Drug toxicity can also be measured by caspase-3 activity. Drug specificity can also be measured, given that we have a panel of related kinase substrates that correspond to specific kinase activities. If combined with gene expression data, alterations in enzyme functions can be correlated with subsequent changes in gene expression. All together, these data would represent a thorough characterization of drug efficacy and specificity.

Second, we can discover novel regulators of cellular processes using CRISPR screens and single-cell biochemistry measurements. There are several single-cell CRISPR screening methods that currently exist, however, the sequencing cost is often prohibitively high for widespread applicability (Adamson et al. 2016; Jaitin et al. 2016; Datlinger et al. 2017; Mimitou et al. 2019; Replogle et al. 2020). To identify novel functional regulators of DNA, protein kinases, protein phosphatases, and protease, we can transfect cells with a genome-wide CRISPR library and measure enzyme function at single-cell resolution. If we design the CRISPR guides to include a unique capture sequence (Mimitou et al. 2019), then we can capture them alongside our substrates. When known regulators of enzyme activities are knocked out, then the activity will decrease or disappear. When unknown or indirect regulators are knocked out, we are likely to measure an intermediate effect. For example, if the DNA glycosylase UNG is knocked out, we expect very little enzymatic activity on the uracil containing DNA repair substrate. However, if the scaffolding DNA repair protein XRCC1 is knocked out, we may see a moderate decrease in many DNA glycosylase activities. Since we are measuring enzymatic activities, we would have direct functional measurements in knockout cells at a much lower sequencing depth. This kind of single-cell CRISPR screen can also be applied to identify synthetic lethal combinations or combinations

that significantly impact overlapping pathways. Overall, a genome-wide screen for regulators of enzymatic activities could act as a blueprint choosing drug targets.

Single-cell biochemistry in developmental biology

The field of developmental biology has benefited from the growing applications of single-cell biology. Developing organisms are a prime example of rapidly changing heterogeneous cell populations. Cells progress along developmental trajectories as they differentiate into specialized cell types that make up diverse tissues. These developmental systems can be monitored at single-cell resolution throughout developmental timepoints to understand how cell fates are determined. In addition to discovering key regulators during the developmental process, the single-cell field has also been working towards developing a reference single-cell atlas that describes all the cells in the body. In a recent review, Dr. Samantha Morris argued that there are three main features to a complete cell atlas (Morris 2019). First, we need a catalogue of cell phenotypes and functions. Second, we need a thorough lineage map for each cell type. Third, and finally, we need an understanding of cell states. Single-cell biochemistry methods can be applied to each of these features in order to create a complete cell atlas (**Figure 3.5c**).

Using single-cell biochemistry to define cell types and functions

Single cell methods are very useful in finding new cell types, especially rare subtypes with unknown markers that are easily lost in bulk and cell sorting experiments (S. Zheng et al. 2018). New cell types are characterized by their gene expression profiles that can leave many unanswered questions about cell phenotype and function. Using an expanded single-cell biochemistry toolkit, we can add to the characterization of novel cell types within the context of existing methods. Our unique position to simultaneously measure enzyme activity and gene expression may rapidly expand our characterization of cell types throughout the body, especially when combined with

bulk datasets, like the human protein atlas (Uhlén et al. 2015; Thul et al. 2017). Additionally, tissue-specific protein functions can lead to tissue-specific regulatory networks (Zitnik and Leskovec 2017). Adding single-cell biochemical data to an existing cell atlas would add direct, functional data to characterize abundant and rare cell types.

Using single-cell biology to uncover the determinant factors of cell trajectories

Single-cell methods have significantly improved our understanding of developmental lineages and trajectories. Many single-cell methods use computational strategies to track and predict developmental trajectories (Qiu, Mao, et al. 2017). There are also methods to directly track developmental lineages using polyadenylated heritable cell barcodes (Guo et al. 2019). Single-cell gene expression measurements of developing organs have been used to develop prediction models for regulatory networks that determine cell fates (Moignard et al. 2015; Lescroart et al. 2018). While these novel methods uncovered transition states and regulatory networks, understanding the functional changes throughout the process is still a challenge. Adding biochemical information to these types of analyses may aid our ability to make fate predictions or help us understand the major functional players of regulatory networks that govern the developmental process (**Figure 3.5c**, middle).

Specifically, our single-cell biochemistry assays can provide insight into key enzyme activities that drive developmental processes. For example, the major endoderm differentiation marker, KLF8, regulates the expression of some MMPs (Chu et al. 2016; X. Wang et al. 2011). If we measure the activity of MMPs in single cells during endoderm differentiation, we may learn what role MMPs play in differentiation and whether specific cells drive the process. Another example is in B cell development where there are rapid changes in the proteome (Bendall et al. 2014). By measuring E3 ubiquitin ligase activities and cellular protease activities, we can learn

about the regulatory processes that enable cells to rapidly change their proteomes. These are just two specific examples of how single-cell biochemical assays can add to our understanding of developmental processes.

Using single-cell biochemistry to define and understand cell states

Cells are dynamic and responsive to environmental cues that can lead to heterogeneous gene expression profiles even within a cell type. To fully achieve a single-cell atlas, cell states must be included in the picture. For example, the cell cycle represents four unique states of dividing cells. In fact, there are several single-cell methods that remove cell cycle variability from single-cell analysis (Butler et al. 2018a; Qiu, Hill, et al. 2017). While the cell cycle is very well defined in gene expression and function, other cell states are less defined. For example, T cells can have many different states depending on their location in the body, what antigens they are exposed to, and from where they originate. Several single-cell studies have characterized T cell states in healthy tissues and in cancers (Zemmour et al. 2018; van der Leun, Thommen, and Schumacher 2020; Sade-Feldman et al. 2018). In regulatory T cells, T cell receptor (TCR) activation can alter or determine a T cell state (Zemmour et al. 2018). The TCR signal transduction pathway is complex and has many overlapping signaling molecules. If we can comprehensively measure protein kinase and phosphatase activities, we may be able to determine how TCR signaling can affect T cell states and heterogeneity in T cell activation during an immune response. Given we can develop a panel of signal transduction substrates, our single-cell biochemistry toolkit could be deployed to effectively measure cell states based on environmental exposures.

One consideration in characterizing cell states is variability in the enzymatic activities. I measure variability in enzymatic activities and gene expression in single cells. The amplitude of these variabilities (i.e. range) in normalized data (see **Chapter II**) is similar between mRNA

expression and repair activities. However, some cell types, like B cells, seem to have a smaller range of repair activities compare to other cell types, like T cells (**Figure 2.8**). These differences could be due to cell states or the presence of subpopulations of related cell types. Some evidence indicates that stochastic gene expression leads to heterogenous responses to DNA damage (Uphoff et al. 2016; Uphoff 2018). Single-cell measurements may uncover that healthy cells have a normal variability range that may change upon the onset of disease. To truly understand the biological significance of the variability in the data, we need to have first achieve a thorough understanding of gene expression and enzymatic activities for each cell type in each cell state.

If everything works, what could we learn about biology?

Imagine a scientific community where the term biomarkers is a thing of the past. Currently, we rely on gene expression or DNA sequence measurements to predict cell function. We use biomarkers to predict disease prognosis. If single-cell or multiplexed biochemistry methods became high-throughput, specific, sensitive, and available at a low cost, then biomarkers and gene expression may become irrelevant for a variety of fields. Rather than look at gene expression to predict cell function, a panel of substrates could be used to directly characterize cell function. Drug responses in cells could be measured using a panel of biochemical readouts. Developmental biologists could track kinase and E3 ubiquitin ligase activity throughout development. Medical providers and cancer biologists could directly measure the effect of small-molecule inhibitors in tumor biopsies and assess their efficacy at single-cell resolution with direct enzymatic data. Variability and stochasticity of cell responses could be thoroughly catalogued at unprecedented resolution. Gene expression and genome-wide measurements will still advance scientific discoveries, but they may no longer have to be distilled into biomarkers and phenotypic predictions.

REFERENCES

- Aamodt, Randi M., Pål Ø. Falnes, Rune F. Johansen, Erling Seeberg, and Magnar Bjørås. 2004. "The *Bacillus Subtilis* Counterpart of the Mammalian 3-Methyladenine DNA Glycosylase Has Hypoxanthine and 1,N6-Ethenoadenine as Preferred Substrates." *The Journal of Biological Chemistry* 279 (14): 13601–6.
- Adamson, Britt, Thomas M. Norman, Marco Jost, Min Y. Cho, James K. Nuñez, Yuwen Chen, Jacqueline E. Villalta, et al. 2016. "A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response." *Cell* 167 (7): 1867–1882.e21.
- Aden, Konrad, Kareen Bartsch, Joseph Dahl, Martin A. M. Reijns, Daniela Esser, Raheleh Sheibani-Tezerji, Anupam Sinha, et al. 2019. "Epithelial RNase H2 Maintains Genome Integrity and Prevents Intestinal Tumorigenesis in Mice." *Gastroenterology* 156 (1): 145–159.e19.
- Agnihotri, Sameer, Kelly Burrell, Pawel Buczkowicz, Marc Remke, Brian Golbourn, Yevgen Chornenkyy, Aaron Gajadhar, et al. 2014. "ATM Regulates 3-Methylpurine-DNA Glycosylase and Promotes Therapeutic Resistance to Alkylating Agents." *Cancer Discovery* 4 (10): 1198–1213.
- Aguilera, Andrés, and Tatiana García-Muse. 2012. "R Loops: From Transcription Byproducts to Threats to Genome Stability." *Molecular Cell* 46 (2): 115–24.
- Akbari, M., M. Otterlei, J. Peña-Diaz, and H. E. Krokan. 2007. "Different Organization of Base Excision Repair of Uracil in DNA in Nuclei and Mitochondria and Selective Upregulation of Mitochondrial Uracil-DNA Glycosylase after Oxidative Stress." *Neuroscience* 145 (4): 1201–12.
- Akbari, Mansour, Javier Peña-Diaz, Sonja Andersen, Nina-Beate Liabakk, Marit Otterlei, and Hans Einar Krokan. 2009. "Extracts of Proliferating and Non-Proliferating Human Cells Display Different Base Excision Pathways and Repair Fidelity." *DNA Repair* 8 (7): 834–43.
- Alexa, Adrian, and Jörg Rahnenführer. 2009. "Gene Set Enrichment Analysis with TopGO." *Bioconductor* *Improv* 27. <https://bioconductor.org/packages/release/bioc/vignettes/topGO/inst/doc/topGO.pdf>.
- Allocca, Mariacarmela, Joshua J. Corrigan, Kimberly R. Fake, Jennifer A. Calvo, and Leona D. Samson. 2017. "PARP Inhibitors Protect against Sex- and AAG-Dependent Alkylation-Induced Neural Degeneration." *Oncotarget* 8 (40): 68707–20.
- Allocca, Mariacarmela, Joshua J. Corrigan, Aprotim Mazumder, Kimberly R. Fake, and Leona D. Samson. 2019. "Inflammation, Necrosis, and the Kinase RIP3 Are Key Mediators of AAG-Dependent Alkylation-Induced Retinal Degeneration." *Science Signaling* 12 (568). <https://doi.org/10.1126/scisignal.aau9216>.
- Al-Tassan, Nada, Nikolas H. Chmiel, Julie Maynard, Nick Fleming, Alison L. Livingston, Geraint T. Williams, Angela K. Hodges, et al. 2002. "Inherited Variants of MYH Associated with Somatic G:C-->T:A Mutations in Colorectal Tumors." *Nature Genetics* 30 (2): 227–32.
- An, Qian, Peter Robins, Tomas Lindahl, and Deborah E. Barnes. 2005. "C --> T Mutagenesis and Gamma-Radiation Sensitivity Due to Deficiency in the Smug1 and Ung DNA Glycosylases." *The EMBO Journal* 24 (12): 2205–13.
- Andersen, Sonja, Tina Heine, Ragnhild Sneve, Imbritt König, Hans E. Krokan, Bernd Epe, and Hilde Nilsen. 2005. "Incorporation of DUMP into DNA Is a Major Source of Spontaneous

- DNA Damage, While Excision of Uracil Is Not Required for Cytotoxicity of Fluoropyrimidines in Mouse Embryonic Fibroblasts.” *Carcinogenesis* 26 (3): 547–55.
- Andersson, M., P. Stenqvist, and B. Hellman. 2007. “Interindividual Differences in Initial DNA Repair Capacity When Evaluating H₂O₂-Induced DNA Damage in Extended-Term Cultures of Human Lymphocytes Using the Comet Assay.” *Cell Biology and Toxicology* 23 (6): 401–11.
- Antoniali, Giulia, Matilde Clarissa Malfatti, and Gianluca Tell. 2017. “Unveiling the Non-Repair Face of the Base Excision Repair Pathway in RNA Processing: A Missing Link between DNA Repair and Gene Expression?” *DNA Repair* 56 (August): 65–74.
- Araya, Paula, Katherine A. Waugh, Kelly D. Sullivan, Nicolás G. Núñez, Emiliano Roselli, Keith P. Smith, Ross E. Granrath, et al. 2019. “Trisomy 21 Dysregulates T Cell Lineages toward an Autoimmunity-Prone State Associated with Interferon Hyperactivity.” *Proceedings of the National Academy of Sciences of the United States of America* 116 (48): 24231–41.
- Arudchandran, A., S. Cerritelli, S. Narimatsu, M. Itaya, D. Y. Shin, Y. Shimada, and R. J. Crouch. 2000. “The Absence of Ribonuclease H1 or H2 Alters the Sensitivity of *Saccharomyces Cerevisiae* to Hydroxyurea, Caffeine and Ethyl Methanesulphonate: Implications for Roles of RNases H in DNA Replication and Repair.” *Genes to Cells: Devoted to Molecular & Cellular Mechanisms* 5 (10): 789–802.
- Azzam, Edouard I., Jean-Paul Jay-Gerin, and Debkumar Pain. 2012. “Ionizing Radiation-Induced Metabolic Oxidative Stress and Prolonged Cell Injury.” *Cancer Letters* 327 (1–2): 48–60.
- Ba, Xueqing, and Istvan Boldogh. 2018. “8-Oxoguanine DNA Glycosylase 1: Beyond Repair of the Oxidatively Modified Base Lesions.” *Redox Biology* 14 (April): 669–78.
- Bandaru, Viswanath, Sirisha Sunkara, Susan S. Wallace, and Jeffrey P. Bond. 2002. “A Novel Human DNA Glycosylase That Removes Oxidative DNA Damage and Is Homologous to *Escherichia Coli* Endonuclease VIII.” *DNA Repair* 1 (7): 517–29.
- Banerjee, Anirban, Wei Yang, Martin Karplus, and Gregory L. Verdine. 2005. “Structure of a Repair Enzyme Interrogating Undamaged DNA Elucidates Recognition of Damaged DNA.” *Nature* 434 (7033): 612–18.
- Banerjee, Dibyendu, Santi M. Mandal, Aditi Das, Muralidhar L. Hegde, Soumita Das, Kishor K. Bhakat, Istvan Boldogh, Partha S. Sarkar, Sankar Mitra, and Tapas K. Hazra. 2011. “Preferential Repair of Oxidized Base Damage in the Transcribed Genes of Mammalian Cells.” *The Journal of Biological Chemistry* 286 (8): 6006–16.
- Barnes, Deborah E., and Tomas Lindahl. 2004. “Repair and Genetic Consequences of Endogenous DNA Base Damage in Mammalian Cells.” *Annual Review of Genetics* 38: 445–76.
- Barnes, Tavish, Wan-Cheol Kim, Anil K. Mantha, Sang-Eun Kim, Tadahide Izumi, Sankar Mitra, and Chow H. Lee. 2009. “Identification of Apurinic/Apyrimidinic Endonuclease 1 (APE1) as the Endoribonuclease That Cleaves c-Myc mRNA.” *Nucleic Acids Research* 37 (12): 3946–58.
- Barnett, G. C., S. L. Kerns, D. J. Noble, A. M. Dunning, C. M. L. West, and N. G. Burnet. 2015. “Incorporating Genetic Biomarkers into Predictive Models of Normal Tissue Toxicity.” *Clinical Oncology* 27 (10): 579–87.
- Basu, A. K., E. L. Loechler, S. A. Leadon, and J. M. Essigmann. 1989. “Genetic Effects of Thymine Glycol: Site-Specific Mutagenesis and Molecular Modeling Studies.” *Proceedings of the National Academy of Sciences of the United States of America* 86 (20): 7677–81.

- Basu, A. K., M. L. Wood, L. J. Niedernhofer, L. A. Ramos, and J. M. Essigmann. 1993. "Mutagenic and Genotoxic Effects of Three Vinyl Chloride-Induced DNA Lesions: 1,N6-Ethenoadenine, 3,N4-Ethenocytosine, and 4-Amino-5-(Imidazol-2-Yl)Imidazole." *Biochemistry* 32 (47): 12793–801.
- Bauer, Nicholas C., Anita H. Corbett, and Paul W. Doetsch. 2015a. "The Current State of Eukaryotic DNA Base Damage and Repair." *Nucleic Acids Research* 43 (21): 10083–101.
- . 2015b. "The Current State of Eukaryotic DNA Base Damage and Repair." *Nucleic Acids Research* 43 (21): 10083–101.
- Becker, David, and Michael D. Sevilla. 1993. "3 - The Chemical Consequences of Radiation Damage to DNA." In *Advances in Radiation Biology*, edited by John T. Lett and Warren K. Sinclair, 17:121–80. Elsevier.
- Bellacosa, A., L. Cicchillitti, F. Schepis, A. Riccio, A. T. Yeung, Y. Matsumoto, E. A. Golemis, M. Genuardi, and G. Neri. 1999. "MED1, a Novel Human Methyl-CpG-Binding Endonuclease, Interacts with DNA Mismatch Repair Protein MLH1." *Proceedings of the National Academy of Sciences of the United States of America* 96 (7): 3969–74.
- Bendall, Sean C., Kara L. Davis, El-Ad David Amir, Michelle D. Tadmor, Erin F. Simonds, Tiffany J. Chen, Daniel K. Shenfeld, Garry P. Nolan, and Dana Pe'er. 2014. "Single-Cell Trajectory Detection Uncovers Progression and Regulatory Coordination in Human B Cell Development." *Cell* 157 (3): 714–25.
- Bennett, Matthew T., M. T. Rodgers, Alexander S. Hebert, Lindsay E. Ruslander, Leslie Eisele, and Alexander C. Drohat. 2006. "Specificity of Human Thymine DNA Glycosylase Depends on N-Glycosidic Bond Stability." *Journal of the American Chemical Society* 128 (38): 12510–19.
- Beranek, D. T. 1990. "Distribution of Methyl and Ethyl Adducts Following Alkylation with Monofunctional Alkylating Agents." *Mutation Research* 231 (1): 11–30.
- Berger, Sondra H., Douglas L. Pittman, and Michael D. Wyatt. 2008. "Uracil in DNA: Consequences for Carcinogenesis and Chemotherapy." *Biochemical Pharmacology* 76 (6): 697–706.
- Bessho, T. 1999. "Nucleotide Excision Repair 3' Endonuclease XPG Stimulates the Activity of Base Excision Repair Enzyme Thymine Glycol DNA Glycosylase." *Nucleic Acids Research* 27 (4): 979–83.
- Bhakat, Kishor K., Sanath K. Mokkapatil, Istvan Boldogh, Tapas K. Hazra, and Sankar Mitra. 2006. "Acetylation of Human 8-Oxoguanine-DNA Glycosylase by P300 and Its Role in 8-Oxoguanine Repair in Vivo." *Molecular and Cellular Biology* 26 (5): 1654–65.
- Birben, Esra, Umit Murat Sahiner, Cansin Sackesen, Serpil Erzurum, and Omer Kalayci. 2012. "Oxidative Stress and Antioxidant Defense." *The World Allergy Organization Journal* 5 (1): 9–19.
- Blainey, Paul C., Antoine M. van Oijen, Anirban Banerjee, Gregory L. Verdine, and X. Sunney Xie. 2006. "A Base-Excision DNA-Repair Protein Finds Intrahelical Lesion Bases by Fast Sliding in Contact with DNA." *Proceedings of the National Academy of Sciences of the United States of America* 103 (15): 5752–57.
- Blair, Ian A. 2008. "DNA Adducts with Lipid Peroxidation Products." *The Journal of Biological Chemistry* 283 (23): 15545–49.
- Blount, B. C., M. M. Mack, C. M. Wehr, J. T. MacGregor, R. A. Hiatt, G. Wang, S. N. Wickramasinghe, R. B. Everson, and B. N. Ames. 1997. "Folate Deficiency Causes Uracil Misincorporation into Human DNA and Chromosome Breakage: Implications for Cancer

- and Neuronal Damage.” *Proceedings of the National Academy of Sciences of the United States of America* 94 (7): 3290–95.
- Branzei, Dana, and Marco Foiani. 2008. “Regulation of DNA Repair throughout the Cell Cycle.” *Nature Reviews. Molecular Cell Biology* 9 (4): 297–308.
- Bravard, Anne, Anna Campalans, Monique Vacher, Barbara Gouget, Céline Levalois, Sylvie Chevillard, and J. Pablo Radicella. 2010. “Inactivation by Oxidation and Recruitment into Stress Granules of HOGG1 but Not APE1 in Human Cells Exposed to Sub-Lethal Concentrations of Cadmium.” *Mutation Research* 685 (1–2): 61–69.
- Bravard, Anne, Monique Vacher, Barbara Gouget, Alexandre Coutant, Florence Hillairet de Boisferon, Stéphanie Marsin, Sylvie Chevillard, and J. Pablo Radicella. 2006. “Redox Regulation of Human OGG1 Activity in Response to Cellular Oxidative Stress.” *Molecular and Cellular Biology* 26 (20): 7430–36.
- Briegert, Manuela, and Bernd Kaina. 2007. “Human Monocytes, but Not Dendritic Cells Derived from Them, Are Defective in Base Excision Repair and Hypersensitive to Methylating Agents.” *Cancer Research* 67 (1): 26–31.
- Brooks, Philip J. 2017. “The Cyclopurine Deoxynucleosides: DNA Repair, Biological Effects, Mechanistic Insights, and Unanswered Questions.” *Free Radical Biology & Medicine* 107 (June): 90–100.
- Brooks, Philip J., Dean S. Wise, David A. Berry, Joseph V. Kosmoski, Michael J. Smerdon, Robert L. Somers, Hugh Mackie, et al. 2000. “The Oxidative DNA Lesion 8,5'-(S)-Cyclo-2'-Deoxyadenosine Is Repaired by the Nucleotide Excision Repair Pathway and Blocks Gene Expression in Mammalian Cells.” *The Journal of Biological Chemistry* 275 (29): 22355–62.
- Brown, Jessica A., and Zucui Suo. 2011. “Unlocking the Sugar ‘Steric Gate’ of DNA Polymerases.” *Biochemistry* 50 (7): 1135–42.
- Bryan, D. Suzi, Monica Ransom, Biniam Adane, Kerri York, and Jay R. Hesselberth. 2014. “High Resolution Mapping of Modified DNA Nucleobases Using Excision Repair Enzymes.” *Genome Research* 24 (9): 1534–42.
- Bubeck, Doryen, Martin A. M. Reijns, Stephen C. Graham, Katy R. Astell, E. Yvonne Jones, and Andrew P. Jackson. 2011. “PCNA Directs Type 2 RNase H Activity on DNA Replication and Repair Substrates.” *Nucleic Acids Research* 39 (9): 3652–66.
- Butler, Andrew, Paul Hoffman, Peter Smibert, Efthymia Papalex, and Rahul Satija. 2018a. “Integrating Single-Cell Transcriptomic Data across Different Conditions, Technologies, and Species.” *Nature Biotechnology* 36 (5): 411–20.
- . 2018b. “Integrating Single-Cell Transcriptomic Data across Different Conditions, Technologies, and Species.” *Nature Biotechnology* 36 (5): 411–20.
- Cabrini, Giulio, Enrica Fabbri, Cristiana Lo Nigro, Maria Cristina Dehecchi, and Roberto Gambari. 2015. “Regulation of Expression of O6-Methylguanine-DNA Methyltransferase and the Treatment of Glioblastoma (Review).” *International Journal of Oncology* 47 (2): 417–28.
- Cadet, Jean, Thierry Douki, and Jean-Luc Ravanat. 2008. “Oxidatively Generated Damage to the Guanine Moiety of DNA: Mechanistic Aspects and Formation in Cells.” *Accounts of Chemical Research* 41 (8): 1075–83.
- . 2015. “Oxidatively Generated Damage to Cellular DNA by UVB and UVA Radiation.” *Photochemistry and Photobiology* 91 (1): 140–55.

- Cai, Shanbao, Yi Xu, Ryan J. Cooper, Michael J. Ferkowicz, Jennifer R. Hartwell, Karen E. Pollok, and Mark R. Kelley. 2005. "Mitochondrial Targeting of Human O6-Methylguanine DNA Methyltransferase Protects against Cell Killing by Chemotherapeutic Alkylating Agents." *Cancer Research* 65 (8): 3319–27.
- Campalans, Anna, Stéphanie Marsin, Yusaku Nakabeppu, Timothy R. O'connor, Serge Boiteux, and J. Pablo Radicella. 2005. "XRCC1 Interactions with Multiple DNA Glycosylases: A Model for Its Recruitment to Base Excision Repair." *DNA Repair* 4 (7): 826–35.
- Candéias, S., B. Pons, M. Viau, S. Caillat, and S. Sauvaigo. 2010. "Direct Inhibition of Excision/Synthesis DNA Repair Activities by Cadmium: Analysis on Dedicated Biochips." *Mutation Research* 694 (1–2): 53–59.
- Canman, C. E., E. H. Radany, L. A. Parsels, M. A. Davis, T. S. Lawrence, and J. Maybaum. 1994. "Induction of Resistance to Fluorodeoxyuridine Cytotoxicity and DNA Damage in Human Tumor Cells by Expression of Escherichia Coli Deoxyuridinetriphosphatase." *Cancer Research* 54 (9): 2296–98.
- Cao, Chunyang, Yu Lin Jiang, James T. Stivers, and Fenhong Song. 2004. "Dynamic Opening of DNA during the Enzymatic Search for a Damaged Base." *Nature Structural & Molecular Biology* 11 (12): 1230–36.
- Cao, Jianli, and Arnold Revzin. 1993. "7 - Interference and Missing Contact Footprinting." In *Footprinting of Nucleic Acid-Protein Complexes*, edited by Arnold Revzin, 173–88. Boston: Academic Press.
- Cao, Junyue, Jonathan S. Packer, Vijay Ramani, Darren A. Cusanovich, Chau Huynh, Riza Daza, Xiaojie Qiu, et al. 2017. "Comprehensive Single-Cell Transcriptional Profiling of a Multicellular Organism." *Science* 357 (6352): 661–67.
- Cao, Junyue, Malte Spielmann, Xiaojie Qiu, Xingfan Huang, Daniel M. Ibrahim, Andrew J. Hill, Fan Zhang, et al. 2019. "The Single-Cell Transcriptional Landscape of Mammalian Organogenesis." *Nature* 566 (7745): 496–502.
- Caple, Fiona, Elizabeth A. Williams, Alison Spiers, John Tyson, Brian Burtle, Ann K. Daly, John C. Mathers, and John E. Hesketh. 2010. "Inter-Individual Variation in DNA Damage and Base Excision Repair in Young, Healthy Non-Smokers: Effects of Dietary Supplementation and Genotype." *The British Journal of Nutrition* 103 (11): 1585–93.
- Castellano-Pozo, Maikel, Tatiana García-Muse, and Andrés Aguilera. 2012. "R-Loops Cause Replication Impairment and Genome Instability during Meiosis." *EMBO Reports* 13 (10): 923–29.
- Cataldo, Anne M., Suzana Petanceska, Corrinne M. Peterhoff, Nicole B. Terio, Charles J. Epstein, Angela Villar, Elaine J. Carlson, Matthias Staufenbiel, and Ralph A. Nixon. 2003. "App Gene Dosage Modulates Endosomal Abnormalities of Alzheimer's Disease in a Segmental Trisomy 16 Mouse Model of down Syndrome." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 23 (17): 6788–92.
- Cerritelli, Susana M., and Robert J. Crouch. 2009. "Ribonuclease H: The Enzymes in Eukaryotes." *The FEBS Journal* 276 (6): 1494–1505.
- . 2016. "The Balancing Act of Ribonucleotides in DNA." *Trends in Biochemical Sciences* 41 (5): 434–45.
- Cerritelli, Susana M., Ella G. Frolova, Chiguang Feng, Alexander Grinberg, Paul E. Love, and Robert J. Crouch. 2003. "Failure to Produce Mitochondrial DNA Results in Embryonic Lethality in Rnaseh1 Null Mice." *Molecular Cell* 11 (3): 807–15.

- Chaim, Isaac A., Zachary D. Nagel, Jennifer J. Jordan, Patrizia Mazzucato, Le P. Ngo, and Leona D. Samson. 2017a. "In Vivo Measurements of Interindividual Differences in DNA Glycosylases and APE1 Activities." *Proceedings of the National Academy of Sciences of the United States of America* 114 (48): E10379–88.
- . 2017b. "In Vivo Measurements of Interindividual Differences in DNA Glycosylases and APE1 Activities." *Proceedings of the National Academy of Sciences of the United States of America* 114 (48): E10379–88.
- . 2017c. "In Vivo Measurements of Interindividual Differences in DNA Glycosylases and APE1 Activities." *Proceedings of the National Academy of Sciences of the United States of America* 114 (48): E10379–88.
- Chance, B., H. Sies, and A. Boveris. 1979. "Hydroperoxide Metabolism in Mammalian Organs." *Physiological Reviews* 59 (3): 527–605.
- Chang, Jyh-Lurn, Gang Chen, Cornelia M. Ulrich, Jeannette Bigler, Irena B. King, Yvonne Schwarz, Shiuying Li, Lin Li, John D. Potter, and Johanna W. Lampe. 2010. "DNA Damage and Repair: Fruit and Vegetable Effects in a Feeding Trial." *Nutrition and Cancer* 62 (3): 329–35.
- Chen, S. Y., L. Wan, C. M. Huang, Y. C. Huang, J. J. C. Sheu, Y. J. Lin, S. P. Liu, et al. 2010. "Genetic Polymorphisms of the DNA Repair Gene MPG May Be Associated with Susceptibility to Rheumatoid Arthritis." *Journal of Applied Genetics* 51 (4): 519–21.
- Chen, Xiaoyue, Minjie Zhang, Haiyun Gan, Heping Wang, Jeong-Heon Lee, Dong Fang, Gaspar J. Kitange, et al. 2018. "A Novel Enhancer Regulates MGMT Expression and Promotes Temozolomide Resistance in Glioblastoma." *Nature Communications* 9 (1): 2949.
- Chen, Yen-Ling, Staffan Eriksson, and Zee-Fen Chang. 2010. "Regulation and Functional Contribution of Thymidine Kinase 1 in Repair of DNA Damage." *The Journal of Biological Chemistry* 285 (35): 27327–35.
- Chiricolo, M., A. R. Musa, D. Monti, M. Zannotti, and C. Franceschi. 1993. "Enhanced DNA Repair in Lymphocytes of Down Syndrome Patients: The Influence of Zinc Nutritional Supplementation." *Mutation Research* 295 (3): 105–11.
- Chon, Hyongi, Alex Vassilev, Melvin L. DePamphilis, Yingming Zhao, Junmei Zhang, Peter M. Burgers, Robert J. Crouch, and Susana M. Cerritelli. 2009. "Contributions of the Two Accessory Subunits, RNASEH2B and RNASEH2C, to the Activity and Properties of the Human RNase H2 Complex." *Nucleic Acids Research* 37 (1): 96–110.
- Chon, James, Martha S. Field, and Patrick J. Stover. 2019. "Deoxyuracil in DNA and Disease: Genomic Signal or Managed Situation?" *DNA Repair* 77 (May): 36–44.
- Chou, Kai-Ming, and Yung-Chi Cheng. 2002. "An Exonucleolytic Activity of Human Apurinic/Apyrimidinic Endonuclease on 3' Mispaiored DNA." *Nature* 415 (6872): 655–59.
- . 2003. "The Exonuclease Activity of Human Apurinic/Apyrimidinic Endonuclease (APE1). Biochemical Properties and Inhibition by the Natural Dinucleotide Gp4G." *The Journal of Biological Chemistry* 278 (20): 18289–96.
- Christmann, Markus, Georg Nagel, Sigrid Horn, Ulrike Krahn, Dorothee Wiewrodt, Clemens Sommer, and Bernd Kaina. 2010. "MGMT Activity, Promoter Methylation and Immunohistochemistry of Pretreatment and Recurrent Malignant Gliomas: A Comparative Study on Astrocytoma and Glioblastoma." *International Journal of Cancer. Journal International Du Cancer* 127 (9): 2106–18.
- Chu, Li-Fang, Ning Leng, Jue Zhang, Zhonggang Hou, Daniel Mamott, David T. Vereide, Jee Choi, Christina Kendziorski, Ron Stewart, and James A. Thomson. 2016. "Single-Cell

- RNA-Seq Reveals Novel Regulators of Human Embryonic Stem Cell Differentiation to Definitive Endoderm.” *Genome Biology* 17 (1): 173.
- Church, D. F., and W. A. Pryor. 1985. “Free-Radical Chemistry of Cigarette Smoke and Its Toxicological Implications.” *Environmental Health Perspectives* 64 (December): 111–26.
- Clausen, Anders R., Scott A. Lujan, Adam B. Burkholder, Clinton D. Orebaugh, Jessica S. Williams, Maryam F. Clausen, Ewa P. Malc, et al. 2015. “Tracking Replication Enzymology in Vivo by Genome-Wide Mapping of Ribonucleotide Incorporation.” *Nature Structural & Molecular Biology* 22 (3): 185–91.
- Clausen, Anders R., Sufang Zhang, Peter M. Burgers, Marietta Y. Lee, and Thomas A. Kunkel. 2013a. “Ribonucleotide Incorporation, Proofreading and Bypass by Human DNA Polymerase δ .” *DNA Repair* 12 (2): 121–27.
- . 2013b. “Ribonucleotide Incorporation, Proofreading and Bypass by Human DNA Polymerase δ .” *DNA Repair* 12 (2): 121–27.
- Collins, A. R., I. M. Fleming, and C. M. Gedik. 1994. “In Vitro Repair of Oxidative and Ultraviolet-Induced DNA Damage in Supercoiled Nucleoid DNA by Human Cell Extract.” *Biochimica et Biophysica Acta* 1219 (3): 724–27.
- Collins, Andrew R. 2004. “The Comet Assay for DNA Damage and Repair: Principles, Applications, and Limitations.” *Molecular Biotechnology* 26 (3): 249–61.
- Connor, Ellen E., and Michael D. Wyatt. 2002. “Active-Site Clashes Prevent the Human 3-Methyladenine DNA Glycosylase from Improperly Removing Bases.” *Chemistry & Biology* 9 (9): 1033–41.
- Cooper, P. K., T. Nospikel, S. G. Clarkson, and S. A. Leadon. 1997. “Defective Transcription-Coupled Repair of Oxidative Base Damage in Cockayne Syndrome Patients from XP Group G.” *Science* 275 (5302): 990–93.
- Cornelio, Deborah A., Hailey N. C. Sedam, Jessica A. Ferrarezi, Nadia M. V. Sampaio, and Juan Lucas Argueso. 2017. “Both R-Loop Removal and Ribonucleotide Excision Repair Activities of RNase H2 Contribute Substantially to Chromosome Stability.” *DNA Repair* 52 (April): 110–14.
- Cortellino, Salvatore, Jinfei Xu, Mara Sannai, Robert Moore, Elena Caretti, Antonio Cigliano, Madeleine Le Coz, et al. 2011. “Thymine DNA Glycosylase Is Essential for Active DNA Demethylation by Linked Deamination-Base Excision Repair.” *Cell* 146 (1): 67–79.
- Crow, Yanick J., Andrea Leitch, Bruce E. Hayward, Anna Garner, Rekha Parmar, Elen Griffith, Manir Ali, et al. 2006. “Mutations in Genes Encoding Ribonuclease H2 Subunits Cause Aicardi-Goutieres Syndrome and Mimic Congenital Viral Brain Infection.” *Nature Genetics* 38 (8): 910–16.
- Cusanovich, Darren A., Riza Daza, Andrew Adey, Hannah A. Pliner, Lena Christiansen, Kevin L. Gunderson, Frank J. Steemers, Cole Trapnell, and Jay Shendure. 2015. “Multiplex Single Cell Profiling of Chromatin Accessibility by Combinatorial Cellular Indexing.” *Science* 348 (6237): 910–14.
- Dalhus, Bjørn, Monika Forsbring, Ina Høydal Helle, Erik Sebastian Vik, Rune Johansen Forstrøm, Paul Hoff Backe, Ingrun Alseth, and Magnar Bjørås. 2011. “Separation-of-Function Mutants Unravel the Dual-Reaction Mode of Human 8-Oxoguanine DNA Glycosylase.” *Structure* 19 (1): 117–27.
- Dantzer, Françoise, Luisa Luna, Magnar Bjørås, and Erling Seeberg. 2002. “Human OGG1 Undergoes Serine Phosphorylation and Associates with the Nuclear Matrix and Mitotic Chromatin in Vivo.” *Nucleic Acids Research* 30 (11): 2349–57.

- Datlinger, Paul, André F. Rendeiro, Thorina Boenke, Thomas Krausgruber, Daniele Barreca, and Christoph Bock. 2019. "Ultra-High Throughput Single-Cell RNA Sequencing by Combinatorial Fluidic Indexing." *BioRxiv*. <https://doi.org/10.1101/2019.12.17.879304>.
- Datlinger, Paul, André F. Rendeiro, Christian Schmidl, Thomas Krausgruber, Peter Traxler, Johanna Klughammer, Linda C. Schuster, Amelie Kuchler, Donat Alpar, and Christoph Bock. 2017. "Pooled CRISPR Screening with Single-Cell Transcriptome Readout." *Nature Methods* 14 (3): 297–301.
- David, Sheila S., Valerie L. O'Shea, and Sucharita Kundu. 2007. "Base-Excision Repair of Oxidative DNA Damage." *Nature* 447 (7147): 941–50.
- De Bont, Rinne, and Nik van Larebeke. 2004. "Endogenous DNA Damage in Humans: A Review of Quantitative Data." *Mutagenesis* 19 (3): 169–85.
- Deasy, Sarah K., Ryo Uehara, Suman K. Vodnala, Howard H. Yang, Randall A. Dass, Ying Hu, Maxwell P. Lee, Robert J. Crouch, and Kent W. Hunter. 2019. "Aicardi-Goutières Syndrome Gene Rnaseh2c Is a Metastasis Susceptibility Gene in Breast Cancer." *PLoS Genetics* 15 (5): e1008020.
- Delaney, Michael O., Carissa J. Wiederholt, and Marc M. Greenberg. 2002. "Fapy.DA Induces Nucleotide Misincorporation Translesionally by a DNA Polymerase." *Angewandte Chemie* 41 (5): 771–73.
- Di Cresce, C., R. Figueredo, P. J. Ferguson, M. D. Vincent, and J. Koropatnick. 2011. "Combining Small Interfering RNAs Targeting Thymidylate Synthase and Thymidine Kinase 1 or 2 Sensitizes Human Tumor Cells to 5-Fluorodeoxyuridine and Pemetrexed." *The Journal of Pharmacology and Experimental Therapeutics* 338 (3): 952–63.
- Di Noia, Javier M., and Michael S. Neuberger. 2007. "Molecular Mechanisms of Antibody Somatic Hypermutation." *Annual Review of Biochemistry* 76: 1–22.
- Di Noia, Javier M., Cristina Rada, and Michael S. Neuberger. 2006. "SMUG1 Is Able to Excise Uracil from Immunoglobulin Genes: Insight into Mutation versus Repair." *The EMBO Journal* 25 (3): 585–95.
- Dianov, G. L., T. V. Timchenko, O. I. Sinitsina, A. V. Kuzminov, O. A. Medvedev, and R. I. Salganik. 1991. "Repair of Uracil Residues Closely Spaced on the Opposite Strands of Plasmid DNA Results in Double-Strand Break and Deletion Formation." *Molecular & General Genetics: MGG* 225 (3): 448–52.
- Dierssen, Mara. 2012. "Down Syndrome: The Brain in Trisomic Mode." *Nature Reviews. Neuroscience* 13 (12): 844–58.
- Dion, Vincent. 2014. "Tissue Specificity in DNA Repair: Lessons from Trinucleotide Repeat Instability." *Trends in Genetics: TIG* 30 (6): 220–29.
- Dizdaroglu, M., B. Karahalil, S. Sentürker, T. J. Buckley, and T. Roldán-Arjona. 1999. "Excision of Products of Oxidative DNA Base Damage by Human NTH1 Protein." *Biochemistry* 38 (1): 243–46.
- Dogliotti, E., P. Fortini, B. Pascucci, and E. Parlanti. 2001. "The Mechanism of Switching among Multiple BER Pathways." *Progress in Nucleic Acid Research and Molecular Biology* 68: 3–27.
- Drabløs, Finn, Emadoldin Feyzi, Per Arne Aas, Cathrine B. Vaagbø, Bodil Kavli, Marit S. Bratlie, Javier Peña-Diaz, Marit Otterlei, Geir Slupphaug, and Hans E. Krokan. 2004. "Alkylation Damage in DNA and RNA--Repair Mechanisms and Medical Significance." *DNA Repair* 3 (11): 1389–1407.

- Duguid, J. R., J. N. Eble, T. M. Wilson, and M. R. Kelley. 1995. "Differential Cellular and Subcellular Expression of the Human Multifunctional Apurinic/Apyrimidinic Endonuclease (APE/Ref-1) DNA Repair Enzyme." *Cancer Research* 55 (24): 6097–6102.
- Duthie, Susan J., George Grant, Lynn P. Pirie, Amanda J. Watson, and Geoffrey P. Margison. 2010. "Folate Deficiency Alters Hepatic and Colon MGMT and OGG-1 DNA Repair Protein Expression in Rats but Has No Effect on Genome-Wide DNA Methylation." *Cancer Prevention Research* 3 (1): 92–100.
- Dyrkheeva, N. S., N. A. Lebedeva, and O. I. Lavrik. 2016. "AP Endonuclease 1 as a Key Enzyme in Repair of Apurinic/Apyrimidinic Sites." *Biochemistry. Biokhimiia* 81 (9): 951–67.
- Egeblad, Mikala, and Zena Werb. 2002. "New Functions for the Matrix Metalloproteinases in Cancer Progression." *Nature Reviews. Cancer* 2 (3): 161–74.
- Eide, L., L. Luna, E. C. Gustad, P. T. Henderson, J. M. Essigmann, B. Demple, and E. Seeberg. 2001. "Human Endonuclease III Acts Preferentially on DNA Damage Opposite Guanine Residues in DNA." *Biochemistry* 40 (22): 6653–59.
- Encell, L. P., and L. A. Loeb. 1999a. "Redesigning the Substrate Specificity of Human O(6)-Alkylguanine-DNA Alkyltransferase. Mutants with Enhanced Repair of O(4)-Methylthymine." *Biochemistry* 38 (37): 12097–103.
- . 1999b. "Redesigning the Substrate Specificity of Human O(6)-Alkylguanine-DNA Alkyltransferase. Mutants with Enhanced Repair of O(4)-Methylthymine." *Biochemistry* 38 (37): 12097–103.
- . 1999c. "Redesigning the Substrate Specificity of Human O(6)-Alkylguanine-DNA Alkyltransferase. Mutants with Enhanced Repair of O(4)-Methylthymine." *Biochemistry* 38 (37): 12097–103.
- Eng, Chee-Huat Linus, Michael Lawson, Qian Zhu, Ruben Dries, Noushin Koulou, Yodai Takei, Jina Yun, et al. 2019. "Transcriptome-Scale Super-Resolved Imaging in Tissues by RNA SeqFISH." *Nature* 568 (7751): 235–39.
- Engelward, B. P., J. M. Allan, A. J. Dreslin, J. D. Kelly, M. M. Wu, B. Gold, and L. D. Samson. 1998. "A Chemical and Genetic Approach Together Define the Biological Consequences of 3-Methyladenine Lesions in the Mammalian Genome." *The Journal of Biological Chemistry* 273 (9): 5412–18.
- Engelward, B. P., A. Dreslin, J. Christensen, D. Huszar, C. Kurahara, and L. Samson. 1996. "Repair-Deficient 3-Methyladenine DNA Glycosylase Homozygous Mutant Mouse Cells Have Increased Sensitivity to Alkylation-Induced Chromosome Damage and Cell Killing." *The EMBO Journal* 15 (4): 945–52.
- Evans, A. R., M. Limp-Foster, and M. R. Kelley. 2000. "Going APE over Ref-1." *Mutation Research* 461 (2): 83–108.
- Evans, Mark D., Miral Dizdaroglu, and Marcus S. Cooke. 2004. "Oxidative DNA Damage and Disease: Induction, Repair and Significance." *Mutation Research* 567 (1): 1–61.
- Everhard, Sibille, Gintian Kaloshi, Emmanuelle Crinière, Alexandra Benouaich-Amiel, Julie Lejeune, Yannick Marie, Marc Sanson, et al. 2006. "MGMT Methylation: A Marker of Response to Temozolomide in Low-Grade Gliomas." *Annals of Neurology* 60 (6): 740–43.
- Fan, C-H, W-L Liu, H. Cao, C. Wen, L. Chen, and G. Jiang. 2013. "O6-Methylguanine DNA Methyltransferase as a Promising Target for the Treatment of Temozolomide-Resistant Gliomas." *Cell Death & Disease* 4 (October): e876.

- Feng, Yuqing, Mariam H. Goubran, Tyson B. Follack, and Linda Chelico. 2017. “Deamination-Independent Restriction of LINE-1 Retrotransposition by APOBEC3H.” *Scientific Reports* 7 (1): 10881.
- Fishel, Melissa L., Ying He, Martin L. Smith, and Mark R. Kelley. 2007. “Manipulation of Base Excision Repair to Sensitize Ovarian Cancer Cells to Alkylating Agent Temozolomide.” *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* 13 (1): 260–67.
- Fleming, Stephen J., John C. Marioni, and Mehrtash Babadi. 2019. “CellBender Remove-Background: A Deep Generative Model for Unsupervised Removal of Background Noise from ScRNA-Seq Datasets.” *BioRxiv*. <https://doi.org/10.1101/791699>.
- Forestier, Anne, Thierry Douki, Sylvie Sauvaigo, Viviana De Rosa, Christine Demeilliers, and Walid Rachidi. 2012a. “Alzheimer’s Disease-Associated Neurotoxic Peptide Amyloid- β Impairs Base Excision Repair in Human Neuroblastoma Cells.” *International Journal of Molecular Sciences* 13 (11): 14766–87.
- . 2012b. “Alzheimer’s Disease-Associated Neurotoxic Peptide Amyloid- β Impairs Base Excision Repair in Human Neuroblastoma Cells.” *International Journal of Molecular Sciences* 13 (11): 14766–87.
- Forestier, Anne, Fanny Sarrazy, Sylvain Caillat, Yves Vandembrouck, and Sylvie Sauvaigo. 2012. “Functional DNA Repair Signature of Cancer Cell Lines Exposed to a Set of Cytotoxic Anticancer Drugs Using a Multiplexed Enzymatic Repair Assay on Biochip.” *PloS One* 7 (12): e51754.
- Fortini, P., E. Parlanti, O. M. Sidorkina, J. Laval, and E. Dogliotti. 1999. “The Type of DNA Glycosylase Determines the Base Excision Repair Pathway in Mammalian Cells.” *The Journal of Biological Chemistry* 274 (21): 15230–36.
- Friedman, Joshua I., Ananya Majumdar, and James T. Stivers. 2009. “Nontarget DNA Binding Shapes the Dynamic Landscape for Enzymatic Recognition of DNA Damage.” *Nucleic Acids Research* 37 (11): 3493–3500.
- Fritz, Gerhard, Sabine Grösch, Maja Tomicic, and Bernd Kaina. 2003. “APE/Ref-1 and the Mammalian Response to Genotoxic Stress.” *Toxicology* 193 (1–2): 67–78.
- Frye, Michaela, Bryan T. Harada, Mikaela Behm, and Chuan He. 2018. “RNA Modifications Modulate Gene Expression during Development.” *Science* 361 (6409): 1346–49.
- Fu, Dragony, Jennifer A. Calvo, and Leona D. Samson. 2012. “Balancing Repair and Tolerance of DNA Damage Caused by Alkylating Agents.” *Nature Reviews. Cancer* 12 (2): 104–20.
- Gad, Helge, Tobias Koolmeister, Ann-Sofie Jemth, Saeed Eshtad, Sylvain A. Jacques, Cecilia E. Ström, Linda M. Svensson, et al. 2014. “MTH1 Inhibition Eradicates Cancer by Preventing Sanitation of the DNTP Pool.” *Nature* 508 (7495): 215–21.
- Gao, Yankun, Sachin Katyal, Youngsoo Lee, Jingfeng Zhao, Jerold E. Rehg, Helen R. Russell, and Peter J. McKinnon. 2011. “DNA Ligase III Is Critical for MtDNA Integrity but Not Xrcc1-Mediated Nuclear DNA Repair.” *Nature* 471 (7337): 240–44.
- Golato, Tyler, Boris Brennerman, Daniel R. McNeill, Jianfeng Li, Robert W. Sobol, and David M. Wilson 3rd. 2017. “Development of a Cell-Based Assay for Measuring Base Excision Repair Responses.” *Scientific Reports* 7 (1): 13007.
- Gonen, Nitzan, and Yehuda G. Assaraf. 2012. “Antifolates in Cancer Therapy: Structure, Activity and Mechanisms of Drug Resistance.” *Drug Resistance Updates: Reviews and Commentaries in Antimicrobial and Anticancer Chemotherapy* 15 (4): 183–210.

- Goulian, M., B. Bleile, and B. Y. Tseng. 1980a. "Methotrexate-Induced Misincorporation of Uracil into DNA." *Proceedings of the National Academy of Sciences of the United States of America* 77 (4): 1956–60.
- . 1980b. "Methotrexate-Induced Misincorporation of Uracil into DNA." *Proceedings of the National Academy of Sciences of the United States of America* 77 (4): 1956–60.
- . 1980c. "The Effect of Methotrexate on Levels of DUTP in Animal Cells." *The Journal of Biological Chemistry* 255 (22): 10630–37.
- Gros, Laurent, Alexander A. Ishchenko, Hiroshi Ide, Rhoderick H. Elder, and Murat K. Saparbaev. 2004. "The Major Human AP Endonuclease (Ape1) Is Involved in the Nucleotide Incision Repair Pathway." *Nucleic Acids Research* 32 (1): 73–81.
- Grösch, S., G. Fritz, and B. Kaina. 1998. "Apurinic Endonuclease (Ref-1) Is Induced in Mammalian Cells by Oxidative Stress and Involved in Clastogenic Adaptation." *Cancer Research* 58 (19): 4410–16.
- Guillet, Marie, Patricia Auffret Van Der Kemp, and Serge Boiteux. 2006. "DUTPase Activity Is Critical to Maintain Genetic Stability in *Saccharomyces Cerevisiae*." *Nucleic Acids Research* 34 (7): 2056–66.
- Guo, Chuner, Wenjun Kong, Kenji Kamimoto, Guillermo C. Rivera-Gonzalez, Xue Yang, Yuhei Kiritani, and Samantha A. Morris. 2019. "CellTag Indexing: Genetic Barcode-Based Sample Multiplexing for Single-Cell Genomics." *Genome Biology* 20 (1): 90.
- Hakem, Razqallah. 2008. "DNA-Damage Repair; the Good, the Bad, and the Ugly." *The EMBO Journal* 27 (4): 589–605.
- Han, Dandan, Lars Schomacher, Katrin M. Schüle, Medhavi Mallick, Michael U. Musheev, Emil Karaulanov, Laura Krebs, Annika von Seggern, and Christof Niehrs. 2019. "NEIL1 and NEIL2 DNA Glycosylases Protect Neural Crest Development against Mitochondrial Oxidative Stress." *ELife* 8 (September). <https://doi.org/10.7554/eLife.49044>.
- Hansen, Erik C., Monica Ransom, Jay R. Hesselberth, Nina N. Hosmane, Adam A. Capoferri, Katherine M. Bruner, Ross A. Pollack, et al. 2016. "Diverse Fates of Uracilated HIV-1 DNA during Infection of Myeloid Lineage Cells." *ELife* 5 (September). <https://doi.org/10.7554/eLife.18447>.
- Hashimoto, Keiji, Yohei Tominaga, Yusaku Nakabeppu, and Masaaki Moriya. 2004. "Futile Short-Patch DNA Base Excision Repair of Adenine:8-Oxoguanine Mismatch." *Nucleic Acids Research* 32 (19): 5928–34.
- He, Yu-Fei, Bin-Zhong Li, Zheng Li, Peng Liu, Yang Wang, Qingyu Tang, Jianping Ding, et al. 2011. "Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA." *Science* 333 (6047): 1303–7.
- Hegde, Muralidhar L., Pavana M. Hegde, Larry J. Bellot, Santi M. Mandal, Tapas K. Hazra, Guo-Min Li, Istvan Boldogh, Alan E. Tomkinson, and Sankar Mitra. 2013. "Prereplicative Repair of Oxidized Bases in the Human Genome Is Mediated by NEIL1 DNA Glycosylase Together with Replication Proteins." *Proceedings of the National Academy of Sciences of the United States of America* 110 (33): E3090-9.
- Heimberg, Graham, Rajat Bhatnagar, Hana El-Samad, and Matt Thomson. 2016. "Low Dimensionality in Gene Expression Data Enables the Accurate Extraction of Transcriptional Programs from Shallow Sequencing." *Cell Systems* 2 (4): 239–50.
- Hemminki, K., K. Peltonen, and P. Vodicka. 1989. "Depurination from DNA of 7-Methylguanine, 7-(2-Aminoethyl)-Guanine and Ring-Opened 7-Methylguanines." *Chemico-Biological Interactions* 70 (3–4): 289–303.

- Hendrich, B., U. Hardeland, H. H. Ng, J. Jiricny, and A. Bird. 1999. "The Thymine Glycosylase MBD4 Can Bind to the Product of Deamination at Methylated CpG Sites." *Nature* 401 (6750): 301–4.
- Hickman, M. J., and L. D. Samson. 1999. "Role of DNA Mismatch Repair and P53 in Signaling Induction of Apoptosis by Alkylating Agents." *Proceedings of the National Academy of Sciences of the United States of America* 96 (19): 10764–69.
- Hickman, Mark J., and Leona D. Samson. 2004. "Apoptotic Signaling in Response to a Single Type of DNA Lesion, O(6)-Methylguanine." *Molecular Cell* 14 (1): 105–16.
- Hill, J. W., T. K. Hazra, T. Izumi, and S. Mitra. 2001. "Stimulation of Human 8-Oxoguanine-DNA Glycosylase by AP-Endonuclease: Potential Coordination of the Initial Steps in Base Excision Repair." *Nucleic Acids Research* 29 (2): 430–38.
- Hiller, Björn, Anja Hoppe, Christa Haase, Christina Hiller, Nadja Schubert, Werner Müller, Martin A. M. Reijns, et al. 2018. "Ribonucleotide Excision Repair Is Essential to Prevent Squamous Cell Carcinoma of the Skin." *Cancer Research* 78 (20): 5917–26.
- Hirmondo, Rita, Anna Lopata, Eva Viola Suranyi, Beata G. Vertessy, and Judit Toth. 2017. "Differential Control of DNTP Biosynthesis and Genome Integrity Maintenance by the DUTPase Superfamily Enzymes." *Scientific Reports* 7 (1): 6043.
- Hirota, Kiichi, Minoru Matsui, Satoshi Iwata, Akira Nishiyama, Kenjiro Mori, and Junji Yodoi. 1997. "AP-1 Transcriptional Activity Is Regulated by a Direct Association between Thioredoxin and Ref-1." *Proceedings of the National Academy of Sciences of the United States of America* 94 (8): 3633–38.
- Horiguchi, Michiko, Jahye Kim, Naoya Matsunaga, Hiroaki Kaji, Takashi Egawa, Kazutaka Makino, Satoru Koyanagi, and Shigehiro Ohdo. 2010. "Glucocorticoid-Dependent Expression of O(6)-Methylguanine-DNA Methyltransferase Gene Modulates Dacarbazine-Induced Hepatotoxicity in Mice." *The Journal of Pharmacology and Experimental Therapeutics* 333 (3): 782–87.
- Houston, Kaiulani M., Adam T. Melvin, Gregory S. Woss, Effrat L. Fayer, Marcey L. Waters, and Nancy L. Allbritton. 2017. "Development of β -Hairpin Peptides for the Measurement of SCF-Family E3 Ligase Activity in Vitro via Ornithine Ubiquitination." *ACS Omega* 2 (3): 1198–1206.
- Hu, Jingping, Syed Z. Imam, Kazunari Hashiguchi, Nadja C. de Souza-Pinto, and Vilhelm A. Bohr. 2005. "Phosphorylation of Human Oxoguanine DNA Glycosylase (Alpha-OGG1) Modulates Its Function." *Nucleic Acids Research* 33 (10): 3271–82.
- Hu, Jingping, Nadja C. de Souza-Pinto, Kazuhiro Haraguchi, Barbara A. Hogue, Pawel Jaruga, Marc M. Greenberg, Miral Dizdaroglu, and Vilhelm A. Bohr. 2005. "Repair of Formamidopyrimidines in DNA Involves Different Glycosylases: Role of the OGG1, NTH1, and NEIL1 Enzymes." *The Journal of Biological Chemistry* 280 (49): 40544–51.
- Hu, Lulu, Junyan Lu, Jingdong Cheng, Qinhui Rao, Ze Li, Haifeng Hou, Zhiyong Lou, et al. 2015. "Structural Insight into Substrate Preference for TET-Mediated Oxidation." *Nature* 527 (7576): 118–22.
- Hu, Shaohui, Jun Wan, Yijing Su, Qifeng Song, Yaxue Zeng, Ha Nam Nguyen, Jaehoon Shin, et al. 2013. "DNA Methylation Presents Distinct Binding Sites for Human Transcription Factors." *ELife* 2 (September): e00726.
- Huang, Chung-Ming, Shih-Yin Chen, Po-Hao Huang, and Fuu-Jen Tsai. 2015. "Effect of MPG Gene Rs2858056 Polymorphism, Copy Number Variation, and Level of Serum MPG Protein on the Risk for Rheumatoid Arthritis." *PLoS One* 10 (3): e0120699.

- Huber, L. C., O. Distler, I. Tarner, R. E. Gay, S. Gay, and T. Pap. 2006. "Synovial Fibroblasts: Key Players in Rheumatoid Arthritis." *Rheumatology* 45 (6): 669–75.
- Huehls, Amelia M., Catherine J. Huntoon, Poorval M. Joshi, Carly A. Baehr, Jill M. Wagner, Xiaoxiao Wang, Marietta Y. Lee, and Larry M. Karnitz. 2016. "Genomically Incorporated 5-Fluorouracil That Escapes UNG-Initiated Base Excision Repair Blocks DNA Replication and Activates Homologous Recombination." *Molecular Pharmacology* 89 (1): 53–62.
- Hustedt, Nicole, Alejandro Álvarez-Quilón, Andrea McEwan, Jing Yi Yuan, Tiffany Cho, Lisa Koob, Traver Hart, and Daniel Durocher. 2019. "A Consensus Set of Genetic Vulnerabilities to ATR Inhibition." *Open Biology* 9 (9): 190156.
- Impellizzeri, K. J., B. Anderson, and P. M. Burgers. 1991. "The Spectrum of Spontaneous Mutations in a *Saccharomyces Cerevisiae* Uracil-DNA-Glycosylase Mutant Limits the Function of This Enzyme to Cytosine Deamination Repair." *Journal of Bacteriology* 173 (21): 6807–10.
- Islam, Saiful, Amit Zeisel, Simon Joost, Gioele La Manno, Pawel Zajac, Maria Kasper, Peter Lönnerberg, and Sten Linnarsson. 2014. "Quantitative Single-Cell RNA-Seq with Unique Molecular Identifiers." *Nature Methods* 11 (2): 163–66.
- Ito, Shinsuke, Li Shen, Qing Dai, Susan C. Wu, Leonard B. Collins, James A. Swenberg, Chuan He, and Yi Zhang. 2011. "Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine." *Science* 333 (6047): 1300–1303.
- Jackson, Stephen P., and Jiri Bartek. 2009. "The DNA-Damage Response in Human Biology and Disease." *Nature* 461 (7267): 1071–78.
- Jacobs, Angelika L., and Primo Schär. 2012. "DNA Glycosylases: In DNA Repair and Beyond." *Chromosoma* 121 (1): 1–20.
- Jaiswal, M., N. F. LaRusso, N. Nishioka, Y. Nakabeppu, and G. J. Gores. 2001. "Human Ogg1, a Protein Involved in the Repair of 8-Oxoguanine, Is Inhibited by Nitric Oxide." *Cancer Research* 61 (17): 6388–93.
- Jaitin, Diego Adhemar, Assaf Weiner, Ido Yofe, David Lara-Astiaso, Hadas Keren-Shaul, Eyal David, Tomer Meir Salame, Amos Tanay, Alexander van Oudenaarden, and Ido Amit. 2016. "Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq." *Cell* 167 (7): 1883-1896.e15.
- Jalal, Shadia, Jennifer N. Earley, and John J. Turchi. 2011. "DNA Repair: From Genome Maintenance to Biomarker and Therapeutic Target." *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* 17 (22): 6973–84.
- Jaruga, Pawel, Mustafa Birincioglu, Thomas A. Rosenquist, and Miral Dizdaroglu. 2004. "Mouse NEIL1 Protein Is Specific for Excision of 2,6-Diamino-4-Hydroxy-5-Formamidopyrimidine and 4,6-Diamino-5-Formamidopyrimidine from Oxidatively Damaged DNA." *Biochemistry* 43 (50): 15909–14.
- Jensen, Anne, Guillaume Calvayrac, Bensus Karahalil, Vilhelm A. Bohr, and Tinna Stevnsner. 2003. "Mammalian 8-Oxoguanine DNA Glycosylase 1 Incises 8-Oxoadenine Opposite Cytosine in Nuclei and Mitochondria, While a Different Glycosylase Incises 8-Oxoadenine Opposite Guanine in Nuclei." *The Journal of Biological Chemistry* 278 (21): 19541–48.
- Jeong, Ho-Sang, Peter S. Backlund, Hao-Chia Chen, Alexander A. Karavanov, and Robert J. Crouch. 2004. "RNase H2 of *Saccharomyces Cerevisiae* Is a Complex of Three Proteins." *Nucleic Acids Research* 32 (2): 407–14.

- Jetson, Rachael R., and Casey J. Krusemark. 2016. "Sensing Enzymatic Activity by Exposure and Selection of DNA-Encoded Probes." *Angewandte Chemie, International Edition* 55 (33): 9562–66.
- Jiang, Yu Lin, and James T. Stivers. 2002. "Mutational Analysis of the Base-Flipping Mechanism of Uracil DNA Glycosylase." *Biochemistry* 41 (37): 11236–47.
- Jin, Ke, Le Ou-Yang, Xing-Ming Zhao, Hong Yan, and Xiao-Fei Zhang. 2020. "ScTSSR: Gene Expression Recovery for Single-Cell RNA Sequencing Using Two-Side Sparse Self-Representation." *Bioinformatics*, February. <https://doi.org/10.1093/bioinformatics/btaa108>.
- Johnson, Jennifer M., and Jean J. Latimer. 2005. "Analysis of DNA Repair Using Transfection-Based Host Cell Reactivation." *Methods in Molecular Biology* 291: 321–35.
- Johnson, Robert E., Sung-Lim Yu, Satya Prakash, and Louise Prakash. 2007. "A Role for Yeast and Human Translesion Synthesis DNA Polymerases in Promoting Replication through 3-Methyl Adenine." *Molecular and Cellular Biology* 27 (20): 7198–7205.
- Kaina, Bernd, Aliko Ziouta, Kirsten Ochs, and Therese Coquerelle. 1997. "Chromosomal Instability, Reproductive Cell Death and Apoptosis Induced by O6-Methylguanine in Mex-, Mex+ and Methylation-Tolerant Mismatch Repair Compromised Cells: Facts and Models." *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 381 (2): 227–41.
- Kakolyris, S., L. Kaklamanis, K. Engels, S. B. Fox, M. Taylor, I. D. Hickson, K. C. Gatter, and A. L. Harris. 1998. "Human AP Endonuclease 1 (HAP1) Protein Expression in Breast Cancer Correlates with Lymph Node Status and Angiogenesis." *British Journal of Cancer* 77 (7): 1169–73.
- Kamiya, H., and H. Kasai. 1995. "2-Hydroxyadenine (Isoguanine) as Oxidative DNA Damage: Its Formation and Mutation Inducibility." *Nucleic Acids Symposium Series*, no. 34: 233–34.
- Kang, D., J. Nishida, A. Iyama, Y. Nakabeppu, M. Furuichi, T. Fujiwara, M. Sekiguchi, and K. Takeshige. 1995. "Intracellular Localization of 8-Oxo-DGTPase in Human Cells, with Special Reference to the Role of the Enzyme in Mitochondria." *The Journal of Biological Chemistry* 270 (24): 14659–65.
- Karahalil, Bensu, Barbara A. Hogue, Nadja C. de Souza-Pinto, and Vilhelm A. Bohr. 2002. "Base Excision Repair Capacity in Mitochondria and Nuclei: Tissue-Specific Variations." *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology* 16 (14): 1895–1902.
- Karran, P., and T. Lindahl. 1980. "Hypoxanthine in Deoxyribonucleic Acid: Generation by Heat-Induced Hydrolysis of Adenine Residues and Release in Free Form by a Deoxyribonucleic Acid Glycosylase from Calf Thymus." *Biochemistry* 19 (26): 6005–11.
- Kasai, H., and S. Nishimura. 1984. "Hydroxylation of Deoxyguanosine at the C-8 Position by Ascorbic Acid and Other Reducing Agents." *Nucleic Acids Research* 12 (4): 2137–45.
- Kasai, Hiroshi. 1997. "Analysis of a Form of Oxidative DNA Damage, 8-Hydroxy-2'-Deoxyguanosine, as a Marker of Cellular Oxidative Stress during Carcinogenesis." *Mutation Research/Reviews in Mutation Research* 387 (3): 147–63.
- Kat, A., W. G. Thilly, W. H. Fang, M. J. Longley, G. M. Li, and P. Modrich. 1993. "An Alkylation-Tolerant, Mutator Human Cell Line Is Deficient in Strand-Specific Mismatch Repair." *Proceedings of the National Academy of Sciences of the United States of America* 90 (14): 6424–28.

- Kavli, Bodil, Marit Otterlei, Geir Slupphaug, and Hans E. Krokan. 2007. "Uracil in DNA--General Mutagen, but Normal Intermediate in Acquired Immunity." *DNA Repair* 6 (4): 505–16.
- Kavli, Bodil, Ottar Sundheim, Mansour Akbari, Marit Otterlei, Hilde Nilsen, Frank Skorpen, Per Arne Aas, Lars Hagen, Hans E. Krokan, and Geir Slupphaug. 2002. "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." *The Journal of Biological Chemistry* 277 (42): 39926–36.
- Kellner, Vanessa, and Brian Luke. 2020. "Molecular and Physiological Consequences of Faulty Eukaryotic Ribonucleotide Excision Repair." *The EMBO Journal* 39 (3): e102309.
- Kennedy, C. H., R. Cueto, S. A. Belinsky, J. F. Lechner, and W. A. Pryor. 1998. "Overexpression of HMTH1 mRNA: A Molecular Marker of Oxidative Stress in Lung Cancer Cells." *FEBS Letters* 429 (1): 17–20.
- Kent, W. James, Charles W. Sugnet, Terrence S. Furey, Krishna M. Roskin, Tom H. Pringle, Alan M. Zahler, and David Haussler. 2002. "The Human Genome Browser at UCSC." *Genome Research* 12 (6): 996–1006.
- Khodyreva, S. N., R. Prasad, E. S. Ilina, M. V. Sukhanova, M. M. Kutuzov, Y. Liu, E. W. Hou, S. H. Wilson, and O. I. Lavrik. 2010a. "Apurinic/Apyrimidinic (AP) Site Recognition by the 5'-DRP/AP Lyase in Poly(ADP-Ribose) Polymerase-1 (PARP-1)." *Proceedings of the National Academy of Sciences of the United States of America* 107 (51): 22090–95.
- . 2010b. "Apurinic/Apyrimidinic (AP) Site Recognition by the 5'-DRP/AP Lyase in Poly(ADP-Ribose) Polymerase-1 (PARP-1)." *Proceedings of the National Academy of Sciences of the United States of America* 107 (51): 22090–95.
- Kim, Kyu-Tae, Hye Won Lee, Hae-Ock Lee, Hye Jin Song, Da Eun Jeong, Sang Shin, Hyunho Kim, et al. 2016. "Application of Single-Cell RNA Sequencing in Optimizing a Combinatorial Therapeutic Strategy in Metastatic Renal Cell Carcinoma." *Genome Biology* 17 (April): 80.
- Kim, N. K., S. H. Lee, K. Y. Cha, and J. S. Seo. 1998. "Tissue-Specific Expression and Activation of N-Methylpurine-DNA Glycosylase in Thymic Carcinomas of Transgenic Mice Expressing the SV40 Large T-Antigen Gene." *Molecules and Cells* 8 (4): 383–87.
- Kim, Nayun, Shar-Yin N. Huang, Jessica S. Williams, Yue C. Li, Alan B. Clark, Jang-Eun Cho, Thomas A. Kunkel, Yves Pommier, and Sue Jinks-Robertson. 2011. "Mutagenic Processing of Ribonucleotides in DNA by Yeast Topoisomerase I." *Science* 332 (6037): 1561–64.
- Kim, Wan-Cheol, Dustin King, and Chow H. Lee. 2010. "RNA-Cleaving Properties of Human Apurinic/Apyrimidinic Endonuclease 1 (APE1)." *International Journal of Biochemistry and Molecular Biology* 1 (1): 12–25.
- Kitange, Gaspar J., Brett L. Carlson, Mark A. Schroeder, Patrick T. Grogan, Jeff D. Lamont, Paul A. Decker, Wenting Wu, C. David James, and Jann N. Sarkaria. 2009. "Induction of MGMT Expression Is Associated with Temozolomide Resistance in Glioblastoma Xenografts." *Neuro-Oncology* 11 (3): 281–91.
- Klungland, A., and T. Lindahl. 1997. "Second Pathway for Completion of Human DNA Base Excision-Repair: Reconstitution with Purified Proteins and Requirement for DNase IV (FEN1)." *The EMBO Journal* 16 (11): 3341–48.
- Klungland, Arne, and Adam B. Robertson. 2017. "Oxidized C5-Methyl Cytosine Bases in DNA: 5-Hydroxymethylcytosine; 5-Formylcytosine; and 5-Carboxycytosine." *Free Radical Biology & Medicine* 107 (June): 62–68.

- Koh, Kyung Duk, Sathya Balachander, Jay R. Hesselberth, and Francesca Storici. 2015. "Ribose-Seq: Global Mapping of Ribonucleotides Embedded in Genomic DNA." *Nature Methods* 12 (3): 251–57, 3 p following 257.
- Kow, Yoke W. 2002. "Repair of Deaminated Bases in DNA." *Free Radical Biology & Medicine* 33 (7): 886–93.
- Kraemer, K. H., M. M. Lee, and J. Scotto. 1984. "DNA Repair Protects against Cutaneous and Internal Neoplasia: Evidence from Xeroderma Pigmentosum." *Carcinogenesis* 5 (4): 511–14.
- Krahn, Joseph M., William A. Beard, Holly Miller, Arthur P. Grollman, and Samuel H. Wilson. 2003. "Structure of DNA Polymerase Beta with the Mutagenic DNA Lesion 8-Oxodeoxyguanine Reveals Structural Insights into Its Coding Potential." *Structure* 11 (1): 121–27.
- Kreth, Simone, Elisabeth Limbeck, Ludwig C. Hinske, Stefanie V. Schütz, Niklas Thon, Kai Hoefig, Rupert Egensperger, and Friedrich W. Kreth. 2013. "In Human Glioblastomas Transcript Elongation by Alternative Polyadenylation and MiRNA Targeting Is a Potent Mechanism of MGMT Silencing." *Acta Neuropathologica* 125 (5): 671–81.
- Krokan, Hans E., and Magnar Bjørås. 2013. "Base Excision Repair." *Cold Spring Harbor Perspectives in Biology* 5 (4): a012583.
- Krokan, Hans E., Finn Drabløs, and Geir Slupphaug. 2002. "Uracil in DNA--Occurrence, Consequences and Repair." *Oncogene* 21 (58): 8935–48.
- Kronenberg, Golo, Christoph Harms, Robert W. Sobol, Fernando Cardozo-Pelaez, Heinz Linhart, Benjamin Winter, Mustafa Balkaya, et al. 2008. "Folate Deficiency Induces Neurodegeneration and Brain Dysfunction in Mice Lacking Uracil DNA Glycosylase." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 28 (28): 7219–30.
- Kunkel, Thomas A., John D. Roberts, and Akio Sugino. 1991. "The Fidelity of DNA Synthesis by the Catalytic Subunit of Yeast DNA Polymerase α Alone and with Accessory Proteins." *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 250 (1): 175–82.
- Kuraoka, Isao, Christina Bender, Anthony Romieu, Jean Cadet, Richard D. Wood, and Tomas Lindahl. 2000. "Removal of Oxygen Free-Radical-Induced 5',8-Purine Cyclodeoxynucleosides from DNA by the Nucleotide Excision-Repair Pathway in Human Cells." *Proceedings of the National Academy of Sciences of the United States of America* 97 (8): 3832–37.
- Ladner, R. D., F. J. Lynch, S. Groshen, Y. P. Xiong, A. Sherrod, S. J. Caradonna, J. Stoehlmacher, and H. J. Lenz. 2000. "DUTP Nucleotidohydrolase Isoform Expression in Normal and Neoplastic Tissues: Association with Survival and Response to 5-Fluorouracil in Colorectal Cancer." *Cancer Research* 60 (13): 3493–3503.
- Landry, Sébastien, Iñigo Narvaiza, Daniel C. Linfesty, and Matthew D. Weitzman. 2011. "APOBEC3A Can Activate the DNA Damage Response and Cause Cell-Cycle Arrest." *EMBO Reports* 12 (5): 444–50.
- Langie, Sabine A. S., Amaya Azqueta, and Andrew R. Collins. 2015a. "The Comet Assay: Past, Present, and Future." *Frontiers in Genetics* 6 (August): 266.
- . 2015b. "The Comet Assay: Past, Present, and Future." *Frontiers in Genetics* 6 (August): 266.

- Langmead, Ben, and Steven L. Salzberg. 2012a. “Fast Gapped-Read Alignment with Bowtie 2.” *Nature Methods* 9 (4): 357–59.
- . 2012b. “Fast Gapped-Read Alignment with Bowtie 2.” *Nature Methods* 9 (4): 357–59.
- Lau, A. Y., M. D. Wyatt, B. J. Glassner, L. D. Samson, and T. Ellenberger. 2000. “Molecular Basis for Discriminating between Normal and Damaged Bases by the Human Alkyladenine Glycosylase, AAG.” *Proceedings of the National Academy of Sciences of the United States of America* 97 (25): 13573–78.
- Lawley, P. D., and W. Warren. 1976. “Removal of Minor Methylation Products 7-Methyl Adenine and 3-Methylguanine from DNA of *Escherichia Coli* Treated with Dimethyl Sulphate.” *Chemico-Biological Interactions* 12 (2): 211–20.
- Lebedeva, Natalia A., Nadejda I. Rechkunova, Sherif F. El-Khamisy, and Olga I. Lavrik. 2012a. “Tyrosyl-DNA Phosphodiesterase 1 Initiates Repair of Apurinic/Apyrimidinic Sites.” *Biochimie* 94 (8): 1749–53.
- . 2012b. “Tyrosyl-DNA Phosphodiesterase 1 Initiates Repair of Apurinic/Apyrimidinic Sites.” *Biochimie* 94 (8): 1749–53.
- Lee, Chun-Yue I., James C. Delaney, Maria Kartalou, Gondichatnahalli M. Lingaraju, Ayelet Maor-Shoshani, John M. Essigmann, and Leona D. Samson. 2009. “Recognition and Processing of a New Repertoire of DNA Substrates by Human 3-Methyladenine DNA Glycosylase (AAG).” *Biochemistry* 48 (9): 1850–61.
- Lee, Shao Chin, and Juliana C. N. Chan. 2015. “Evidence for DNA Damage as a Biological Link between Diabetes and Cancer.” *Chinese Medical Journal* 128 (11): 1543–48.
- Leitner-Dagan, Yael, Ziv Sevilya, Mila Pinchev, Ran Kremer, Dalia Elinger, Hedy S. Rennert, Edna Schechtman, et al. 2014. “Enzymatic MPG DNA Repair Assays for Two Different Oxidative DNA Lesions Reveal Associations with Increased Lung Cancer Risk.” *Carcinogenesis* 35 (12): 2763–70.
- Lescroart, Fabienne, Xiaonan Wang, Xionghui Lin, Benjamin Swedlund, Souhir Gargouri, Adriana Sánchez-Dànes, Victoria Moignard, et al. 2018. “Defining the Earliest Step of Cardiovascular Lineage Segregation by Single-Cell RNA-Seq.” *Science* 359 (6380): 1177–81.
- Leun, Anne M. van der, Daniela S. Thommen, and Ton N. Schumacher. 2020. “CD8+ T Cell States in Human Cancer: Insights from Single-Cell Analysis.” *Nature Reviews. Cancer* 20 (4): 218–32.
- Levy, M., and S. L. Miller. 1998. “The Stability of the RNA Bases: Implications for the Origin of Life.” *Proceedings of the National Academy of Sciences of the United States of America* 95 (14): 7933–38.
- Li, Fengling, Steven Kennedy, Taraneh Hajian, Elisa Gibson, Alma Seitova, Chao Xu, Cheryl H. Arrowsmith, and Masoud Vedadi. 2016. “A Radioactivity-Based Assay for Screening Human M6A-RNA Methyltransferase, METTL3-METTL14 Complex, and Demethylase ALKBH5.” *Journal of Biomolecular Screening* 21 (3): 290–97.
- Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. 2009a. “The Sequence Alignment/Map Format and SAMtools.” *Bioinformatics* 25 (16): 2078–79.
- . 2009b. “The Sequence Alignment/Map Format and SAMtools.” *Bioinformatics* 25 (16): 2078–79.

- Li, Yingfu, and Ronald R. Breaker. 1999. "Kinetics of RNA Degradation by Specific Base Catalysis of Transesterification Involving the 2'-Hydroxyl Group." *Journal of the American Chemical Society* 121 (23): 5364–72.
- Lian, Christine Guo, Yufei Xu, Craig Ceol, Feizhen Wu, Allison Larson, Karen Dresser, Wenqi Xu, et al. 2012. "Loss of 5-Hydroxymethylcytosine Is an Epigenetic Hallmark of Melanoma." *Cell* 150 (6): 1135–46.
- Lima, Walt F., John B. Rose, Josh G. Nichols, Hongjiang Wu, Michael T. Migawa, Tadeusz K. Wyrzykiewicz, Andrew M. Siwkowski, and Stanley T. Crooke. 2007. "Human RNase H1 Discriminates between Subtle Variations in the Structure of the Heteroduplex Substrate." *Molecular Pharmacology* 71 (1): 83–91.
- Lima, Walt F., John B. Rose, Josh G. Nichols, Hongjiang Wu, Michael T. Migawa, Tadeusz K. Wyrzykiewicz, Guillermo Vasquez, Eric E. Swayze, and Stanley T. Crooke. 2007. "The Positional Influence of the Helical Geometry of the Heteroduplex Substrate on Human RNase H1 Catalysis." *Molecular Pharmacology* 71 (1): 73–82.
- Lindahl, T. 1993. "Instability and Decay of the Primary Structure of DNA." *Nature* 362 (6422): 709–15.
- Lindahl, T., and O. Karlström. 1973. "Heat-Induced Depyrimidination of Deoxyribonucleic Acid in Neutral Solution." *Biochemistry* 12 (25): 5151–54.
- Lindahl, T., and B. Nyberg. 1972. "Rate of Depurination of Native Deoxyribonucleic Acid." *Biochemistry* 11 (19): 3610–18.
- Liu, Ce, Yanyang Tu, Jun Yuan, Xinggang Mao, Shiming He, Liang Wang, Guoqiang Fu, Jianhai Zong, and Yongsheng Zhang. 2012. "Aberrant Expression of N-Methylpurine-DNA Glycosylase Influences Patient Survival in Malignant Gliomas." *Journal of Biomedicine & Biotechnology* 2012 (February): 760679.
- Lockhart, Arianna, Vanessa Borges Pires, Fabio Bento, Vanessa Kellner, Sarah Luke-Glaser, George Yakoub, Helle D. Ulrich, and Brian Luke. 2019. "RNase H1 and H2 Are Differentially Regulated to Process RNA-DNA Hybrids." *Cell Reports* 29 (9): 2890-2900.e5.
- Loeb, L. A., and B. D. Preston. 1986. "Mutagenesis by Apurinic/Apyrimidinic Sites." *Annual Review of Genetics* 20: 201–30.
- Loechler, E. L., C. L. Green, and J. M. Essigmann. 1984. "In Vivo Mutagenesis by O6-Methylguanine Built into a Unique Site in a Viral Genome." *Proceedings of the National Academy of Sciences of the United States of America* 81 (20): 6271–75.
- London, Robert E. 2015. "The Structural Basis of XRCC1-Mediated DNA Repair." *DNA Repair* 30 (June): 90–103.
- Loon, Barbara van, and Ulrich Hübscher. 2009. "An 8-Oxo-Guanine Repair Pathway Coordinated by MUTYH Glycosylase and DNA Polymerase Lambda." *Proceedings of the National Academy of Sciences of the United States of America* 106 (43): 18201–6.
- Lühnsdorf, Bork, Bernd Epe, and Andriy Khobta. 2014. "Excision of Uracil from Transcribed DNA Negatively Affects Gene Expression." *The Journal of Biological Chemistry* 289 (32): 22008–18.
- Lujan, Scott A., Jessica S. Williams, Anders R. Clausen, Alan B. Clark, and Thomas A. Kunkel. 2013. "Ribonucleotides Are Signals for Mismatch Repair of Leading-Strand Replication Errors." *Molecular Cell* 50 (3): 437–43.
- Lundin, Elin, Chenglin Wu, Albin Widmark, Mikaela Behm, Jens Hjerling-Leffler, Chammiran Daniel, Marie Öhman, and Mats Nilsson. 2020. "Spatiotemporal Mapping of RNA Editing

- in the Developing Mouse Brain Using in Situ Sequencing Reveals Regional and Cell-Type-Specific Regulation.” *BMC Biology* 18 (1): 6.
- MacFarlane, Amanda J., Donald D. Anderson, Per Flodby, Cheryll A. Perry, Robert H. Allen, Sally P. Stabler, and Patrick J. Stover. 2011. “Nuclear Localization of de Novo Thymidylate Biosynthesis Pathway Is Required to Prevent Uracil Accumulation in DNA.” *The Journal of Biological Chemistry* 286 (51): 44015–22.
- Mackenzie, Karen J., Paula Carroll, Laura Lettice, Žygimantė Tarnauskaitė, Kaalak Reddy, Flora Dix, Ailsa Revuelta, et al. 2016. “Ribonuclease H2 Mutations Induce a CGAS/STING-Dependent Innate Immune Response.” *The EMBO Journal* 35 (8): 831–44.
- Macosko, Evan Z., Anindita Basu, Rahul Satija, James Nemesh, Karthik Shekhar, Melissa Goldman, Itay Tirosh, et al. 2015. “Highly Parallel Genome-Wide Expression Profiling of Individual Cells Using Nanoliter Droplets.” *Cell* 161 (5): 1202–14.
- Mainz, Emilie R., D. Stephen Serafin, Tuong T. Nguyen, Teresa K. Tarrant, Christopher E. Sims, and Nancy L. Allbritton. 2016. “Single Cell Chemical Cytometry of Akt Activity in Rheumatoid Arthritis and Normal Fibroblast-like Synoviocytes in Response to Tumor Necrosis Factor α .” *Analytical Chemistry* 88 (15): 7786–92.
- Maiti, Atanu, and Alexander C. Drohat. 2011. “Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine: Potential Implications for Active Demethylation of CpG Sites.” *The Journal of Biological Chemistry* 286 (41): 35334–38.
- Malinge, Sébastien, Meghan Bliss-Moreau, Gina Kirsammer, Lauren Diebold, Timothy Chlon, Sandeep Gurbuxani, and John D. Crispino. 2012. “Increased Dosage of the Chromosome 21 Ortholog Dyrk1a Promotes Megakaryoblastic Leukemia in a Murine Model of Down Syndrome.” *The Journal of Clinical Investigation* 122 (3): 948–62.
- Marathe, Arvind, and Manju Bansal. 2010. “The 5-Methyl Group in Thymine Dynamically Influences the Structure of A-Tracts in DNA at the Local and Global Level.” *The Journal of Physical Chemistry. B* 114 (16): 5534–46.
- Markkanen, Enni, Julia Dorn, and Ulrich Hübscher. 2013. “MUTYH DNA Glycosylase: The Rationale for Removing Undamaged Bases from the DNA.” *Frontiers in Genetics* 4 (February): 18.
- Marsin, Stéphanie, Antonio E. Vidal, Marguerite Sossou, Josiane Ménissier-de Murcia, Florence Le Page, Serge Boiteux, Gilbert de Murcia, and J. Pablo Radicella. 2003. “Role of XRCC1 in the Coordination and Stimulation of Oxidative DNA Damage Repair Initiated by the DNA Glycosylase HOGG1.” *The Journal of Biological Chemistry* 278 (45): 44068–74.
- Martin, Sarah A., Afshan McCarthy, Louise J. Barber, Darren J. Burgess, Suzanne Parry, Christopher J. Lord, and Alan Ashworth. 2009. “Methotrexate Induces Oxidative DNA Damage and Is Selectively Lethal to Tumour Cells with Defects in the DNA Mismatch Repair Gene MSH2.” *EMBO Molecular Medicine* 1 (6–7): 323–37.
- Marullo, Rossella, Erica Werner, Natalya Degtyareva, Bryn Moore, Giuseppe Altavilla, Suresh S. Ramalingam, and Paul W. Doetsch. 2013. “Cisplatin Induces a Mitochondrial-ROS Response That Contributes to Cytotoxicity Depending on Mitochondrial Redox Status and Bioenergetic Functions.” *PloS One* 8 (11): e81162.
- Matsuda, Toshiro, Brian J. Vande Berg, Katarzyna Bebenek, Wendy P. Osheroff, Samuel H. Wilson, and Thomas A. Kunkel. 2003. “The Base Substitution Fidelity of DNA Polymerase Beta-Dependent Single Nucleotide Base Excision Repair.” *The Journal of Biological Chemistry* 278 (28): 25947–51.

- Matsumoto, Y., Q. M. Zhang, M. Takao, A. Yasui, and S. Yonei. 2001. "Escherichia Coli Nth and Human HNT1 DNA Glycosylases Are Involved in Removal of 8-Oxoguanine from 8-Oxoguanine/Guanine Mispairs in DNA." *Nucleic Acids Research* 29 (9): 1975–81.
- Maul, Robert W., Hyongi Chon, Kiran Sakhujia, Susana M. Cerritelli, Lina A. Gugliotti, Patricia J. Gearhart, and Robert J. Crouch. 2017. "R-Loop Depletion by Over-Expressed RNase H1 in Mouse B Cells Increases Activation-Induced Deaminase Access to the Transcribed Strand without Altering Frequency of Isotype Switching." *Journal of Molecular Biology* 429 (21): 3255–63.
- McElhinny, Stephanie A. Nick, Dinesh Kumar, Alan B. Clark, Danielle L. Watt, Brian E. Watts, Else-Britt Lundström, Erik Johansson, Andrei Chabes, and Thomas A. Kunkel. 2010. "Genome Instability Due to Ribonucleotide Incorporation into DNA." *Nature Chemical Biology* 6 (10): 774–81.
- McGinnis, Christopher S., David M. Patterson, Juliane Winkler, Daniel N. Conrad, Marco Y. Hein, Vasudha Srivastava, Jennifer L. Hu, et al. 2019. "MULTI-Seq: Sample Multiplexing for Single-Cell RNA Sequencing Using Lipid-Tagged Indices." *Nature Methods* 16 (7): 619–26.
- McIlwain, David R., Thorsten Berger, and Tak W. Mak. 2013. "Caspase Functions in Cell Death and Disease." *Cold Spring Harbor Perspectives in Biology* 5 (4): a008656.
- McKenna, P. G., and I. Hickey. 1981. "UV Sensitivity in Thymidine Kinase Deficient Mouse Erythroleukaemia Cells." *Cell Biology International Reports* 5 (6): 555–61.
- McKenna, P. G., V. J. McKelvey, and T. L. Frew. 1988. "Sensitivity to Cell Killing and the Induction of Cytogenetic Damage Following Gamma Irradiation in Wild-Type and Thymidine Kinase-Deficient Friend Mouse Erythroleukaemia Cells." *Mutation Research* 200 (1–2): 231–42.
- Meikrantz, W., M. A. Bergom, A. Memisoglu, and L. Samson. 1998. "O6-Alkylguanine DNA Lesions Trigger Apoptosis." *Carcinogenesis* 19 (2): 369–72.
- Melvin, Adam T., Gregory S. Woss, Jessica H. Park, Marcey L. Waters, and Nancy L. Allbritton. 2013. "Measuring Activity in the Ubiquitin-Proteasome System: From Large Scale Discoveries to Single Cells Analysis." *Cell Biochemistry and Biophysics* 67 (1): 75–89.
- Mendez, Pedro, Miquel Taron, Teresa Moran, Marco A. Fernandez, Gerard Requena, and Rafael Rosell. 2011. "A Modified Host-Cell Reactivation Assay to Quantify DNA Repair Capacity in Cryopreserved Peripheral Lymphocytes." *DNA Repair* 10 (6): 603–10.
- Métivier, Raphaël, Rozenn Gallais, Christophe Tiffocche, Christine Le Péron, Renata Z. Jurkowska, Richard P. Carmouche, David Ibberson, et al. 2008. "Cyclical DNA Methylation of a Transcriptionally Active Promoter." *Nature* 452 (7183): 45–50.
- Mikhed, Yuliya, Agnes Görlach, Ulla G. Knaus, and Andreas Daiber. 2015. "Redox Regulation of Genome Stability by Effects on Gene Expression, Epigenetic Pathways and DNA Damage/Repair." *Redox Biology* 5 (August): 275–89.
- Millau, J-F, A-L Raffin, S. Caillat, C. Claudet, G. Arras, N. Ugolin, T. Douki, et al. 2008. "A Microarray to Measure Repair of Damaged Plasmids by Cell Lysates." *Lab on a Chip* 8 (10): 1713–22.
- Mimitou, Eleni P., Anthony Cheng, Antonino Montalbano, Stephanie Hao, Marlon Stoeckius, Mateusz Legut, Timothy Roush, et al. 2019. "Multiplexed Detection of Proteins, Transcriptomes, Clonotypes and CRISPR Perturbations in Single Cells." *Nature Methods* 16 (5): 409–12.

- Mjelle, Robin, Siv Anita Hegre, Per Arne Aas, Geir Slupphaug, Finn Drabløs, Pål Saetrom, and Hans E. Krokan. 2015. "Cell Cycle Regulation of Human DNA Repair and Chromatin Remodeling Genes." *DNA Repair* 30 (June): 53–67.
- Moffitt, Jeffrey R., Dhananjay Bambah-Mukku, Stephen W. Eichhorn, Eric Vaughn, Karthik Shekhar, Julio D. Perez, Nimrod D. Rubinstein, et al. 2018. "Molecular, Spatial, and Functional Single-Cell Profiling of the Hypothalamic Preoptic Region." *Science* 362 (6416). <https://doi.org/10.1126/science.aau5324>.
- Moffitt, Jeffrey R., Junjie Hao, Guiping Wang, Kok Hao Chen, Hazen P. Babcock, and Xiaowei Zhuang. 2016. "High-Throughput Single-Cell Gene-Expression Profiling with Multiplexed Error-Robust Fluorescence in Situ Hybridization." *Proceedings of the National Academy of Sciences of the United States of America* 113 (39): 11046–51.
- Mohamed, Mona Mostafa, and Bonnie F. Sloane. 2006. "Cysteine Cathepsins: Multifunctional Enzymes in Cancer." *Nature Reviews. Cancer* 6 (10): 764–75.
- Moignard, Victoria, Steven Woodhouse, Laleh Haghverdi, Andrew J. Lilly, Yosuke Tanaka, Adam C. Wilkinson, Florian Buettner, et al. 2015. "Decoding the Regulatory Network of Early Blood Development from Single-Cell Gene Expression Measurements." *Nature Biotechnology* 33 (3): 269–76.
- Molinie, Benoit, and Cosmas C. Giallourakis. 2017. "Genome-Wide Location Analyses of N6-Methyladenosine Modifications (M6A-Seq)." *Methods in Molecular Biology* 1562: 45–53.
- Mongia, Aanchal, Debarka Sengupta, and Angshul Majumdar. 2019. "McImpute: Matrix Completion Based Imputation for Single Cell RNA-Seq Data." *Frontiers in Genetics* 10 (January): 9.
- Morawiec, Zbigniew, Katarzyna Janik, Michal Kowalski, Tomasz Stetkiewicz, Jerzy Szaflik, Alina Morawiec-Bajda, Anna Sobczuk, and Janusz Blasiak. 2008. "DNA Damage and Repair in Children with Down's Syndrome." *Mutation Research* 637 (1–2): 118–23.
- Morita, Yoko, Toshihiro Shibutani, Nozomi Nakanishi, Kazuko Nishikura, Shigenori Iwai, and Isao Kuraoka. 2013. "Human Endonuclease V Is a Ribonuclease Specific for Inosine-Containing RNA." *Nature Communications* 4: 2273.
- Morreall, Jordan, Kristin Limpose, Clayton Sheppard, Yoke Wah Kow, Erica Werner, and Paul W. Doetsch. 2015. "Inactivation of a Common OGG1 Variant by TNF-Alpha in Mammalian Cells." *DNA Repair* 26 (February): 15–22.
- Morris, Samantha A. 2019. "The Evolving Concept of Cell Identity in the Single Cell Era." *Development* 146 (12). <https://doi.org/10.1242/dev.169748>.
- Murante, R. S., L. A. Henricksen, and R. A. Bambara. 1998. "Junction Ribonuclease: An Activity in Okazaki Fragment Processing." *Proceedings of the National Academy of Sciences of the United States of America* 95 (5): 2244–49.
- Nagel, Zachary D., Andrew A. Beharry, Patrizia Mazzucato, Gaspar J. Kitange, Jann N. Sarkaria, Eric T. Kool, and Leona D. Samson. 2019. "Fluorescent Reporter Assays Provide Direct, Accurate, Quantitative Measurements of MGMT Status in Human Cells." *PloS One* 14 (2): e0208341.
- Nagel, Zachary D., Isaac A. Chaim, and Leona D. Samson. 2014. "Inter-Individual Variation in DNA Repair Capacity: A Need for Multi-Pathway Functional Assays to Promote Translational DNA Repair Research." *DNA Repair* 19 (July): 199–213.
- Nagel, Zachary D., Gaspar J. Kitange, Shiv K. Gupta, Brian A. Joughin, Isaac A. Chaim, Patrizia Mazzucato, Douglas A. Lauffenburger, Jann N. Sarkaria, and Leona D. Samson. 2016.

- “DNA Repair Capacity in Multiple Pathways Predicts Chemoresistance in Glioblastoma Multiforme.” *Cancer Research* 77 (1): 198–206.
- . 2017. “DNA Repair Capacity in Multiple Pathways Predicts Chemoresistance in Glioblastoma Multiforme.” *Cancer Research* 77 (1): 198–206.
- Nagel, Zachary D., Carrie M. Margulies, Isaac A. Chaim, Siobhan K. McRee, Patrizia Mazzucato, Anwaar Ahmad, Ryan P. Abo, Vincent L. Butty, Anthony L. Forget, and Leona D. Samson. 2014. “Multiplexed DNA Repair Assays for Multiple Lesions and Multiple Doses via Transcription Inhibition and Transcriptional Mutagenesis.” *Proceedings of the National Academy of Sciences* 111 (18): E1823–32.
- Nair, J., A. Barbin, I. Velic, and H. Bartsch. 1999. “Etheno DNA-Base Adducts from Endogenous Reactive Species.” *Mutation Research* 424 (1–2): 59–69.
- Nakabeppu, Yusaku. 2014. “Cellular Levels of 8-Oxoguanine in Either DNA or the Nucleotide Pool Play Pivotal Roles in Carcinogenesis and Survival of Cancer Cells.” *International Journal of Molecular Sciences* 15 (7): 12543–57.
- Nakamura, J., and J. A. Swenberg. 1999. “Endogenous Apurinic/Apyrimidinic Sites in Genomic DNA of Mammalian Tissues.” *Cancer Research* 59 (11): 2522–26.
- Nakamura, Tomohumi, Kouichi Murakami, Haruto Tada, Yoshihiko Uehara, Jumpei Nogami, Kazumitsu Maehara, Yasuyuki Ohkawa, et al. 2017. “Thymine DNA Glycosylase Modulates DNA Damage Response and Gene Expression by Base Excision Repair-Dependent and Independent Mechanisms.” *Genes to Cells: Devoted to Molecular & Cellular Mechanisms* 22 (4): 392–405.
- Narvaiza, Iñigo, Sébastien Landry, and Matthew D. Weitzman. 2012. “APOBEC3 Proteins and Genomic Stability: The High Cost of a Good Defense.” *Cell Cycle* 11 (1): 33–38.
- Necchi, Daniela, Antonella Pinto, Micol Tillhon, Ilaria Dutto, Melania Maria Serafini, Cristina Lanni, Stefano Govoni, Marco Racchi, and Ennio Proserpi. 2015. “Defective DNA Repair and Increased Chromatin Binding of DNA Repair Factors in Down Syndrome Fibroblasts.” *Mutation Research* 780 (October): 15–23.
- Neeley, William L., and John M. Essigmann. 2006. “Mechanisms of Formation, Genotoxicity, and Mutation of Guanine Oxidation Products.” *Chemical Research in Toxicology* 19 (4): 491–505.
- Neri, Monica, Daniele Milazzo, Donatella Ugolini, Mirta Milic, Alessandra Campolongo, Patrizio Pasqualetti, and Stefano Bonassi. 2015. “Worldwide Interest in the Comet Assay: A Bibliometric Study.” *Mutagenesis* 30 (1): 155–63.
- Nick McElhinny, Stephanie A., Brian E. Watts, Dinesh Kumar, Danielle L. Watt, Else-Britt Lundström, Peter M. J. Burgers, Erik Johansson, Andrei Chabes, and Thomas A. Kunkel. 2010. “Abundant Ribonucleotide Incorporation into DNA by Yeast Replicative Polymerases.” *Proceedings of the National Academy of Sciences of the United States of America* 107 (11): 4949–54.
- Nilsen, H., M. Otterlei, T. Haug, K. Solum, T. A. Nagelhus, F. Skorpen, and H. E. Krokan. 1997. “Nuclear and Mitochondrial Uracil-DNA Glycosylases Are Generated by Alternative Splicing and Transcription from Different Positions in the UNG Gene.” *Nucleic Acids Research* 25 (4): 750–55.
- Nilsen, H., I. Rosewell, P. Robins, C. F. Skjelbred, S. Andersen, G. Slupphaug, G. Daly, H. E. Krokan, T. Lindahl, and D. E. Barnes. 2000. “Uracil-DNA Glycosylase (UNG)-Deficient Mice Reveal a Primary Role of the Enzyme during DNA Replication.” *Molecular Cell* 5 (6): 1059–65.

- Nilsen, H., S. P. Yazdankhah, I. Eftedal, and H. E. Krokan. 1995. "Sequence Specificity for Removal of Uracil from U.A Pairs and U.G Mismatches by Uracil-DNA Glycosylase from *Escherichia Coli*, and Correlation with Mutational Hotspots." *FEBS Letters* 362 (2): 205–9.
- Nishant, Koodali T., Nadia D. Singh, and Eric Alani. 2009. "Genomic Mutation Rates: What High-Throughput Methods Can Tell Us." *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* 31 (9): 912–20.
- Nordlund, Pär, and Peter Reichard. 2006. "Ribonucleotide Reductases." *Annual Review of Biochemistry* 75: 681–706.
- Nowotny, Marcin, Sergei A. Gaidamakov, Rodolfo Ghirlando, Susana M. Cerritelli, Robert J. Crouch, and Wei Yang. 2007. "Structure of Human RNase H1 Complexed with an RNA/DNA Hybrid: Insight into HIV Reverse Transcription." *Molecular Cell* 28 (2): 264–76.
- O'Brien, Patrick J., and Tom Ellenberger. 2003. "Human Alkyladenine DNA Glycosylase Uses Acid-Base Catalysis for Selective Excision of Damaged Purines." *Biochemistry* 42 (42): 12418–29.
- . 2004. "Dissecting the Broad Substrate Specificity of Human 3-Methyladenine-DNA Glycosylase." *The Journal of Biological Chemistry* 279 (11): 9750–57.
- Obtulowicz, Tomasz, Maja Swoboda, Elzbieta Speina, Daniel Gackowski, Rafal Rozalski, Agnieszka Siomek, Justyna Janik, et al. 2010. "Oxidative Stress and 8-Oxoguanine Repair Are Enhanced in Colon Adenoma and Carcinoma Patients." *Mutagenesis* 25 (5): 463–71.
- Ohno, Mizuki, Kunihiro Sakumi, Ryutaro Fukumura, Masato Furuichi, Yuki Iwasaki, Masaaki Hokama, Toshimichi Ikemura, Teruhisa Tsuzuki, Yoichi Gondo, and Yusaku Nakabeppu. 2014. "8-Oxoguanine Causes Spontaneous de Novo Germline Mutations in Mice." *Scientific Reports* 4 (April): 4689.
- Oka, S., J. Leon, D. Tsuchimoto, K. Sakumi, and Y. Nakabeppu. 2014. "MUTYH, an Adenine DNA Glycosylase, Mediates P53 Tumor Suppression via PARP-Dependent Cell Death." *Oncogenesis* 3 (October): e121.
- Oka, Sugako, and Yusaku Nakabeppu. 2011. "DNA Glycosylase Encoded by MUTYH Functions as a Molecular Switch for Programmed Cell Death under Oxidative Stress to Suppress Tumorigenesis." *Cancer Science* 102 (4): 677–82.
- Oka, Sugako, Mizuki Ohno, Daisuke Tsuchimoto, Kunihiro Sakumi, Masato Furuichi, and Yusaku Nakabeppu. 2008. "Two Distinct Pathways of Cell Death Triggered by Oxidative Damage to Nuclear and Mitochondrial DNAs." *The EMBO Journal* 27 (2): 421–32.
- Olinski, Ryszard, Marek Jurgowiak, and Tomasz Zaremba. 2010. "Uracil in DNA--Its Biological Significance." *Mutation Research* 705 (3): 239–45.
- Oliva, Claudia R., Susan E. Nozell, Anne Diers, Samuel G. McClugage 3rd, Jann N. Sarkaria, James M. Markert, Victor M. Darley-Usmar, et al. 2010. "Acquisition of Temozolomide Chemoresistance in Gliomas Leads to Remodeling of Mitochondrial Electron Transport Chain." *The Journal of Biological Chemistry* 285 (51): 39759–67.
- Olive, P. L., J. P. Banáth, and R. E. Durand. 1990. "Heterogeneity in Radiation-Induced DNA Damage and Repair in Tumor and Normal Cells Measured Using the 'Comet' Assay." *Radiation Research* 122 (1): 86–94.
- Olive, P. L., C. M. Vikse, and J. P. Banath. 1996. "Use of the Comet Assay to Identify Cells Sensitive to Tirapazamine in Multicell Spheroids and Tumors in Mice." *Cancer Research* 56 (19): 4460–63.

- Orthwein, Alexandre, Amélie Fradet-Turcotte, Sylvie M. Noordermeer, Marella D. Canny, Catherine M. Brun, Jonathan Strecker, Cristina Escribano-Diaz, and Daniel Durocher. 2014. "Mitosis Inhibits DNA Double-Strand Break Repair to Guard against Telomere Fusions." *Science* 344 (6180): 189–93.
- Osheroff, Wendy P., Hai Kwan Jung, William A. Beard, Samuel H. Wilson, and Thomas A. Kunkel. 1999. "The Fidelity of DNA Polymerase β during Distributive and Processive DNA Synthesis." *The Journal of Biological Chemistry* 274 (6): 3642–50.
- Osterod, M., S. Hollenbach, J. G. Hengstler, D. E. Barnes, T. Lindahl, and B. Epe. 2001. "Age-Related and Tissue-Specific Accumulation of Oxidative DNA Base Damage in 7,8-Dihydro-8-Oxoguanine-DNA Glycosylase (Ogg1) Deficient Mice." *Carcinogenesis* 22 (9): 1459–63.
- Ostling, O., and K. J. Johanson. 1984. "Microelectrophoretic Study of Radiation-Induced DNA Damages in Individual Mammalian Cells." *Biochemical and Biophysical Research Communications* 123 (1): 291–98.
- . 1987. "Bleomycin, in Contrast to Gamma Irradiation, Induces Extreme Variation of DNA Strand Breakage from Cell to Cell." *International Journal of Radiation Biology and Related Studies in Physics, Chemistry, and Medicine* 52 (5): 683–91.
- Otterlei, M., E. Warbrick, T. A. Nagelhus, T. Haug, G. Slupphaug, M. Akbari, P. A. Aas, K. Steinsbekk, O. Bakke, and H. E. Krokan. 1999. "Post-Replicative Base Excision Repair in Replication Foci." *The EMBO Journal* 18 (13): 3834–44.
- Paik, Johanna, Tod Duncan, Tomas Lindahl, and Barbara Sedgwick. 2005. "Sensitization of Human Carcinoma Cells to Alkylating Agents by Small Interfering RNA Suppression of 3-Alkyladenine-DNA Glycosylase." *Cancer Research* 65 (22): 10472–77.
- Pan, Lang, Bing Zhu, Wenjing Hao, Xianlu Zeng, Spiros A. Vlahopoulos, Tapas K. Hazra, Muralidhar L. Hegde, et al. 2016. "Oxidized Guanine Base Lesions Function in 8-Oxoguanine DNA Glycosylase-1-Mediated Epigenetic Regulation of Nuclear Factor KB-Driven Gene Expression." *The Journal of Biological Chemistry* 291 (49): 25553–66.
- Pang, Bo, Jose L. McFaline, Nicholas E. Burgis, Min Dong, Koli Taghizadeh, Matthew R. Sullivan, C. Eric Elmquist, Richard P. Cunningham, and Peter C. Dedon. 2012. "Defects in Purine Nucleotide Metabolism Lead to Substantial Incorporation of Xanthine and Hypoxanthine into DNA and RNA." *Proceedings of the National Academy of Sciences of the United States of America* 109 (7): 2319–24.
- Paone, A., M. Marani, A. Fiascarelli, S. Rinaldo, G. Giardina, R. Contestabile, A. Paiardini, and F. Cutruzzolà. 2014. "SHMT1 Knockdown Induces Apoptosis in Lung Cancer Cells by Causing Uracil Misincorporation." *Cell Death & Disease* 5 (November): e1525.
- Parker, Nicole R., Amanda L. Hudson, Peter Khong, Jonathon F. Parkinson, Trisha Dwight, Rowan J. Ikin, Ying Zhu, et al. 2016. "Intratumoral Heterogeneity Identified at the Epigenetic, Genetic and Transcriptional Level in Glioblastoma." *Scientific Reports* 6 (March): 22477.
- Parsels, L. A., J. D. Parsels, L. M. Wagner, T. L. Loney, E. H. Radany, and J. Maybaum. 1998. "Mechanism and Pharmacological Specificity of DUTPase-Mediated Protection from DNA Damage and Cytotoxicity in Human Tumor Cells." *Cancer Chemotherapy and Pharmacology* 42 (5): 357–62.
- Patterson, David, and Diane C. Cabelof. 2012. "Down Syndrome as a Model of DNA Polymerase Beta Haploinsufficiency and Accelerated Aging." *Mechanisms of Ageing and Development* 133 (4): 133–37.

- Pazin, Michael J. 2015. "Using the ENCODE Resource for Functional Annotation of Genetic Variants." *Cold Spring Harbor Protocols* 2015 (6): 522–36.
- Persano, L., F. Pistollato, E. Rampazzo, A. Della Puppa, S. Abbadi, C. Frasson, F. Volpin, S. Indraccolo, R. Scienza, and G. Basso. 2012. "BMP2 Sensitizes Glioblastoma Stem-like Cells to Temozolomide by Affecting HIF-1 α Stability and MGMT Expression." *Cell Death & Disease* 3 (October): e412.
- Picelli, Simone, Omid R. Faridani, Asa K. Björklund, Gösta Winberg, Sven Sagasser, and Rickard Sandberg. 2014. "Full-Length RNA-Seq from Single Cells Using Smart-Seq2." *Nature Protocols* 9 (1): 171–81.
- Pollyea, Daniel A., Brett M. Stevens, Courtney L. Jones, Amanda Winters, Shanshan Pei, Mohammad Minhajuddin, Angelo D'Alessandro, et al. 2018. "Venetoclax with Azacitidine Disrupts Energy Metabolism and Targets Leukemia Stem Cells in Patients with Acute Myeloid Leukemia." *Nature Medicine* 24 (12): 1859–66.
- Pons, Bénédicte, Anne-Sophie Belmont, Gwénaëlle Masson-Genteuil, Violaine Chapuis, Thierry Oddos, and Sylvie Sauvaigo. 2010a. "Age-Associated Modifications of Base Excision Repair Activities in Human Skin Fibroblast Extracts." *Mechanisms of Ageing and Development* 131 (11–12): 661–65.
- . 2010b. "Age-Associated Modifications of Base Excision Repair Activities in Human Skin Fibroblast Extracts." *Mechanisms of Ageing and Development* 131 (11–12): 661–65.
- Pouget, J-P, S. Frelon, J-L Ravanat, I. Testard, F. Odin, and J. Cadet. 2002. "Formation of Modified DNA Bases in Cells Exposed Either to Gamma Radiation or to High-LET Particles." *Radiation Research* 157 (5): 589–95.
- Pourquier, Philippe, Mary-Ann Bjornsti, and Yves Pommier. 1998. "Induction of Topoisomerase I Cleavage Complexes by the Vinyl Chloride Adduct 1,N 6-Ethenoadenine." *The Journal of Biological Chemistry* 273 (42): 27245–49.
- Prasad, Rajendra, Nadezhda Dyrkheeva, Jason Williams, and Samuel H. Wilson. 2015a. "Mammalian Base Excision Repair: Functional Partnership between PARP-1 and APE1 in AP-Site Repair." *PloS One* 10 (5): e0124269.
- . 2015b. "Mammalian Base Excision Repair: Functional Partnership between PARP-1 and APE1 in AP-Site Repair." *PloS One* 10 (5): e0124269.
- Proctor, Angela, Qunzhao Wang, David S. Lawrence, and Nancy L. Allbritton. 2012. "Metabolism of Peptide Reporters in Cell Lysates and Single Cells." *The Analyst* 137 (13): 3028–38.
- Qiu, Xiaojie, Andrew Hill, Jonathan Packer, Dejun Lin, Yi-An Ma, and Cole Trapnell. 2017. "Single-Cell mRNA Quantification and Differential Analysis with Census." *Nature Methods* 14 (3): 309–15.
- Qiu, Xiaojie, Qi Mao, Ying Tang, Li Wang, Raghav Chawla, Hannah A. Pliner, and Cole Trapnell. 2017. "Reversed Graph Embedding Resolves Complex Single-Cell Trajectories." *Nature Methods* 14 (10): 979–82.
- Quinlan, Aaron R., and Ira M. Hall. 2010a. "BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features." *Bioinformatics* 26 (6): 841–42.
- . 2010b. "BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features." *Bioinformatics* 26 (6): 841–42.
- Radany, E. H., K. J. Dornfeld, R. J. Sanderson, M. K. Savage, A. Majumdar, M. M. Seidman, and D. W. Mosbaugh. 2000. "Increased Spontaneous Mutation Frequency in Human Cells Expressing the Phage PBS2-Encoded Inhibitor of Uracil-DNA Glycosylase." *Mutation Research* 461 (1): 41–58.

- Rasmussen, Kasper Dindler, and Kristian Helin. 2016. "Role of TET Enzymes in DNA Methylation, Development, and Cancer." *Genes & Development* 30 (7): 733–50.
- Reardon, Joyce T., and Aziz Sancar. 2005. "Nucleotide Excision Repair." In *Progress in Nucleic Acid Research and Molecular Biology*, 79:183–235. Academic Press.
- Rebhandl, Stefan, Michael Huemer, Richard Greil, and Roland Geisberger. 2015. "AID/APOBEC Deaminases and Cancer." *Oncoscience* 2 (4): 320–33.
- Regnell, Christine Elisabeth, Gunn Annette Hildrestrand, Yngve Sejersted, Tirill Medin, Olve Moldestad, Veslemøy Rolseth, Silje Zandstra Krokeide, et al. 2012. "Hippocampal Adult Neurogenesis Is Maintained by Neil3-Dependent Repair of Oxidative DNA Lesions in Neural Progenitor Cells." *Cell Reports* 2 (3): 503–10.
- Reijns, Martin A. M., Doryen Bubeck, Lucien C. D. Gibson, Stephen C. Graham, George S. Baillie, E. Yvonne Jones, and Andrew P. Jackson. 2011. "The Structure of the Human RNase H2 Complex Defines Key Interaction Interfaces Relevant to Enzyme Function and Human Disease." *The Journal of Biological Chemistry* 286 (12): 10530–39.
- Reijns, Martin A. M., Björn Rabe, Rachel E. Rigby, Pleasantine Mill, Katy R. Astell, Laura A. Lettice, Shelagh Boyle, et al. 2012. "Enzymatic Removal of Ribonucleotides from DNA Is Essential for Mammalian Genome Integrity and Development." *Cell* 149 (5): 1008–22.
- Replogle, Joseph M., Thomas M. Norman, Albert Xu, Jeffrey A. Hussmann, Jin Chen, J. Zachery Cogan, Elliott J. Meer, et al. 2020. "Combinatorial Single-Cell CRISPR Screens by Direct Guide RNA Capture and Targeted Sequencing." *Nature Biotechnology*, March. <https://doi.org/10.1038/s41587-020-0470-y>.
- Revy, P., T. Muto, Y. Levy, F. Geissmann, A. Plebani, O. Sanal, N. Catalan, et al. 2000. "Activation-Induced Cytidine Deaminase (AID) Deficiency Causes the Autosomal Recessive Form of the Hyper-IgM Syndrome (HIGM2)." *Cell* 102 (5): 565–75.
- Reyffman, Paul A., James M. Walter, Nikita Joshi, Kishore R. Anekalla, Alexandra C. McQuattie-Pimentel, Stephen Chiu, Ramiro Fernandez, et al. 2019. "Single-Cell Transcriptomic Analysis of Human Lung Provides Insights into the Pathobiology of Pulmonary Fibrosis." *American Journal of Respiratory and Critical Care Medicine* 199 (12): 1517–36.
- Rice, Gillian, Teresa Patrick, Rekha Parmar, Claire F. Taylor, Alec Aeby, Jean Aicardi, Rafael Artuch, et al. 2007. "Clinical and Molecular Phenotype of Aicardi-Goutieres Syndrome." *American Journal of Human Genetics* 81 (4): 713–25.
- Richer, A. L., J. M. Cala, K. O'Brien, V. M. Carson, and L. J. Inge. 2017. "WEE1 Kinase Inhibitor AZD1775 Has Preclinical Efficacy in LKB1-Deficient Non-Small Cell Lung Cancer." *Cancer Research*. <https://cancerres.aacrjournals.org/content/77/17/4663.short>.
- Richer, Amanda L., Jacqueline M. Friel, Vashti M. Carson, Landon J. Inge, and Timothy G. Whitsett. 2015. "Genomic Profiling toward Precision Medicine in Non-Small Cell Lung Cancer: Getting beyond EGFR." *Pharmacogenomics and Personalized Medicine* 8 (February): 63–79.
- Richer, Amanda L., Kent A. Riemondy, Lakotah Hardie, and Jay R. Hesselberth. 2020a. "Simultaneous Measurement of Biochemical Phenotypes and Gene Expression in Single Cells." *Nucleic Acids Research*, April. <https://doi.org/10.1093/nar/gkaa240>.
- . 2020b. "Simultaneous Measurement of Biochemical Phenotypes and Gene Expression in Single Cells." *Nucleic Acids Research*, April. <https://doi.org/10.1093/nar/gkaa240>.
- Riemondy, Kent A., Austin E. Gillen, Emily A. White, Lori K. Bogren, Jay R. Hesselberth, and Sandra L. Martin. 2018. "Dynamic Temperature-Sensitive A-to-I RNA Editing in the Brain of a Heterothermic Mammal during Hibernation." *RNA* 24 (11): 1481–95.

- Rodrigues, Samuel G., Robert R. Stickels, Aleksandrina Goeva, Carly A. Martin, Evan Murray, Charles R. Vanderburg, Joshua Welch, Linlin M. Chen, Fei Chen, and Evan Z. Macosko. 2019. "Slide-Seq: A Scalable Technology for Measuring Genome-Wide Expression at High Spatial Resolution." *Science* 363 (6434): 1463–67.
- Rohman, Muhammad S., Yuichi Koga, Kazufumi Takano, Hyongi Chon, Robert J. Crouch, and Shigenori Kanaya. 2008. "Effect of the Disease-Causing Mutations Identified in Human Ribonuclease (RNase) H2 on the Activities and Stabilities of Yeast RNase H2 and Archaeal RNase HII." *The FEBS Journal* 275 (19): 4836–49.
- Rolseth, Veslemøy, Luisa Luna, Ann Karin Olsen, Rajikala Suganthan, Katja Scheffler, Christine G. Neurauter, Ying Esbensen, et al. 2017. "No Cancer Predisposition or Increased Spontaneous Mutation Frequencies in NEIL DNA Glycosylases-Deficient Mice." *Scientific Reports* 7 (1): 4384.
- Rood, Brian R., Huizhen Zhang, and Philip H. Cogen. 2004. "Intercellular Heterogeneity of Expression of the MGMT DNA Repair Gene in Pediatric Medulloblastoma." *Neuro-Oncology* 6 (3): 200–207.
- Roos, W. P., L. F. Z. Batista, S. C. Naumann, W. Wick, M. Weller, C. F. M. Menck, and B. Kaina. 2007. "Apoptosis in Malignant Glioma Cells Triggered by the Temozolomide-Induced DNA Lesion O6-Methylguanine." *Oncogene* 26 (2): 186–97.
- Rosenberg, Alexander B., Charles M. Roco, Richard A. Muscat, Anna Kuchina, Paul Sample, Zizhen Yao, Lucas T. Graybuck, et al. 2018. "Single-Cell Profiling of the Developing Mouse Brain and Spinal Cord with Split-Pool Barcoding." *Science* 360 (6385): 176–82.
- Ruggieri, V., E. Pin, M. T. Russo, F. Barone, P. Degan, M. Sanchez, M. Quaia, et al. 2013. "Loss of MUTYH Function in Human Cells Leads to Accumulation of Oxidative Damage and Genetic Instability." *Oncogene* 32 (38): 4500–4508.
- Russo, Maria Teresa, Monica Francesca Blasi, Federica Chiera, Paola Fortini, Paolo Degan, Peter Macpherson, Masato Furuichi, et al. 2004. "The Oxidized Deoxynucleoside Triphosphate Pool Is a Significant Contributor to Genetic Instability in Mismatch Repair-Deficient Cells." *Molecular and Cellular Biology* 24 (1): 465–74.
- Rychlik, Monika P., Hyongi Chon, Susana M. Cerritelli, Paulina Klimek, Robert J. Crouch, and Marcin Nowotny. 2010. "Crystal Structures of RNase H2 in Complex with Nucleic Acid Reveal the Mechanism of RNA-DNA Junction Recognition and Cleavage." *Molecular Cell* 40 (4): 658–70.
- Rydberg, B., and T. Lindahl. 1982. "Nonenzymatic Methylation of DNA by the Intracellular Methyl Group Donor S-Adenosyl-L-Methionine Is a Potentially Mutagenic Reaction." *The EMBO Journal* 1 (2): 211–16.
- Sade-Feldman, Moshe, Keren Yizhak, Stacey L. Bjorgaard, John P. Ray, Carl G. de Boer, Russell W. Jenkins, David J. Lieb, et al. 2018. "Defining T Cell States Associated with Response to Checkpoint Immunotherapy in Melanoma." *Cell* 175 (4): 998-1013.e20.
- Sakai, Yasunari, Masato Furuichi, Masayuki Takahashi, Masaki Mishima, Shigenori Iwai, Masahiro Shirakawa, and Yusaku Nakabeppu. 2002. "A Molecular Basis for the Selective Recognition of 2-Hydroxy-DATP and 8-Oxo-DGTP by Human MTH1." *The Journal of Biological Chemistry* 277 (10): 8579–87.
- Sakamoto, Katsumi, Yohei Tominaga, Kazumi Yamauchi, Yoshimichi Nakatsu, Kunihiko Sakumi, Kaoru Yoshiyama, Akinori Egashira, et al. 2007. "MUTYH-Null Mice Are Susceptible to Spontaneous and Oxidative Stress Induced Intestinal Tumorigenesis." *Cancer Research* 67 (14): 6599–6604.

- Sakumi, K., M. Furuichi, T. Tsuzuki, T. Kakuma, S. Kawabata, H. Maki, and M. Sekiguchi. 1993. "Cloning and Expression of cDNA for a Human Enzyme That Hydrolyzes 8-Oxo-DGTP, a Mutagenic Substrate for DNA Synthesis." *The Journal of Biological Chemistry* 268 (31): 23524–30.
- Sakumi, Kunihiko, Yohei Tominaga, Masato Furuichi, Ping Xu, Teruhisa Tsuzuki, Mutsuo Sekiguchi, and Yusaku Nakabeppu. 2003. "Ogg1 Knockout-Associated Lung Tumorigenesis and Its Suppression by Mth1 Gene Disruption." *Cancer Research* 63 (5): 902–5.
- Sancar, Aziz, Laura A. Lindsey-Boltz, Keziban Unsal-Kaçmaz, and Stuart Linn. 2004. "Molecular Mechanisms of Mammalian DNA Repair and the DNA Damage Checkpoints." *Annual Review of Biochemistry* 73: 39–85.
- Santos-Pereira, José M., and Andrés Aguilera. 2015. "R Loops: New Modulators of Genome Dynamics and Function." *Nature Reviews. Genetics* 16 (10): 583–97.
- Sattler, Ulrike, Philippe Frit, Bernard Salles, and Patrick Calsou. 2003. "Long-Patch DNA Repair Synthesis during Base Excision Repair in Mammalian Cells." *EMBO Reports* 4 (4): 363–67.
- Sauvaigo, S., M. Bonnet-Duquennoy, F. Odin, F. Hazane-Puch, N. Lachmann, F. Bonté, R. Kurfürst, and A. Favier. 2007. "DNA Repair Capacities of Cutaneous Fibroblasts: Effect of Sun Exposure, Age and Smoking on Response to an Acute Oxidative Stress." *The British Journal of Dermatology* 157 (1): 26–32.
- Sauvaigo, Sylvie, Sylvain Caillat, Francette Odin, Alex Nkengne, Christiane Bertin, and Thierry Oddos. 2010. "Effect of Aging on DNA Excision/Synthesis Repair Capacities of Human Skin Fibroblasts." *The Journal of Investigative Dermatology* 130 (6): 1739–41.
- Schier, Alexander F. 2020. "Single-Cell Biology: Beyond the Sum of Its Parts." *Nature Methods* 17 (1): 17–20.
- Schiesser, Stefan, Benjamin Hackner, Toni Pfaffeneder, Markus Müller, Christian Hagemeyer, Matthias Truss, and Thomas Carell. 2012. "Mechanism and Stem-Cell Activity of 5-Carboxycytosine Decarboxylation Determined by Isotope Tracing." *Angewandte Chemie* 51 (26): 6516–20.
- Schomacher, Lars, Dandan Han, Michael U. Musheev, Khelifa Arab, Sabine Kienhöfer, Annika von Seggern, and Christof Niehrs. 2016. "Neil DNA Glycosylases Promote Substrate Turnover by Tdg during DNA Demethylation." *Nature Structural & Molecular Biology* 23 (2): 116–24.
- Sczepanski, Jonathan T., Remus S. Wong, Jeffrey N. McKnight, Gregory D. Bowman, and Marc M. Greenberg. 2010. "Rapid DNA-Protein Cross-Linking and Strand Scission by an Abasic Site in a Nucleosome Core Particle." *Proceedings of the National Academy of Sciences of the United States of America* 107 (52): 22475–80.
- Sedgwick, Barbara, Paul A. Bates, Johanna Paik, Susan C. Jacobs, and Tomas Lindahl. 2007. "Repair of Alkylated DNA: Recent Advances." *DNA Repair* 6 (4): 429–42.
- Sekiguchi, J., and S. Shuman. 1997. "Site-Specific Ribonuclease Activity of Eukaryotic DNA Topoisomerase I." *Molecular Cell* 1 (1): 89–97.
- Selby, Christopher P., and Aziz Sancar. 2006. "A Cryptochrome/Photolyase Class of Enzymes with Single-Stranded DNA-Specific Photolyase Activity." *Proceedings of the National Academy of Sciences of the United States of America* 103 (47): 17696–700.

- Sevilya, Ziv, Yael Leitner-Dagan, Mila Pinchev, Ran Kremer, Dalia Elinger, Hedy S. Rennert, Edna Schechtman, et al. 2014. "Low Integrated DNA Repair Score and Lung Cancer Risk." *Cancer Prevention Research* 7 (4): 398–406.
- Shalek, Alex K., and Mikael Benson. 2017. "Single-Cell Analyses to Tailor Treatments." *Science Translational Medicine* 9 (408). <https://doi.org/10.1126/scitranslmed.aan4730>.
- Shapiro, R., and S. J. Shiuey. 1969. "Reaction of Nitrous Acid with Alkylaminopurines." *Biochimica et Biophysica Acta* 174 (1): 403–5.
- Shen, Jiang-Cheng, Edward J. Fox, Eun Hyun Ahn, and Lawrence A. Loeb. 2014. "A Rapid Assay for Measuring Nucleotide Excision Repair by Oligonucleotide Retrieval." *Scientific Reports* 4 (May): 4894.
- Shen, Z., W. Wu, and S. L. Hazen. 2000. "Activated Leukocytes Oxidatively Damage DNA, RNA, and the Nucleotide Pool through Halide-Dependent Formation of Hydroxyl Radical." *Biochemistry* 39 (18): 5474–82.
- Shibutani, S., M. Takeshita, and A. P. Grollman. 1991. "Insertion of Specific Bases during DNA Synthesis Past the Oxidation-Damaged Base 8-OxodG." *Nature* 349 (6308): 431–34.
- Shin, Dongju, Wookjae Lee, Ji Hyun Lee, and Duhee Bang. 2019. "Multiplexed Single-Cell RNA-Seq via Transient Barcoding for Simultaneous Expression Profiling of Various Drug Perturbations." *Science Advances* 5 (5): eaav2249.
- Sidorenko, Viktoriya S., Georgy A. Nevinsky, and Dmitry O. Zharkov. 2008. "Specificity of Stimulation of Human 8-Oxoguanine-DNA Glycosylase by AP Endonuclease." *Biochemical and Biophysical Research Communications* 368 (1): 175–79.
- Siebenlist, U., and W. Gilbert. 1980. "Contacts between Escherichia Coli RNA Polymerase and an Early Promoter of Phage T7." *Proceedings of the National Academy of Sciences of the United States of America* 77 (1): 122–26.
- Sieber, Oliver M., Lara Lipton, Michael Crabtree, Karl Heinimann, Paulo Fidalgo, Robin K. S. Phillips, Marie-Luise Bisgaard, et al. 2003. "Multiple Colorectal Adenomas, Classic Adenomatous Polyposis, and Germ-Line Mutations in MYH." *The New England Journal of Medicine* 348 (9): 791–99.
- Singh, Amanpreet, and Yong-Jie Xu. 2016. "The Cell Killing Mechanisms of Hydroxyurea." *Genes* 7 (11). <https://doi.org/10.3390/genes7110099>.
- Slupphaug, G., I. Eftedal, B. Kavli, S. Bharati, N. M. Helle, T. Haug, D. W. Levine, and H. E. Krokan. 1995. "Properties of a Recombinant Human Uracil-DNA Glycosylase from the UNG Gene and Evidence That UNG Encodes the Major Uracil-DNA Glycosylase." *Biochemistry* 34 (1): 128–38.
- Slupphaug, G., C. D. Mol, B. Kavli, A. S. Arvai, H. E. Krokan, and J. A. Tainer. 1996. "A Nucleotide-Flipping Mechanism from the Structure of Human Uracil-DNA Glycosylase Bound to DNA." *Nature* 384 (6604): 87–92.
- Slyskova, Jana, Sabine A. S. Langie, Andrew R. Collins, and Pavel Vodicka. 2014. "Functional Evaluation of DNA Repair in Human Biopsies and Their Relation to Other Cellular Biomarkers." *Frontiers in Genetics* 5 (May): 116.
- Slyskova, Jana, Yolanda Lorenzo, Anette Karlsen, Monica H. Carlsen, Vendula Novosadova, Rune Blomhoff, Pavel Vodicka, and Andrew R. Collins. 2014. "Both Genetic and Dietary Factors Underlie Individual Differences in DNA Damage Levels and DNA Repair Capacity." *DNA Repair* 16 (April): 66–73.

- Smith, Tom, Andreas Heger, and Ian Sudbery. 2017a. “UMI-Tools: Modeling Sequencing Errors in Unique Molecular Identifiers to Improve Quantification Accuracy.” *Genome Research* 27 (3): 491–99.
- . 2017b. “UMI-Tools: Modeling Sequencing Errors in Unique Molecular Identifiers to Improve Quantification Accuracy.” *Genome Research* 27 (3): 491–99.
- Song, Shanshan, Guichun Xing, Lin Yuan, Jian Wang, Shan Wang, Yuxin Yin, Chunyan Tian, Fuchu He, and Lingqiang Zhang. 2012. “N-Methylpurine DNA Glycosylase Inhibits P53-Mediated Cell Cycle Arrest and Coordinates with P53 to Determine Sensitivity to Alkylating Agents.” *Cell Research* 22 (8): 1285–1303.
- Sparks, Justin L., and Peter M. Burgers. 2015. “Error-Free and Mutagenic Processing of Topoisomerase 1-Provoked Damage at Genomic Ribonucleotides.” *The EMBO Journal* 34 (9): 1259–69.
- Sparks, Justin L., Hyongi Chon, Susana M. Cerritelli, Thomas A. Kunkel, Erik Johansson, Robert J. Crouch, and Peter M. Burgers. 2012. “RNase H2-Initiated Ribonucleotide Excision Repair.” *Molecular Cell* 47 (6): 980–86.
- Spencer, J. P., A. Jenner, K. Chimel, O. I. Aruoma, C. E. Cross, R. Wu, and B. Halliwell. 1995. “DNA Damage in Human Respiratory Tract Epithelial Cells: Damage by Gas Phase Cigarette Smoke Apparently Involves Attack by Reactive Nitrogen Species in Addition to Oxygen Radicals.” *FEBS Letters* 375 (3): 179–82.
- Spencer, J. P., M. Whiteman, A. Jenner, and B. Halliwell. 2000. “Nitrite-Induced Deamination and Hypochlorite-Induced Oxidation of DNA in Intact Human Respiratory Tract Epithelial Cells.” *Free Radical Biology & Medicine* 28 (7): 1039–50.
- Srivatsan, Sanjay R., José L. McFaline-Figueroa, Vijay Ramani, Lauren Saunders, Junyue Cao, Jonathan Packer, Hannah A. Pliner, et al. 2020. “Massively Multiplex Chemical Transcriptomics at Single-Cell Resolution.” *Science* 367 (6473): 45–51.
- Ståhl, Patrik L., Fredrik Salmén, Sanja Vickovic, Anna Lundmark, José Fernández Navarro, Jens Magnusson, Stefania Giacomello, et al. 2016. “Visualization and Analysis of Gene Expression in Tissue Sections by Spatial Transcriptomics.” *Science* 353 (6294): 78–82.
- Steiner, Marie E., and William G. Woods. 1982. “Normal Formation and Repair of γ -Radiation-Induced Single and Double Strand DNA Breaks in Down Syndrome Fibroblasts.” *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 95 (2): 515–23.
- Stoeckius, Marlon, Christoph Hafemeister, William Stephenson, Brian Houck-Loomis, Pratip K. Chattopadhyay, Harold Swerdlow, Rahul Satija, and Peter Smibert. 2017. “Simultaneous Epitope and Transcriptome Measurement in Single Cells.” *Nature Methods* 14 (9): 865–68.
- Stoeckius, Marlon, Shiwei Zheng, Brian Houck-Loomis, Stephanie Hao, Bertrand Z. Yeung, William M. Mauck 3rd, Peter Smibert, and Rahul Satija. 2018. “Cell Hashing with Barcoded Antibodies Enables Multiplexing and Doublet Detection for Single Cell Genomics.” *Genome Biology* 19 (1): 224.
- Strauss, B., and T. Hill. 1970. “The Intermediate in the Degradation of DNA Alkylated with a Monofunctional Alkylating Agent.” *Biochimica et Biophysica Acta* 213 (1): 14–25.
- Stuart, Tim, Andrew Butler, Paul Hoffman, Christoph Hafemeister, Efthymia Papalexi, William M. Mauck 3rd, Yuhao Hao, Marlon Stoeckius, Peter Smibert, and Rahul Satija. 2019a. “Comprehensive Integration of Single-Cell Data.” *Cell* 177 (7): 1888-1902.e21.
- . 2019b. “Comprehensive Integration of Single-Cell Data.” *Cell* 177 (7): 1888-1902.e21.
- . 2019c. “Comprehensive Integration of Single-Cell Data.” *Cell* 177 (7): 1888-1902.e21.

- Stuart, Tim, and Rahul Satija. 2019a. "Integrative Single-Cell Analysis." *Nature Reviews. Genetics* 20 (5): 257–72.
- . 2019b. "Integrative Single-Cell Analysis." *Nature Reviews. Genetics* 20 (5): 257–72.
- Studebaker, A. W., W. P. Lafuse, R. Kloesel, and M. V. Williams. 2005. "Modulation of Human DUTPase Using Small Interfering RNA." *Biochemical and Biophysical Research Communications* 327 (1): 306–10.
- Sullivan, Kelly D., Donald Evans, Ahwan Pandey, Thomas H. Hraha, Keith P. Smith, Neil Markham, Angela L. Rachubinski, et al. 2017. "Trisomy 21 Causes Changes in the Circulating Proteome Indicative of Chronic Autoinflammation." *Scientific Reports* 7 (1): 14818.
- Sullivan, Kelly D., Hannah C. Lewis, Amanda A. Hill, Ahwan Pandey, Leisa P. Jackson, Joseph M. Cabral, Keith P. Smith, et al. 2016. "Trisomy 21 Consistently Activates the Interferon Response." *ELife* 5 (July). <https://doi.org/10.7554/eLife.16220>.
- Suzuki, T., Y. Matsumura, H. Ide, K. Kanaori, K. Tajima, and K. Makino. 1997. "Deglycosylation Susceptibility and Base-Pairing Stability of 2'-Deoxyoxanosine in Oligodeoxynucleotide." *Biochemistry* 36 (26): 8013–19.
- Swenberg, James A., Kun Lu, Benjamin C. Moeller, Lina Gao, Patricia B. Upton, Jun Nakamura, and Thomas B. Starr. 2011. "Endogenous versus Exogenous DNA Adducts: Their Role in Carcinogenesis, Epidemiology, and Risk Assessment." *Toxicological Sciences: An Official Journal of the Society of Toxicology* 120 Suppl 1 (March): S130-45.
- Tajiri, T., H. Maki, and M. Sekiguchi. 1995. "Functional Cooperation of MutT, MutM and MutY Proteins in Preventing Mutations Caused by Spontaneous Oxidation of Guanine Nucleotide in Escherichia Coli." *Mutation Research* 336 (3): 257–67.
- Tentori, Lucio, and Grazia Graziani. 2009. "Recent Approaches to Improve the Antitumor Efficacy of Temozolomide." *Current Medicinal Chemistry* 16 (2): 245–57.
- Terzidis, Michael A., Carla Ferreri, and Chryssostomos Chatgililoglu. 2015. "Radiation-Induced Formation of Purine Lesions in Single and Double Stranded DNA: Revised Quantification." *Frontiers in Chemistry* 3 (March): 18.
- Theruvathu, Jacob A., Pawel Jaruga, Miral Dizdaroglu, and P. J. Brooks. 2007. "The Oxidatively Induced DNA Lesions 8,5'-Cyclo-2'-Deoxyadenosine and 8-Hydroxy-2'-Deoxyadenosine Are Strongly Resistant to Acid-Induced Hydrolysis of the Glycosidic Bond." *Mechanisms of Ageing and Development* 128 (9): 494–502.
- Thiviyanathan, Varatharasa, Anoma Somasunderam, David E. Volk, and David G. Gorenstein. 2005. "5-Hydroxyuracil Can Form Stable Base Pairs with All Four Bases in a DNA Duplex." *Chemical Communications*, no. 3 (January): 400–402.
- Thul, Peter J., Lovisa Åkesson, Mikaela Wiking, Diana Mahdessian, Aikaterini Geladaki, Hammou Ait Blal, Tove Alm, et al. 2017. "A Subcellular Map of the Human Proteome." *Science* 356 (6340). <https://doi.org/10.1126/science.aal3321>.
- Tinkelenberg, Beverly A., Michael J. Hansbury, and Robert D. Ladner. 2002. "DUTPase and Uracil-DNA Glycosylase Are Central Modulators of Antifolate Toxicity in *Saccharomyces Cerevisiae*." *Cancer Research* 62 (17): 4909–15.
- Torgovnick, Alessandro, and Björn Schumacher. 2015. "DNA Repair Mechanisms in Cancer Development and Therapy." *Frontiers in Genetics* 6 (April): 157.
- Toyokuni, S., T. Mori, and M. Dizdaroglu. 1994. "DNA Base Modifications in Renal Chromatin of Wistar Rats Treated with a Renal Carcinogen, Ferric Nitrilotriacetate." *International Journal of Cancer. Journal International Du Cancer* 57 (1): 123–28.

- Tran, Hoa Thi Nhu, Kok Siong Ang, Marion Chevrier, Xiaomeng Zhang, Nicole Yee Shin Lee, Michelle Goh, and Jinmiao Chen. 2020. "A Benchmark of Batch-Effect Correction Methods for Single-Cell RNA Sequencing Data." *Genome Biology* 21 (1): 12.
- Trapnell, Cole, Davide Cacchiarelli, Jonna Grimsby, Prapti Pokharel, Shuqiang Li, Michael Morse, Niall J. Lennon, Kenneth J. Livak, Tarjei S. Mikkelsen, and John L. Rinn. 2014. "The Dynamics and Regulators of Cell Fate Decisions Are Revealed by Pseudotemporal Ordering of Single Cells." *Nature Biotechnology* 32 (4): 381–86.
- Traut, T. W. 1994. "Physiological Concentrations of Purines and Pyrimidines." *Molecular and Cellular Biochemistry* 140 (1): 1–22.
- Trivedi, Ram N., Karen H. Almeida, Jamie L. Fornsgaglio, Sandra Schamus, and Robert W. Sobol. 2005. "The Role of Base Excision Repair in the Sensitivity and Resistance to Temozolomide-Mediated Cell Death." *Cancer Research* 65 (14): 6394–6400.
- Trzeciak, Andrzej R., Janice Barnes, and Michele K. Evans. 2008. "A Modified Alkaline Comet Assay for Measuring DNA Repair Capacity in Human Populations." *Radiation Research* 169 (1): 110–21.
- Uehara, Ryo, Susana M. Cerritelli, Naushaba Hasin, Kiran Sakhuja, Mariya London, Jaime Iranzo, Hyongi Chon, Alexander Grinberg, and Robert J. Crouch. 2018. "Two RNase H2 Mutants with Differential RNMP Processing Activity Reveal a Threshold of Ribonucleotide Tolerance for Embryonic Development." *Cell Reports* 25 (5): 1135-1145.e5.
- Uhlén, Mathias, Linn Fagerberg, Björn M. Hallström, Cecilia Lindskog, Per Oksvold, Adil Mardinoglu, Åsa Sivertsson, et al. 2015. "Proteomics. Tissue-Based Map of the Human Proteome." *Science* 347 (6220): 1260419.
- Uno, Miyuki, Sueli Mieko Oba-Shinjo, Anamaria Aranha Camargo, Ricardo Pereira Moura, Paulo Henrique de Aguiar, Hector Navarro Cabrera, Marcos Begnami, Sérgio Rosemberg, Manoel Jacobsen Teixeira, and Suely Kazue Nagahashi Marie. 2011. "Correlation of MGMT Promoter Methylation Status with Gene and Protein Expression Levels in Glioblastoma." *Clinics* 66 (10): 1747–55.
- Uphoff, Stephan. 2018. "Real-Time Dynamics of Mutagenesis Reveal the Chronology of DNA Repair and Damage Tolerance Responses in Single Cells." *Proceedings of the National Academy of Sciences of the United States of America* 115 (28): E6516–25.
- Uphoff, Stephan, Nathan D. Lord, Burak Okumus, Laurent Potvin-Trottier, David J. Sherratt, and Johan Paulsson. 2016. "Stochastic Activation of a DNA Damage Response Causes Cell-to-Cell Mutation Rate Variation." *Science* 351 (6277): 1094–97.
- Verri, A., P. Mazzeo, G. Biamonti, S. Spadari, and F. Fochoer. 1990. "The Specific Binding of Nuclear Protein(s) to the CAMP Responsive Element (CRE) Sequence (TGACGTCA) Is Reduced by the Misincorporation of U and Increased by the Deamination of C." *Nucleic Acids Research* 18 (19): 5775–80.
- Verstegen, Ruud H. J., Maaikje A. A. Kusters, Eugénie F. A. Gemen, and Esther DE Vries. 2010. "Down Syndrome B-Lymphocyte Subpopulations, Intrinsic Defect or Decreased T-Lymphocyte Help." *Pediatric Research* 67 (5): 563–69.
- Vinson, Robert K., and Barbara F. Hales. 2002. "Expression and Activity of the DNA Repair Enzyme Uracil DNA Glycosylase during Organogenesis in the Rat Conceptus and Following Methotrexate Exposure in Vitro." *Biochemical Pharmacology* 64 (4): 711–21.
- Visnes, Torkild, Berit Doseth, Henrik Sahlin Pettersen, Lars Hagen, Mirta M. L. Sousa, Mansour Akbari, Marit Otterlei, Bodil Kavli, Geir Slupphaug, and Hans E. Krokan. 2009. "Uracil

- in DNA and Its Processing by Different DNA Glycosylases.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 364 (1517): 563–68.
- Vlahopoulos, Spiros, Maria Adamaki, Nikolas Khoury, Vassilis Zoumpourlis, and Istvan Boldogh. 2019. “Roles of DNA Repair Enzyme OGG1 in Innate Immunity and Its Significance for Lung Cancer.” *Pharmacology & Therapeutics* 194 (February): 59–72.
- Voliotis, Margaritis, Rebecca M. Perrett, Chris McWilliams, Craig A. McArdle, and Clive G. Bowsher. 2014. “Information Transfer by Leaky, Heterogeneous, Protein Kinase Signaling Systems.” *Proceedings of the National Academy of Sciences of the United States of America* 111 (3): E326–33.
- Waals, Lizet M. van der, Jamila Laoukili, Jennifer M. J. Jongen, Danielle A. Raats, Inne H. M. Borel Rinkes, and Onno Kranenburg. 2019. “Differential Anti-Tumour Effects of MTH1 Inhibitors in Patient-Derived 3D Colorectal Cancer Cultures.” *Scientific Reports* 9 (1): 819.
- Wallace, Susan S. 2013. “DNA Glycosylases Search for and Remove Oxidized DNA Bases.” *Environmental and Molecular Mutagenesis* 54 (9): 691–704.
- Wang, Dong, Leina Ma, Bin Wang, Jia Liu, and Wenyi Wei. 2017. “E3 Ubiquitin Ligases in Cancer and Implications for Therapies.” *Cancer Metastasis Reviews* 36 (4): 683–702.
- Wang, Shiliang, and Anguang Hu. 2016. “Comparative Study of Spontaneous Deamination of Adenine and Cytosine in Unbuffered Aqueous Solution at Room Temperature.” *Chemical Physics Letters* 653 (June): 207–11.
- Wang, X., H. Lu, A. M. Urvalek, T. Li, L. Yu, J. Lamar, C. M. DiPersio, P. J. Feustel, and J. Zhao. 2011. “KLF8 Promotes Human Breast Cancer Cell Invasion and Metastasis by Transcriptional Activation of MMP9.” *Oncogene* 30 (16): 1901–11.
- Wang, Y. J., Y. S. Ho, M. J. Lo, and J. K. Lin. 1995. “Oxidative Modification of DNA Bases in Rat Liver and Lung during Chemical Carcinogenesis and Aging.” *Chemico-Biological Interactions* 94 (2): 135–45.
- Wang, Yuru, Sehee Park, and Peter A. Beal. 2018. “Selective Recognition of RNA Substrates by ADAR Deaminase Domains.” *Biochemistry* 57 (10): 1640–51.
- Wardle, Josephine, Peter M. J. Burgers, Isaac K. O. Cann, Kate Darley, Pauline Heslop, Erik Johansson, Li-Jung Lin, et al. 2008. “Uracil Recognition by Replicative DNA Polymerases Is Limited to the Archaea, Not Occurring with Bacteria and Eukarya.” *Nucleic Acids Research* 36 (3): 705–11.
- Waters, T. R., P. Gallinari, J. Jiricny, and P. F. Swann. 1999. “Human Thymine DNA Glycosylase Binds to Apurinic Sites in DNA but Is Displaced by Human Apurinic Endonuclease 1.” *The Journal of Biological Chemistry* 274 (1): 67–74.
- Weeks, L. D., G. E. Zentner, P. C. Scacheri, and S. L. Gerson. 2014. “Uracil DNA Glycosylase (UNG) Loss Enhances DNA Double Strand Break Formation in Human Cancer Cells Exposed to Pemetrexed.” *Cell Death & Disease* 5 (February): e1045.
- Weeks, Lachelle D., Pingfu Fu, and Stanton L. Gerson. 2013. “Uracil-DNA Glycosylase Expression Determines Human Lung Cancer Cell Sensitivity to Pemetrexed.” *Molecular Cancer Therapeutics* 12 (10): 2248–60.
- Wenger, Anna, Sandra Ferreyra Vega, Teresia Kling, Thomas Olsson Bontell, Asgeir Store Jakola, and Helena Carén. 2019. “Intratumor DNA Methylation Heterogeneity in Glioblastoma: Implications for DNA Methylation-Based Classification.” *Neuro-Oncology* 21 (5): 616–27.
- Whitaker, Amy M., and Bret D. Freudenthal. 2018. “APE1: A Skilled Nucleic Acid Surgeon.” *DNA Repair* 71 (November): 93–100.

- Wick, Wolfgang, Michael Platten, Christoph Meisner, Jörg Felsberg, Ghazaleh Tabatabai, Matthias Simon, Guido Nikkhah, et al. 2012. “Temozolomide Chemotherapy Alone versus Radiotherapy Alone for Malignant Astrocytoma in the Elderly: The NOA-08 Randomised, Phase 3 Trial.” *The Lancet Oncology* 13 (7): 707–15.
- Wiederholt, Carissa J., and Marc M. Greenberg. 2002. “Fapy.DG Instructs Klenow Exo(-) to Misincorporate Deoxyadenosine.” *Journal of the American Chemical Society* 124 (25): 7278–79.
- Willems, Luc, and Nicolas Albert Gillet. 2015. “APOBEC3 Interference during Replication of Viral Genomes.” *Viruses* 7 (6): 2999–3018.
- Williams, Jessica S., Anders R. Clausen, Scott A. Lujan, Lisette Marjavaara, Alan B. Clark, Peter M. Burgers, Andrei Chabes, and Thomas A. Kunkel. 2015. “Evidence That Processing of Ribonucleotides in DNA by Topoisomerase 1 Is Leading-Strand Specific.” *Nature Structural & Molecular Biology* 22 (4): 291–97.
- Williams, Jessica S., Daniel B. Gehle, and Thomas A. Kunkel. 2017. “The Role of RNase H2 in Processing Ribonucleotides Incorporated during DNA Replication.” *DNA Repair* 53 (May): 52–58.
- Williams, Jessica S., Dana J. Smith, Lisette Marjavaara, Scott A. Lujan, Andrei Chabes, and Thomas A. Kunkel. 2013. “Topoisomerase 1-Mediated Removal of Ribonucleotides from Nascent Leading-Strand DNA.” *Molecular Cell* 49 (5): 1010–15.
- Wong, Donny, Michael S. DeMott, and Bruce Demple. 2003. “Modulation of the 3'→5'-Exonuclease Activity of Human Apurinic Endonuclease (Ape1) by Its 5'-Incised Abasic DNA Product.” *The Journal of Biological Chemistry* 278 (38): 36242–49.
- Wu, Hao, Xiaoji Wu, Li Shen, and Yi Zhang. 2014. “Single-Base Resolution Analysis of Active DNA Demethylation Using Methylase-Assisted Bisulfite Sequencing.” *Nature Biotechnology* 32 (12): 1231–40.
- Wu, Hao, and Yi Zhang. 2014. “Reversing DNA Methylation: Mechanisms, Genomics, and Biological Functions.” *Cell* 156 (1–2): 45–68.
- Wyatt, M. D., and D. M. Wilson 3rd. 2009. “Participation of DNA Repair in the Response to 5-Fluorouracil.” *Cellular and Molecular Life Sciences: CMLS* 66 (5): 788–99.
- Xanthoudakis, S., and T. Curran. 1992. “Identification and Characterization of Ref-1, a Nuclear Protein That Facilitates AP-1 DNA-Binding Activity.” *The EMBO Journal* 11 (2): 653–65.
- Xu, Yufei, Feizhen Wu, Li Tan, Lingchun Kong, Lijun Xiong, Jie Deng, Andrew J. Barbera, et al. 2011. “Genome-Wide Regulation of 5hmC, 5mC, and Gene Expression by Tet1 Hydroxylase in Mouse Embryonic Stem Cells.” *Molecular Cell* 42 (4): 451–64.
- Xu, Zhaoguo, Li Yu, and Xiaoye Zhang. 2013. “Association between the HOGG1 Ser326Cys Polymorphism and Lung Cancer Susceptibility: A Meta-Analysis Based on 22,475 Subjects.” *Diagnostic Pathology* 8 (August): 144.
- Yan, Nan, Elizabeth O'Day, Lee Adam Wheeler, Alan Engelman, and Judy Lieberman. 2011. “HIV DNA Is Heavily Uracilated, Which Protects It from Autointegration.” *Proceedings of the National Academy of Sciences of the United States of America* 108 (22): 9244–49.
- Yan, Yan, Xiangzi Han, Yulan Qing, Allison G. Condie, Shashank Gorityala, Shuming Yang, Yan Xu, Youwei Zhang, and Stanton L. Gerson. 2016. “Inhibition of Uracil DNA Glycosylase Sensitizes Cancer Cells to 5-Fluorodeoxyuridine through Replication Fork Collapse-Induced DNA Damage.” *Oncotarget* 7 (37): 59299–313.

- Yang, Haotian, Rehan M. Villani, Haolu Wang, Matthew J. Simpson, Michael S. Roberts, Min Tang, and Xiaowen Liang. 2018. “The Role of Cellular Reactive Oxygen Species in Cancer Chemotherapy.” *Journal of Experimental & Clinical Cancer Research: CR* 37 (1): 266.
- Yang, Shan, Angela Proctor, Lauren L. Cline, Kaiulani M. Houston, Marcey L. Waters, and Nancy L. Allbritton. 2013. “ β -Turn Sequences Promote Stability of Peptide Substrates for Kinases within the Cytosolic Environment.” *The Analyst* 138 (15): 4305–11.
- Ye, N., G. P. Holmquist, and T. R. O’Connor. 1998. “Heterogeneous Repair of N-Methylpurines at the Nucleotide Level in Normal Human Cells.” *Journal of Molecular Biology* 284 (2): 269–85.
- Yoon, Jung-Hoon, Gita Bhatia, Satya Prakash, and Louise Prakash. 2010a. “Error-Free Replicative Bypass of Thymine Glycol by the Combined Action of DNA Polymerases Kappa and Zeta in Human Cells.” *Proceedings of the National Academy of Sciences of the United States of America* 107 (32): 14116–21.
- . 2010b. “Error-Free Replicative Bypass of Thymine Glycol by the Combined Action of DNA Polymerases Kappa and Zeta in Human Cells.” *Proceedings of the National Academy of Sciences of the United States of America* 107 (32): 14116–21.
- Yoon, Jung-Hoon, Louise Prakash, and Satya Prakash. 2009. “Highly Error-Free Role of DNA Polymerase Eta in the Replicative Bypass of UV-Induced Pyrimidine Dimers in Mouse and Human Cells.” *Proceedings of the National Academy of Sciences of the United States of America* 106 (43): 18219–24.
- Yoshioka, Ken-Ichi, Yoshiko Yoshioka, and Peggy Hsieh. 2006. “ATR Kinase Activation Mediated by MutS α and MutL α in Response to Cytotoxic O6-Methylguanine Adducts.” *Molecular Cell* 22 (4): 501–10.
- You, Changjun, Xiaoxia Dai, Bifeng Yuan, Jin Wang, Jianshuang Wang, Philip J. Brooks, Laura J. Niedernhofer, and Yinsheng Wang. 2012. “A Quantitative Assay for Assessing the Effects of DNA Lesions on Transcription.” *Nature Chemical Biology* 8 (10): 817–22.
- Young, Matthew D., and Sam Behjati. 2018. “SoupX Removes Ambient RNA Contamination from Droplet Based Single Cell RNA Sequencing Data.” *BioRxiv*. <https://doi.org/10.1101/303727>.
- Yu, Kefei, Frederic Chedin, Chih-Lin Hsieh, Thomas E. Wilson, and Michael R. Lieber. 2003. “R-Loops at Immunoglobulin Class Switch Regions in the Chromosomes of Stimulated B Cells.” *Nature Immunology* 4 (5): 442–51.
- Zemmour, David, Rapolas Zilionis, Evgeny Kiner, Allon M. Klein, Diane Mathis, and Christophe Benoist. 2018. “Single-Cell Gene Expression Reveals a Landscape of Regulatory T Cell Phenotypes Shaped by the TCR.” *Nature Immunology* 19 (3): 291–301.
- Zhang, Ruobing, Gilbert O. Fruhwirth, Oana Coban, James E. Barrett, Thomas Burgoyne, Sang Hak Lee, Paul Dennis Simonson, et al. 2017. “Probing the Heterogeneity of Protein Kinase Activation in Cells by Super-Resolution Microscopy.” *ACS Nano* 11 (1): 249–57.
- Zhang, Zhang, Li-Qian Chen, Yu-Li Zhao, Cai-Guang Yang, Ian A. Roundtree, Zijie Zhang, Jian Ren, Wei Xie, Chuan He, and Guan-Zheng Luo. 2019. “Single-Base Mapping of M6A by an Antibody-Independent Method.” *Science Advances* 5 (7): eaax0250.
- Zhao, Boxuan Simen, Ian A. Roundtree, and Chuan He. 2017. “Post-Transcriptional Gene Regulation by mRNA Modifications.” *Nature Reviews. Molecular Cell Biology* 18 (1): 31–42.

- Zheng, Grace X. Y., Jessica M. Terry, Phillip Belgrader, Paul Ryvkin, Zachary W. Bent, Ryan Wilson, Solongo B. Ziraldo, et al. 2017. “Massively Parallel Digital Transcriptional Profiling of Single Cells.” *Nature Communications* 8 (January): 14049.
- Zheng, Li, and Binghui Shen. 2011. “Okazaki Fragment Maturation: Nucleases Take Centre Stage.” *Journal of Molecular Cell Biology* 3 (1): 23–30.
- Zheng, Shiwei, Efthymia Papalexi, Andrew Butler, William Stephenson, and Rahul Satija. 2018. “Molecular Transitions in Early Progenitors during Human Cord Blood Hematopoiesis.” *Molecular Systems Biology* 14 (3): e8041.
- Zimmer, Anjali D., and Douglas Koshland. 2016. “Differential Roles of the RNases H in Preventing Chromosome Instability.” *Proceedings of the National Academy of Sciences of the United States of America* 113 (43): 12220–25.
- Zimmermann, Michal, Olga Murina, Martin A. M. Reijns, Angelo Agathangelou, Rachel Challis, Žygimantė Tarnauskaitė, Morwenna Muir, et al. 2018. “CRISPR Screens Identify Genomic Ribonucleotides as a Source of PARP-Trapping Lesions.” *Nature* 559 (7713): 285–89.
- Zitnik, Marinka, and Jure Leskovec. 2017. “Predicting Multicellular Function through Multi-Layer Tissue Networks.” *ArXiv [Cs.LG]*. arXiv. <http://arxiv.org/abs/1707.04638>.

APPENDIX A

MEASURING DNA REPAIR ACTIVITIES IN SINGLE CELLS

Abstract

DNA repair is essential for cellular homeostasis and is carried out through multiple and often redundant pathways. The current methods to study DNA repair provide a coarse-grained view of DNA repair capacity and are often limited to studying a single or few types of DNA repair at a time. We developed a sequencing-based functional assay for measuring DNA repair activities that can be scaled to measure repair events catalyzed by endogenous activities in single cells on substrates containing any synthetically available DNA modification. We show that the assay quantitatively measures repair of substrates for base excision repair, ribonucleotide excision repair, and direct reversal. We have used this method to measure base excision repair and ribonucleotide repair at single-cell resolution.

Introduction

Endogenous and exogenous sources can lead to a number of different DNA damage events. These DNA lesions can be the product of oxidation, alkylation, deamination, misincorporation, and more (Bauer, Corbett, and Doetsch 2015b). Cells have developed many, and sometimes overlapping pathways to repair a slew of DNA damage events (Bauer, Corbett, and Doetsch 2015b). Some of the most abundant DNA damage events are the products of DNA replication where deoxyuracil and ribonucleotides are incorporated into nascent DNA strands (Goulian, Bleile, and Tseng 1980b; Clausen et al. 2013b). While these incorporations aren't inherently mutagenic, their presence in the genome can alter the stability of the DNA strand and genome regulation by altering the affinity of DNA binding proteins. The cell's ability to repair its DNA across many different DNA repair pathways is termed the cell's DNA repair capacity.

There are many methods to measure DNA repair capacity and each has strengths and weaknesses. One of the most common methods is the comet assay, which measures either nonspecific or endonuclease-specific DNA damage events (Langie, Azqueta, and Collins 2015b). The comet assay has been useful to understand DNA repair on a genomic level and has been used to measure DNA repair capacity in cells in response to drug treatments or genetic knockdown of DNA repair factors (Langie, Azqueta, and Collins 2015b). Novel methods like fluorescence-based host cell reactivation assays use fluorescent protein-expressing plasmids that contain a DNA lesion (Nagel et al. 2016). After transfection into cells and repair of the DNA lesion, the cell will express the fluorescent protein. Multiplexed HCR assays have been developed to measure multiple repair pathways simultaneously (Nagel et al. 2016; Chaim et al. 2017b). These assays have been used to measure interindividual differences in DNA repair capacity and to predict DNA damage chemotherapy sensitivity (Chaim et al. 2017b; Nagel et al. 2016). However, HCR assays require transient transfection of a reporter plasmid in the cells and it is not suitable for many primary tissue samples. Finally, DNA repair enzyme activity can be measured in cell-free extracts. One cell-free extract method is a DNA repair microarray where oligonucleotides containing a single DNA lesion site and a 5'-fluorophore are immobilized on a slide. These substrates are exposed to cell free extract and upon repair of the lesion, the fluorophore is washed away and DNA repair can be measured as a function of fluorescence (Sylvie Sauvaigo et al. 2010). DNA repair microarrays have been used to measure DNA repair capacity in primary samples and have found that DNA repair capacity varies between individuals and can change with age and upon the onset of disease (Sylvie Sauvaigo et al. 2010; Forestier, Douki, et al. 2012b). While each of these methods have been used to help expand our knowledge of DNA repair in healthy and disease states, our understanding of DNA repair in highly heterogeneous populations like cancer is lacking.

Single-cell methods have greatly expanded over the last 10 years and as the methodology becomes more accessible and affordable, our understanding of cell heterogeneity has greatly expanded as well. Single-cell mRNA sequencing has been used to find new cell populations and track developmental lineages of organs and whole organisms (Stuart and Satija 2019b). In addition to mRNA sequencing, we can also measure chromatin accessibility, protein abundance, and DNA sequence at single-cell resolution. All of these methods focus on measuring the abundance of molecules in single cells (Stuart and Satija 2019b). However, abundance does not always correlate to function. For example, post-translational modifications or sub-cellular localization can have a much greater impact on phenotypes than abundance. Because of this, we developed a method to measure enzymatic function in single cells.

We developed a droplet-based single-cell method that can measure the activity of dozens of DNA repair enzymes simultaneously (**Figure A.1**). We immobilized DNA repair substrates, in the form of DNA hairpins that contain a single DNA lesion, on toyopearl beads that were uniquely barcoded by split-pool synthesis to contain a bead barcoded and a unique molecular identifier (**Figure A.1a-b**). Multiple DNA repair substrates were ligated to a single bead via splint ligation. To ensure that our method was specific and sensitive, all of the substrates were tested in bulk cell-free extract (**Figure A.1c**). We measured single-cell DNA repair of a uracil-containing substrate and a ribonucleotide-containing substrate in a mixture of UNG- and RNASEH2C-deficient cell lines. We also measured DNA repair activity on BER, RER, and DR substrates in cell-free extracts across a range of substrate concentrations, cell extract concentrations, and time points. Additionally, we measured DNA repair activities at single-cell resolution. Both the multiplexed bulk validation method and the single-cell method could be valuable methods to answer questions about DNA repair capacity in primary tissues and complex cell mixtures.

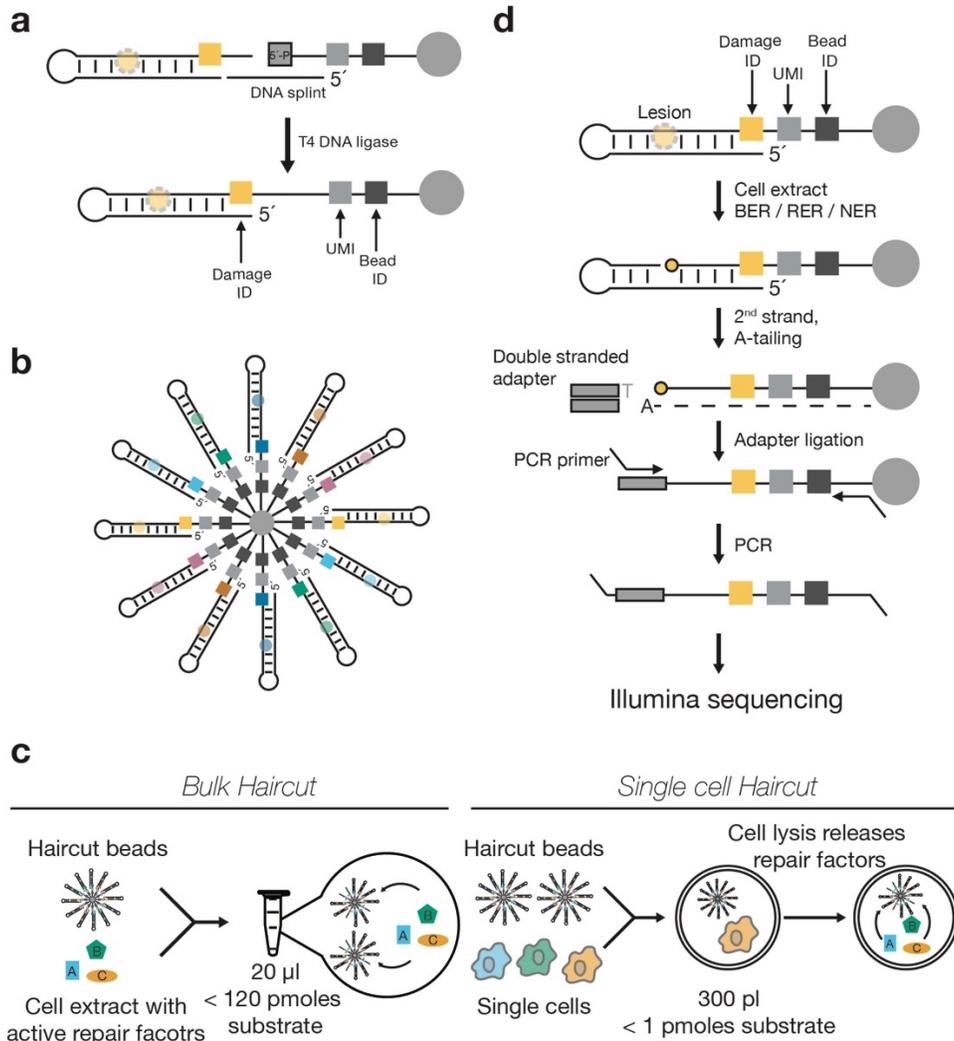


Figure A.1: Multiplexed DNA repair assay to measure DNA repair in bulk cell extracts and single cells

a. DNA repair substrates in the form of DNA hairpins that contain a single DNA lesion are splinted and ligated onto toyopearl beads. Each bead contains a unique bead ID and a UMI.

b. Many substrates can be ligated to a single bead, enabling multiplexed measurement of DNA repair activities in a single cell.

c. Beads containing DNA repair substrates can be used to measure DNA repair activities in bulk cell extract (left) or in single cells (right).

d. DNA repair activities are measured by measuring strand incision events catalyzed by DNA repair factors via a standard DNA library preparation. Following sequencing, the repair activity is identified by the sequence of the substrate or substrate ID. Individual cells are identified by the cell ID and repair events can be quantified using the UMIs.

Materials and methods

Preparation of DNA repair substrates for substrate validation

Oligonucleotides were purchased from IDT. Large hairpin substrates for haircut were prepared via a multi-unit splinted ligation. The common 3'-amino, pieces 1-3, and the split oligo were mixed in equimolar concentrations (20 μ M each) in 50 mM NaCl. The oligonucleotide mixture was then heated to 95 °C for 5 minutes and cooled for 1 hour to 25 °C, then placed on ice. 10X T4 DNA ligase buffer (Enzymatics) was added to the annealed oligonucleotides to 1X concentration (50 mM Tris-HCl, 10 mM MgCl₂, 5 mM DTT, 1 mM ATP, pH 7.6). 1800 units of Rapid T4 DNA ligase (Enzymatics) was added to the annealed oligonucleotides and incubated at 17 °C overnight. Ligated hairpins were gel purified in the same manner as the single cell substrates described above.

Conjugation of DNA repair substrates to beads for substrate validation

For each sample, 100 μ g of NHS-activated magnetic beads (Pierce) were washed once with 1 mM ice-cold HCl then incubated with 150 pmol of purified oligonucleotide hairpins with a 3'-amino group diluted in 10 μ L 0.1 M sodium carbonate/sodium bicarbonate buffer (pH 9.0) for 2 hours at room temperature. If multiple oligonucleotide hairpin substrates were added to the beads, they were mixed in equal mole ratios. Beads were then washed twice with 0.1 M glycine (pH 2.0), then open NHS sites were blocked with 1M 3-amino-2-propanol in 0.1 M sodium carbonate/sodium bicarbonate buffer (pH 9.0) for 1 hour at room temperature or overnight at 4 °C. The beads were washed three times with 100 mM NaCl, 5 mM EDTA, 50 mM Tris-HCL pH 7.5 (WB), then stored in WB at 4 °C until use.

Cell lysis

Adherent cells were grown until 70-80% confluent, then isolated via trypsin. 1 million cells were washed once with PBS and lysed for 10 minutes on ice with 100 μ L M-PER (Thermo Scientific) supplemented with 1x Halt protease inhibitor cocktail (Thermo Scientific). Lysate was cleared by centrifugation at 10,000 g for 10 min at 4 °C. Relative protein concentration was determined using absorbance at 280 nm on a Nanodrop 2000 (Thermo Scientific). Lysates were diluted to 2.5 mg/ml total protein (unless otherwise noted) and immediately incubated with Haircut beads.

Haircut library preparation for substrate validation

50 μ g (20 μ l) total protein from freshly made cell lysate was added to 100 μ g of prepared Haircut beads in MPER lysis buffer supplemented with Halt protease inhibitor cocktail (Thermo Scientific) and 2.5 mM MgCl₂ (unless otherwise noted). The mixture was incubated for 2 hours (unless otherwise noted) at 37 °C with shaking at 350 rpm. The beads were washed three times with detergent wash buffer (100 mM NaCl, 5 mM EDTA, 50 mM Tris-HCl pH 7.5, 0.5 % SDS, 1 M Urea; DWB) and twice with WB. To test for direct reversal repair activity, those samples were digested with Pst1 with the following 20 μ l reaction: 1x Cutsmart (NEB), 1 μ l Pst1 (NEB) for 1 hour at 37 °C with shaking. The beads were then washed three times with DWB and twice with WB. The beads were then incubated in 20 μ l second strand reaction (1 μ M second strand primer, 1x Blue buffer (Enzymatics, 50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl₂, 1 mM DTT, pH 7.9), 400 μ M dNTPs, 10 units Klenow 3'-5' exo- (Enzymatics)) for 1 hour at 37 °C with shaking at 350 rpm. The beads were then washed 3 times with DWB and 2 times with WB. The beads were then incubated in 20 μ L A-tailing reaction (1x Blue Buffer (Enzymatics), 400 μ M dATP, 10 units Klenow 3'-5' exo- (Enzymatics)) for 30 minutes at 37° C with shaking at 350 rpm. The beads were

washed three times with DWB and 2 times with WB. The beads were then incubated in 30 μ L Illumina Y adapter ligation reaction (66 mM Tris-HCl, 10 mM MgCl₂, 1 mM DTT, 1 mM ATP, 7.5% PEG 6000, pH 7.6, 1 μ M annealed Y adapters, 1200 units Rapid T4 DNA Ligase (Enzymatics)) and incubated at 25 °C for 30 minutes. The beads were washed 6 times with DWB and 2 times with WB, then resuspended in 20 μ L water. 5 μ L of beads were added to 45 μ L Illumina PCR reaction (1x Phusion HF buffer (NEB), 200 μ M dNTPs, 0.6 μ M ILMN PCR primers (F and R), 2 units Phusion High Fidelity DNA polymerase) and 23 cycles of PCR with 98 °C melting temperature for 15 seconds, 65 °C annealing temperature for 15 seconds, and 72 °C extension temperature for 15 seconds. The PCR reaction was purified using Agencourt AMPure XP (Beckman Coulter) beads as described by manufacture. The purified PCR reaction was quantified using the Qubit HS dsDNA fluorometric quantitation kit (Thermo Scientific) and paired end sequenced on a MiSeq with 75 base pair read lengths using a MiSeq 150 cycle reagent kit v3 (Illumina) at the University of Colorado Anschutz Medical Campus Genomics and Microarray core.

Data analysis for substrate validation

UMIs were extracted from fastq files from read 2 and appended to the read names. Fastq files were aligned to hairpin reference fasta files using bowtie2 (v2.3.2) (Langmead and Salzberg 2012b). Some substrates had the same sequence, in these cases the fastq files were separated based on a hairpin ID that was extracted from read 2. Following alignment, BAM files were combined using samtools merge (v1.9) (H. Li et al. 2009b). Aligned BAM files were deduplicated using umi_tools dedup (Smith, Heger, and Sudbery 2017b). 5' alignment counts were calculated for each sample using bedtools genomecov (2.26.0) (Quinlan and Hall 2010b). Counts were normalized to total sequencing depth per sample unless standards were included.

5'-phosphate standards for normalization

To include an internal standard to normalize DNA repair activity measurements, oligonucleotides of similar length to repair products containing 5' phosphates were included at 3000, 600, 120, 25, and 6 fmoles per sample in bulk DNA repair assays. In the data processing, the total counts per substrate was summed. The lowest concentration standard (6 fmoles) did not fit the linear model and so it was dropped from normalization calculations. For each sample, a linear model was calculated (lm) for the log₁₀ transformed counts for each substrate vs the log₁₀ transformed picomoles of each substrate added. The calculated linear model was used to calculate the normalized counts for each sample for each hairpin position using the following formula:

$$\log_{10}(\text{Normalized data (pmol repaired)}) \\ = \frac{-\log_{10}(\text{counts at position}) - \text{Intercept of standard model}}{\text{Slope of standard model}}$$

Single-cell bead synthesis

Single-cell haircut beads were synthesized on an Expedite oligonucleotide synthesizer rebuilt by Biolytic Lab Performance. Phosphoramidites and synthesizer chemicals were purchased from Glen Research. Oligonucleotides were synthesized on Tyoparl HW-65S beads. Cell barcodes were synthesized by doing 12 rounds of split-pool synthesis where the synthesis was split into 4 columns and each column was coupled with a different amidite. Following base coupling, the beads were pooled and re-distributed into four columns. Following synthesis, beads were deprotected for 24 hours at room temperature in ammonium hydroxide and methyl amine. The beads were then dried in a speed vac and washed with 100% ethanol at least three times. The beads were resuspended in 10 mM Tris, pH 7.5, 1 mM EDTA, with 0.01% Tween-20 and stored in the fridge until use. Beads were washed before use.

Single-cell bead ligation

To attach DNA repair substrates to single-cell beads, 4.8 nmoles of each hairpin substrate (U:A, riboG:C, Abasic:G) were splint ligated onto 350,000 beads. Hairpins were mixed with 14.8 nmoles splint in 50 mM NaCl and added to 350,000 single-cell haircut beads. The splint, beads, and hairpins were annealed by heating to 95° C and cooling at 0.1 °C per second to 4° C. 2X Rapid Ligase Buffer (Enzymatics) was added to 1X concentration along with 1000 U T4 DNA ligase (Enzymatics) and incubated at room temperature for 30 minutes. The beads were washed twice with H₂O, twice with WB, twice with DWB, twice with WB and twice with H₂O.

Single-cell Haircut on Dolomite system

UNGKO Hap1 and RNASEH2CKO cells were isolated using the same method as above (CHAPTER 2). UNGKO and RNASEH2CKO cells were mixed 1:1 and were counted and diluted to a concentration of 300 cells/μl in PBS. Haircut beads were resuspended in 2X lysis buffer (2% NP40, 100 mM Tris, 200 mM NaCl, 2X ProteaseHalt (Invitrogen), 5 mM MgCl₂) to 300 beads/μl. Beads were loaded 300 μl beads in 2X lysis buffer into the sample loading loop. Ran the beads at 30 μl/minute, cells at 30 μl/minute, and the oil (Droplet Generation Oil for EvaGreen; BioRad) at 200 μl/minute. Ran for the system until beads stopped flowing (~20 minutes). Incubated emulsion at 37° C for 2 hours with end-over-end turning. Following incubation at 37° C, the emulsion was broken by adding 3 mL perfluorooctanol (Synquest) with 10% PicoBreak (Sphere Fluidics) for color. Emulsion was separated by centrifuging at 100 XG for 1 minute. The top lay was removed and the bottom layer was washed with 6X SSC (0.9 M NaCl, 90 mM sodium citrate, pH 7.0) several times to remove beads from the interface. Beads were then transferred to the Eppendorf tube and then washed three times with DWB, three times with WB and three times with H₂O. Recovered beads were counted and then split into two tubes with ~17,000 beads each for the

remainder to the library construction as described above. Size selected libraries on 2% agarose E-gel (Thermo) collecting samples between 150 and 300 base pairs and then re-PCR amplified for sequencing.

Results

Substrate validation with knockout cell lines

Base excision repair, ribonucleotide excision repair, and direct reversal substrates were validated using knockout cell lines. I measured cleavage events on uracil containing substrates (U:A and U:G) one base downstream of where the uracil was present in the substrate (**Figure A.2a,b**). This signal was dependent on the presence of UNG. Lysate from UNGKO cell lines did not have cleavage events on the A:U substrate and only moderate cleavage on the G:U substrate presumably due to the presence of SMUG1 and MBD4 that specifically recognize U:G mismatches (**Figure A.2b**).

Ribonucleotide containing substrates undergo sequential cleavage events catalyzed by RNASEH2 and Pole δ and Fen1 resulting in cleavage on the 5' and 3' side of the ribonucleotide respectively (**Figure A.2c**). I measured cleavage events on the ribonucleotide containing substrates (rG, rA, rC, and rU) consistent with the downstream cleavage on the 3' side of the ribonucleotide. This signal was dependent on RNASEH2 activity, but not secondary repair pathways mediated by Pol β or Top1 (**Figure A.2c**).

The substrate containing the alkylation lesion, ethenoA, is repaired by the glycosylase MPG. Cleavage of the ethenoA substrate is dependent on MPG and is significantly reduced in lysate from MPGKO cells (**Figure A.2d**). MPG also removes the adenine deamination product, inosine, however, we did not measure the inosine-containing substrates in the MPGKO lysate, although we would predict that their activity is dependent on MPG.

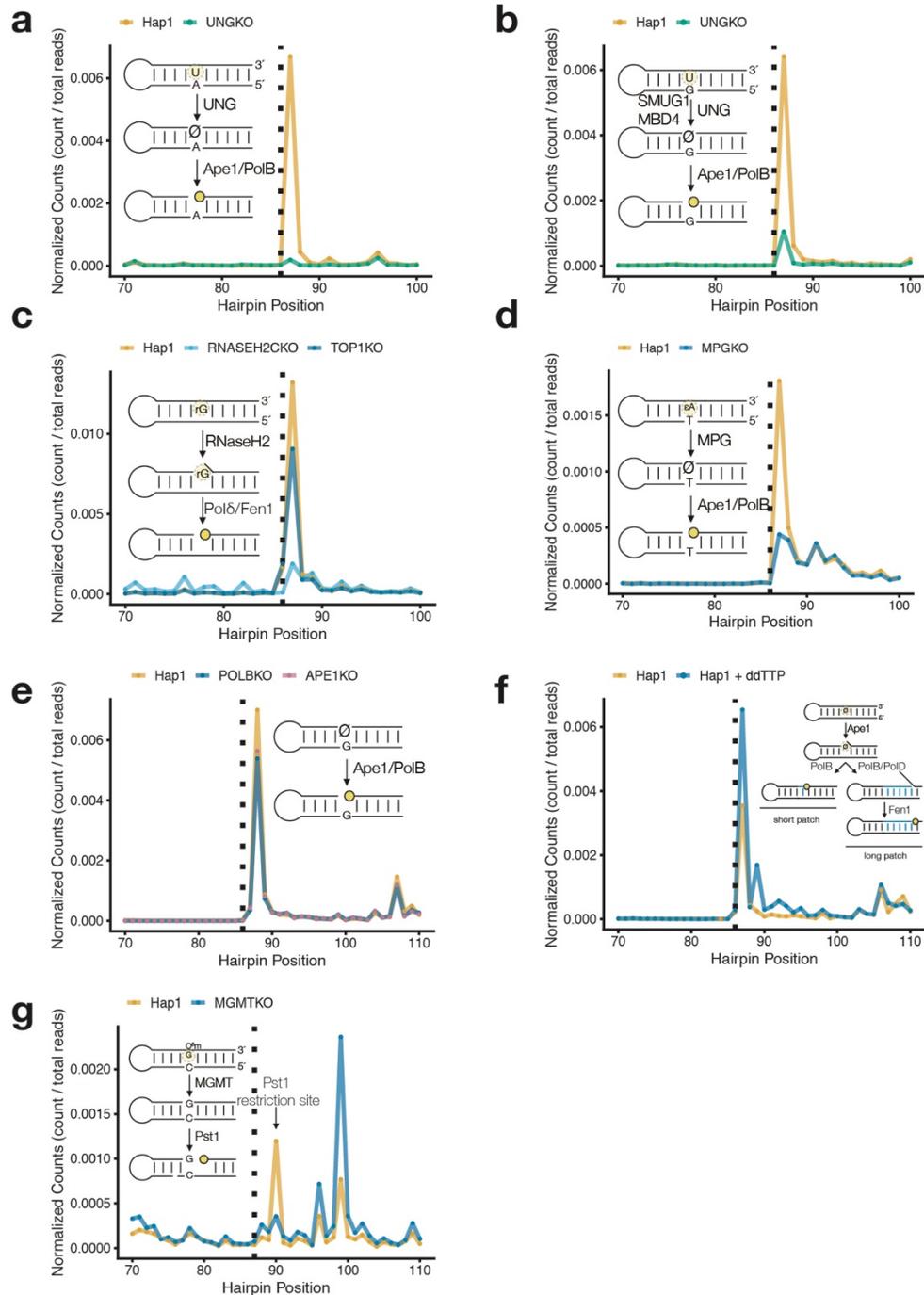


Figure A.2: DNA repair activities are validated in bulk cell extract from knockout cell lines

a. Repair of a substrate containing a uracil:adenine (U:A) base pair initiates with UNG-mediated removal of the uracil nucleobase followed by processing of the abasic site by Ape1 and Pol β (inset). Normalized counts of incision and processing (counts / sequencing depth) are plotted against the position of the hairpin. Incision counts are located one base downstream of where the uracil was located in the substrate, consistent with the DNA repair mechanism. The signal is dependent on the presence of UNG.

b. Repair of a substrate containing a uracil:adenine (U:G) base pair initiates with UNG-mediated (or SMUG1, MBD4, TDG) removal of the uracil nucleobase followed by processing of the abasic site by Ape1 and Pol β (inset). Normalized counts of incision and processing (counts / sequencing depth) are plotted against the position of the hairpin. Incision counts are located one base downstream of where the uracil was located in the substrate, consistent with the DNA repair mechanism. The signal is mostly dependent on the presence of UNG.

c. Repair of a substrate containing a riboG:C base pair initiates with RNASEH2-mediated incision 5' of the ribonucleotide followed by processing by Pol δ and Fen1 (inset). Normalized counts of incision and processing (counts / sequencing depth) are plotted against the position of the hairpin. Incision counts are located one base downstream of where the rG was located in the substrate, consistent with the DNA repair mechanism. The signal is mostly dependent on the presence of RNASEH2C, but not on the secondary ribonucleotide repair factor TOP1.

d. Repair of a substrate containing a etheno-adenine:thymine (ϵ A:T) base pair initiates with MPG-mediated removal of the ϵ A nucleobase followed by processing of the abasic site by Ape1 and Pol β (inset). Normalized counts of incision and processing (counts / sequencing depth) are plotted against the position of the hairpin. Incision counts are located one base downstream of where the ϵ A was located in the substrate, consistent with the DNA repair mechanism. The signal is dependent on the presence of MPG.

e. Repair of substrates containing a abasic:guanine base pair initiates with Ape-1-mediated incision followed by processing of the single-base gap by either Pol β (short-patch repair) (inset). Normalized counts of incision and processing (counts / sequencing depth) are plotted against the position of the hairpin. Incision counts are located two bases downstream of where the abasic site was located in the substrate, consistent with the DNA repair mechanism. The signal is not dependent on the presence of Pol β or Ape1, possibly due to redundant AP endonuclease activities in cells.

f. Repair of substrates containing a abasic:guanine base pair initiates with Ape-1-mediated incision followed by processing of the single-base gap by either Pol β (short-patch repair) or Pol δ/β and Fen1 (long-patch repair) (inset). Normalized counts of incision and processing (counts / sequencing depth) are plotted against the position of the hairpin. Incision counts are located one or more bases downstream of where the abasic site was located in the substrate, consistent with the short-patch and long-patch BER mechanisms. The long-patch BER signal >1 base downstream of the abasic site is inhibited by the inclusion of ddTTP. The addition of ddTTP causes additional downstream peaks after a ddTTP is incorporated into the strand.

g. Repair of substrates containing an O6-methyl-guanine:cytosine (O6mG:C) base pair is completed by the direct reversal enzyme MGMT (inset). Repair is measured by restriction digest of the repaired substrate by Pst1. Normalized counts of incision and processing (counts / sequencing depth) are plotted against the position of the hairpin. Incision counts are located two bases downstream of where the O6mG site was located in the substrate, consistent with the restriction site. The restriction digest is dependent on the presence of MGMT, since the presence of the methyl group on O6mG inhibits Pst1 incision.

Ape1 and Pol β activities can be complemented by other AP lyases and polymerases in the cell (Lebedeva et al. 2012b; Khodyreva et al. 2010b; Prasad et al. 2015b), so activity on the abasic containing substrate was not significantly reduced when exposed to lysate from APE1KO or POLBKO cells (**Figure A.2e**). On the abasic containing substrate we measured both short-patch and long-patch base excision repair activity. Long-patch base excision repair could be halted at specific locations with the addition of ddNTPs in the assay (**Figure A.2f**). Specifically, the addition of ddTTP led to excision products at locations where a T was incorporated into the substrate following incision of the abasic site and several bases incorporated.

The direct reversal substrate O6mG was used to test the activity of MGMT. MGMT activity does not result in a strand incision event, however, the presence of the alkylation lesions O6mG blocks the activity of some restriction enzymes. The O6mG substrate was designed to contain a Pst1 restriction site that contains the O6mG lesion (Encell and Loeb 1999c). Upon removal of O6mG by MGMT, the substrate can be cleaved using Pst1. Following incubation with lysate from Hap1 cells, the O6mG substrate was cleaved by Pst1. This cleavage was dependent on MGMT (**Figure A.2g**).

Other substrates tested did not show enzyme-specific activity either because the signal did not diminish in knockout cell lines (Tg, TT dimer) or we did not measure any activity at expected positions (8-oxo-G, carboxyC, formylC, mismatch) (**Figure A.3**). Tg and TT-dimers are bulky lesions that would stall DNA polymerases, so it is possible that we are unable to distinguish between repair activity and truncated second-strand products. If we included translesion polymerases in the second-strand reaction, we may be able to bypass the lesions and only capture repair-mediated strand incision events.

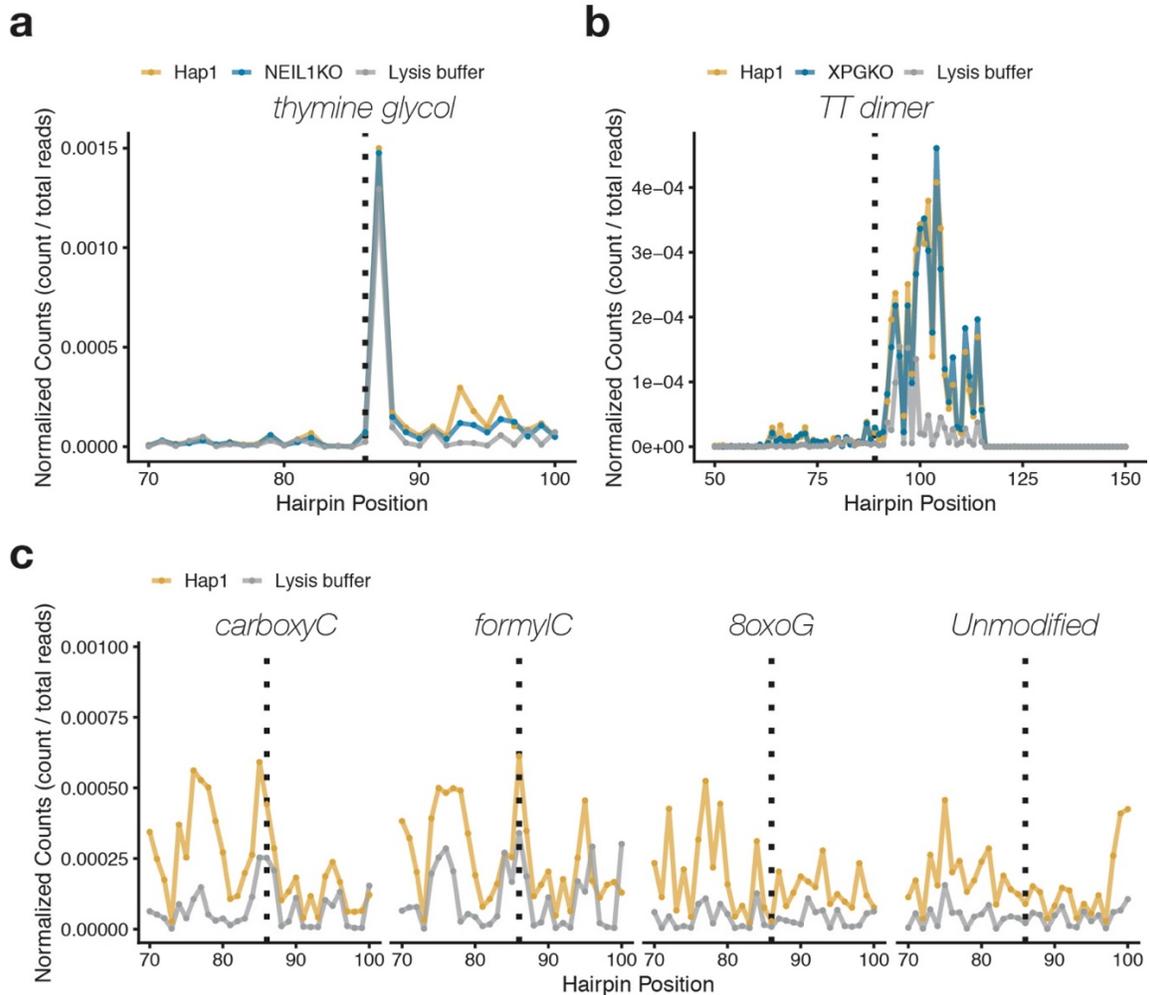


Figure A.3: Nonspecific signals in Haircut

a. Repair of substrates containing an thymine glycol:adenine (Tg:A) base pair is initiated by DNA glycosylases, Neil1, Neil2, Neil3, and others. Normalized counts of incision and processing (counts / sequencing depth) are plotted against the position of the hairpin. Counts located one base downstream, consistent with known repair mechanisms, are not dependent on NEIL1 glycosylase activity (blue) or the presence of cell extract (grey).

b. Repair of substrates containing a pyrimidine dimer (TT dimer) is initiated by nucleotide excision repair machinery. Normalized counts of incision and processing (counts / sequencing depth) are plotted against the position of the hairpin. Counts located several bases downstream, are consistent with known repair mechanisms, but are not dependent on the NER factor, XPG (blue) or the presence of cell extract (grey).

c. Repair of substrates containing a 8-oxoguanine:cytosine, carboxycytosine:guanine, or formylcytosine:guanine are initiated by BER machinery. Normalized counts of incision and processing (counts / sequencing depth) are plotted against the position of the hairpin. Counts around the lesion are not dependent on cell extract (grey), nor are they above background signal on the unmodified substrate (right).

Testing sensitivity and repeatability

To determine if the enzyme signals identified on the hairpin substrates were reproducible and quantitative, I did replicate, time course, and dilution experiments (**Figure A.4**). Hap1 cells were lysed in triplicate (1 million cells per lysate) and the lysate was incubated with Haircut beads. Signal was normalized to internal 5'-phosphate containing standards. Two replicates had measurements that were the same +/- 10% for all repair activities (**Figure A.4a**). The third replicate had the same data trends but had ~50% fewer repair events for some repair activities than the other replicates, indicating that the variability of the bulk assay is greater than expected. Technical and biological replicates are necessary to make quantitative claims on DNA repair capacity if using the bulk assay.

To determine if the signal was dependent on enzyme concentration and incubation time, I diluted the cell extract 100-fold and incubated the extract with DNA repair substrates for 15 minutes to 2 hours. Repair signals were dependent on incubate time and lysate concentration (**Figure A.4b,c**). However, the highest lysate concentration (2 mg/ml) and longest time point (2 hours) had fewer incision events than shorter timepoints (1 hour) and some dilutions (1 mg/ml). This reduction in signal is most likely due to complete repair events occurring on the substrate that are invisible to the assay. I also included five separately barcoded uracil, ribonucleotide, and abasic containing substrates and measured the relationship between the repaired substrates versus the amount of substrate added to the assay. I found that the amount of substrate added compared to the amount of substrate repaired in the assay was nearly perfectly correlated ($R^2 > 0.95$), indicating that repair measurements are consistent across a 50-fold substrate concentration range (**Figure A.4d**).

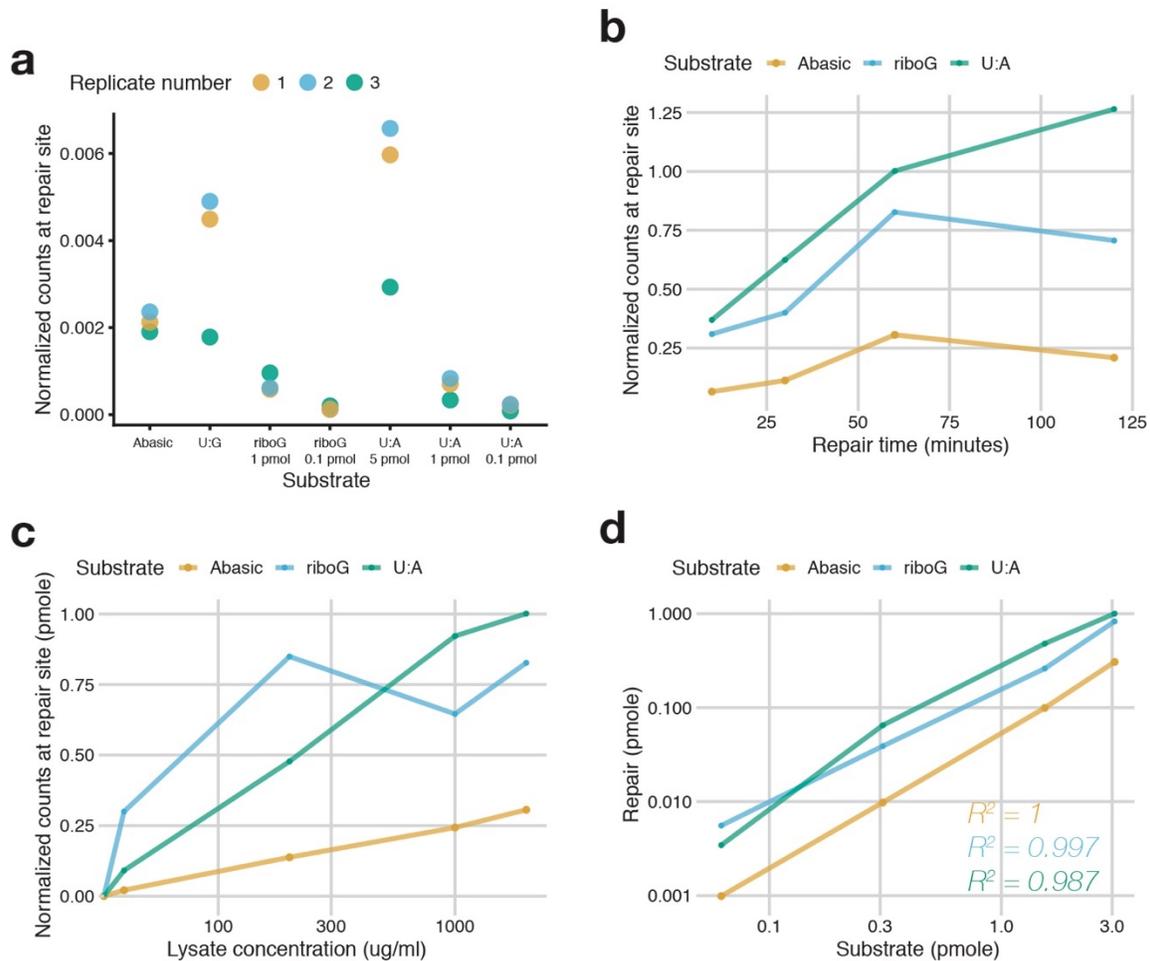


Figure A.4: Repeatability and consistency of Haircut signals

a. Haircut was done in biological triplicate. Repair signal at defined repair positions is plotted by substrate. Replicates 1 and 2 show good agreement but replicate 3 varies in uracil containing substrates only.

b. Haircut substrates were incubated with cell extract for 15-120 minutes. Repair signal at defined repair positions is plotted over time. Generally, repair signal increases up to 60 minutes and then decreases slightly at 120 minutes.

c. Haircut substrates were incubated with cell extracts spanning a 100-fold protein concentration range. Repair signal at defined repair positions is plotted by lysate concentration. Generally, repair signal increases as lysate concentration increases except repair of the ribonucleotide substrate (blue).

d. Haircut substrates spanning a 50-fold concentration range were incubated with cell extracts. Repair signal at defined repair positions is plotted by substrate concentration. Generally, repair signal is directly correlated with substrate concentration.

Single-cell barnyard experiment

To determine if Haircut was sensitive at single-cell resolution, I ligated DNA repair substrates onto single-cell Haircut beads that contained a cell barcode and UMI. I only included three substrates onto the beads: U:A, riboG:C, Abasic:G. I mixed two knockout cell lines 1:1; UNGKO and RNASEH2CKO Hap1 cells (**Figure A.5**). Using the Dolomite single-cell RNAseq system, I encapsulated single-cell Haircut beads and single cells in oil-water emulsion drops. I measured UNG or RNASEH2 activity in droplets for ~100 cells. While the cell number was low, DNA repair activity was segregated, where cells that had uracil repair activity did not have RER and vice versa, indicating we were measuring repair activity in single cells (**Figure A.5**, right).

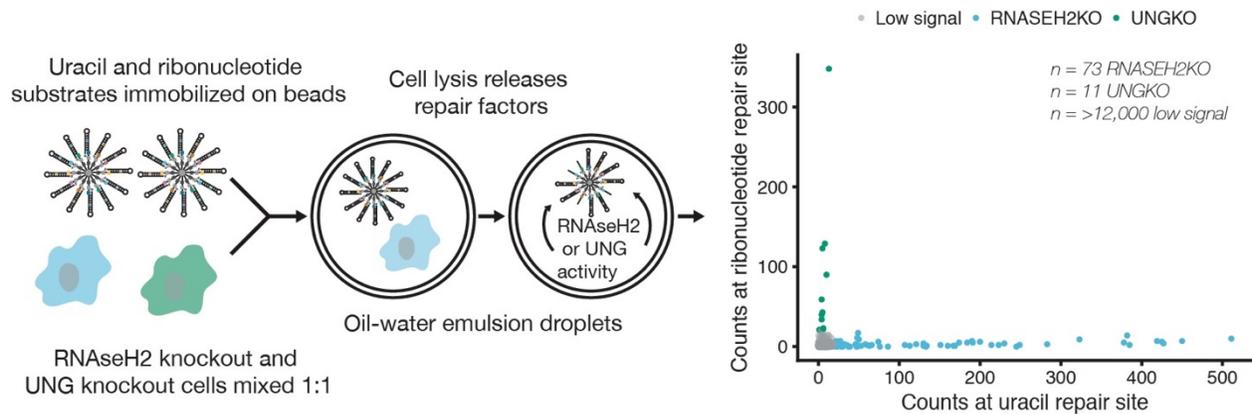


Figure A.5: Measuring DNA repair in single cells

Single-cell Haircut beads containing uracil, ribonucleotide, and abasic substrates were encapsulated with single cells from a single-cell suspension of a mixture of Hap1 cells containing null alleles of either UNG or RNASEH2C (right). Repair signals at defined repair positions for the ribonucleotide substrate are plotted by the repair signal at defined repair positions for the uracil substrate. The level of strand incision for uracil and ribonucleotide substrates was used to classify cells as either UNGKO (green) or RNASEH2CKO (blue) based on strand incision activity (UMI counts at position 87 for ribonucleotide, position 87 for uracil) greater than 5% of the maximum for all cells. Cells that fall on the x- or y-axis are single RNASEH2CKO and UNGKO cells, respectively; and cells with low signal (<5% of the max for both activities) are in grey.

Discussion

Through bulk and single cell experiments I validated DNA repair substrates. Base excision repair, ribonucleotide excision repair, and direct reversal substrates were validated through these experiments. However, due to different substrate and enzymatic requirements I was unable to measure repair activity on nucleotide excision repair and mismatch repair substrates. Additionally, the bulk experiments indicated that this assay functions across a wide range of enzyme concentration, however, some enzymes (e.g. OGG1, TDG) may be at too low an abundance in cells and so we were unable to measure that activity in our assay (**Figure A.3c**). Our inability to measure activity on the 8-oxoG substrate may be an example of an enzyme that is not present in high enough concentrations in order for us to measure its activity. Alternatively, it could be that our substrate design or buffer components were not suitable for OGG1 enzymatic activity. The latter is unlikely because similar small double-stranded substrates have been used to measure OGG1 activity *in vitro* and in cell-free extracts (Bravard et al. 2006; Pons et al. 2010b). OGG1 is active in similar buffers used in our assay. Paradoxically, qRT-PCR results indicate that OGG1 is expressed in our cell lines (data not shown). OGG1 activity can be inhibited with nitric oxide stress (Jaiswal et al. 2001), so it is possible that inhibitory post-translational modifications inactivate OGG1 in our cell lines or assay conditions.

The substrate thymine glycol (Tg) has a large signal, but it's not specific to NEIL1 glycosylase activity (**Figure A.4a**). Thymine glycol is a large bulky lesion that could cause polymerase stalling during the second strand reaction producing a truncated extension product that is competent for ligation. These truncation products look identical to glycosylase activity; however, they are also present in the lysis buffer only control. While signal on the Tg containing substrate was high in cell extract, there was also significant signal in the lysis buffer control. Due

to the polymerase-halting nature of Tg (Yoon et al. 2010b), it is difficult, if not impossible, to truly measure enzymatic activity on the Tg containing substrate. These same problems are also found on the thymine dimer substrate. TT dimers stall DNA polymerases that could lead to signals around the TT dimer independent of DNA repair activity. Additionally, NER does not occur at precise single-base sites around the lesion like BER. NER cleavage events occur upstream (~18-20 bases) and downstream (2-4 bases) of the lesion (Reardon and Sancar 2005). The diffuse nature of these incision events makes it difficult to quantify. Nucleotide excision repair is also difficult to measure for a number of other reasons. First, because ~25 bases are excised from the substrate, NER substrates need to be quite large in comparison to the 20-base pair BER substrates. Second, NER requires assembly of many DNA repair factors at the site of damage in order for incision to occur. The assembly of NER factors may require a larger substrate. Our largest NER substrate was only ~80 base pairs that may not facilitate assembly and activity of NER factors. Finally, NER factors require supplemental cofactors, like ATP. We were unable to measure NER in our assay, but it may be possible to develop substrates and assay conditions that can measure NER in the future.

This version of the single-cell DNA repair assay provided a valuable steppingstone in developing an assay to simultaneously measure DNA repair enzyme activity and mRNA expression in single cells. Some of the major limitations to this method are as follows. While I was able to measure single-cell DNA repair activities in a barnyard experiment, it would be difficult to identify cell types using DNA repair activities alone in a more complex cell mixture. This limitation is even more abundant when using the data from **Chapter II**. The range of signal we measured in cells with high repair activity (B cells) to low repair activity (monocytes) that differ only as a factor of two with variation sufficient to bridge the gap (**Figure 2.7**). Inability to distinguish cell types with high and low repair significantly reduces the applicability of this assay

in its current form. The Dolomite single-cell RNAseq platform is an easily customizable system, however, the difficulties of isolating single cells reproducibly did not make it an ideal system to use if trying to measure DNA repair activity in more than a few thousand cells. Because of these limitations, I developed a method to measure DNA repair activities in single cells using the 10x Genomics single-cell mRNA expression platform (**Chapter II**).

APPENDIX B

DNA REPAIR HETEROGENEITY IN INDIVIDUALS WITH TRISOMY 21

Abstract

We measured single-cell gene expression and DNA repair capacity in peripheral blood mononuclear cells from healthy individuals with trisomy and compared them to age and sex-matched disomic individuals. We identified several known differences in cell type abundance in T21 PBCMs (fewer B cells) and gene expression differences in type I interferon response genes in CD14 monocytes. Additionally, we found differences in some DNA repair activities. Namely, we found increased abundance of DNA repair intermediates in our assay that is consistent with previously published data that decreased expression of DNA polymerase beta in individuals with trisomy 21 leads to a deficiency in single strand break repair and an increase in DNA strand breaks. Due to our small cohort (2 T21, 2 D21) it is difficult to make conclusions on these data, but with more samples and/or integration with additional datasets (cy-TOF, bulk RNAseq, etc) we can begin to understand DNA repair heterogeneity in individuals with trisomy 21.

Introduction

Trisomy 21 (T21) is the most common chromosomal condition in humans and is the cause of Down Syndrome (DS) (Dierssen 2012). People born with DS have both mental and physical challenges, as well as predispositions for a number of conditions and diseases including Alzheimer's disease, autoimmune disorders, and some types of cancers. People with DS are also less likely to get solid tumors and hypertension, leading to the question of the molecular mechanisms driving the predispositions and protective effects of T21. Few direct gene-dosage links have been made that link disease predispositions to specific genes on chromosome 21. For example, the gene *DYRK1A* located on chromosome 21 promotes the development of leukemia

in murine models of DS (Malinge et al. 2012) and the extra dosage of the APP gene promotes Alzheimer's disease (Cataldo et al. 2003). However, much of the molecular mechanisms driving the disease predispositions are still unknown.

DNA repair is an essential process in maintaining cellular homeostasis. Alterations in cell metabolism and redox regulation can alter the rate of DNA damage and DNA repair (Mikhed et al. 2015). At the cellular level, T21 causes global changes in gene expression and the cellular proteome (Sullivan et al. 2016, 2017). Some T21 cells have consistent activation of the interferon response (Sullivan et al. 2016) that may contribute to the development of leukemia through altered cell metabolism and changes in DNA damage and repair in the hematopoietic system, although little direct experimental evidence for this has been measured. Activation of the interferon response can lead to an increase in steady-state DNA damage levels, however, studies have found no increased sensitivity to DNA damage in cells from individuals with T21 (Morawiec et al. 2008). There have been a number of studies that measured DNA repair in T21 cells, but the results have been inconclusive. Some studies found that there was no difference in DNA repair rates (Steiner and Woods 1982). While other studies have found a decrease in DNA repair rates (Morawiec et al. 2008; Necchi et al. 2015), and other studies found an increase in DNA repair rates in T21 cells (Chiricolo et al. 1993).

Given the heterogeneity in DNA repair between cell types, it is possible that the inconclusive DNA repair data in T21 cells is dependent on the type of DNA damage and the cell type. Some gene expression differences in individuals with T21 are cell-type specific (Sullivan et al. 2016) indicating that DNA repair may also differ by cell type. We have developed a sensitive single-cell DNA repair assay that measures the activity of several DNA repair enzymes simultaneously with single-cell gene expression (Amanda L. Richer et al. 2020a) called Haircut.

Haircut measures cell-specific differences in base excision repair enzyme activity from single-cell suspensions. We used Haircut to measure single-cell gene expression and DNA repair capacity in peripheral blood mononuclear cells (PBMCs) from healthy individuals with T21 and compared them to age and sex-matched D21 individuals. The single-cell resolution of our data, allowed us to identify several known differences in cell type abundance in T21 PBCMs (fewer B cells, altered T cell populations) (Verstegen et al. 2010; Araya et al. 2019) and gene expression differences in type I interferon response genes in CD14 monocytes (Sullivan et al. 2016). Additionally, we found very small differences in some DNA repair activities. Namely, we found increased abundance of some DNA repair intermediates in our assay that is consistent with previously published data that decreased expression of DNA polymerase beta ($Pol\beta$) in individuals with T21 leads to a deficiency in single strand break repair and an increase in DNA strand breaks (Patterson and Cabelof 2012). Due to our small cohort (2 T21, 2 D21) it is difficult to make conclusions on these data, but with more samples and/or integration with additional datasets (cy-TOF, bulk RNAseq, etc) we can begin to understand DNA repair heterogeneity in individuals with T21.

Methods

PBMC collection

Fresh PBMCs were isolated from four healthy individuals (females, between 18-35 years old; 2 D21, 2 T21) through the Linda CRNIC Center Human Trisome Project. PBMCs were collected from 4-5 mL of fresh peripheral blood collected in sodium heparin tubes. PBMCs were isolated using a Ficoll gradient as before (Amanda L. Richer et al. 2020b) within 6 hours of blood draw.

10X Haircut

Single-cell mRNA and single-cell DNA repair measurements were made using the single-cell Haircut method (Amanda L. Richer et al. 2020b). The following substrates were added to the master mix to a final concentration of 5 nM each (50 nM total): U:A, G:U, riboG:C, I:T, I:C, Abasic:G, hydroxymethylU:A, ethenoA:T, O6mG:C, unmodified. Substrates were designed to be captured in 10x Genomics 3' mRNA capture kits V3 with feature barcoding technology. The hairpins contained a single strand tail complementary to capture sequence 1. Libraries were prepared as described previously (Amanda L. Richer et al. 2020b) with the addition of Pst1 digestion to detect O6mG direct reversal repair and using PCR primers that are consistent with capture sequence design.

Data analysis

Fastq processing

Single-cell mRNA and repair fastqs were processed as previously described (Amanda L. Richer et al. 2020b) with the following exception: cell barcodes from the mRNA analysis from cellranger (v3.0.2) were matched to the cell barcodes on the DNA repair substrates using the cell barcode reference file located in the cellranger (v3.0.2) software files (cellranger-3.0.2/cellranger-cs/3.0.1/lib/python/cellranger/barcodes/translation/3M-february-2018.txt.gz).

Single-cell analysis

Seurat objects were created and normalized as previously described ((Amanda L. Richer et al. 2020b), **Chapter II**). Samples were integrated using Seurat functions (FindAnchors and IntegrateData) (Stuart et al. 2019c).

Autosomal expression

Processed fastq files were input and normalized into Seurat objects as previously described ((Amanda L. Richer et al. 2020b), **Chapter II**). Each sample was integrated into a single Seurat object using Seurat functions (FindAnchors and IntegrateData). Using the integrated or nonintegrated gene expression data, the mean gene expression was calculated for all genes on chromosome 19 and chromosome 21. Genes and gene location were identified using Encode annotations (Pazin 2015).

Differential gene expression and GO analysis

Differential gene expression for D21 and T21 cells were calculated using Seurat function (FindMarkers, FindAllMarkers) (Butler et al. 2018b; Stuart et al. 2019c). Differential gene expression was calculated by cell type. Differential genes and p values were used as the input for GO term analysis using TopGO (Alexa and Rahnenführer 2009).

Results and discussion

Single-cell gene expression in PBMCs from people with T21

We collected PBMC samples from four adults between the ages of 18 and 35, two individuals with T21 and two individuals with D21. We captured between 1,700 and 5,500 cells per person and were able to identify cell types using gene expression. Samples were integrated using Seurat functions (Stuart et al. 2019c) and UMAP plots show that cells cluster according to cell type, rather than individual or chromosome 21 status (**Figure B.1a**). However, when we looked at gene expression on chromosome 21 and chromosome 19, we found higher average gene expression on chromosome 21 and chromosome 19 (**Figure B.1b**). It is possible that differences in sequencing depth affected our metric of chromosomal gene expression. We also integrated samples to remove sample and batch variations, but this integration method also removed variation

in average autosomal expression on chromosome 21 between individuals (data not shown). We expect that given higher and more constant sequencing depth for each sample; we would measure chromosome 21 gene expression difference.

Confirming known alterations in PBMCs and gene expression in people with T21

We were able to measure known differences in immune cell populations from individuals with T21. We confirmed the previously shown phenotype that people with T21 are deficient in the B cell population compared to people with D21 (Verstegen et al. 2010) (**Figure B.2a**). We were also able to identify CD8⁺ and CD4⁺ T cells and found that individuals with T21 had a higher proportion of CD8⁺ T cells and a lower proportion of naive CD4⁺ T cells (**Figure B.2b**), which was previously described (Araya et al. 2019). Even with our small dataset, these trends were consistent across individuals and confirmed known phenotypes. In addition to differences in cell populations, we also were able to recapitulate some previous data describing an altered innate immune response in some cell types from people with T21 (Sullivan et al. 2016). Our sequencing depth was low (~8,000 reads per cell) and our cohort of four individuals was quite small, however, we were able to identify GO terms associated with the innate immune response in monocytes from people with T21 (**Figure B.2c**). It is possible that with a higher sequencing depth for every sample, we may be able to find more differences in immune cell populations and gene expression differences in PBMCs from individuals with T21.

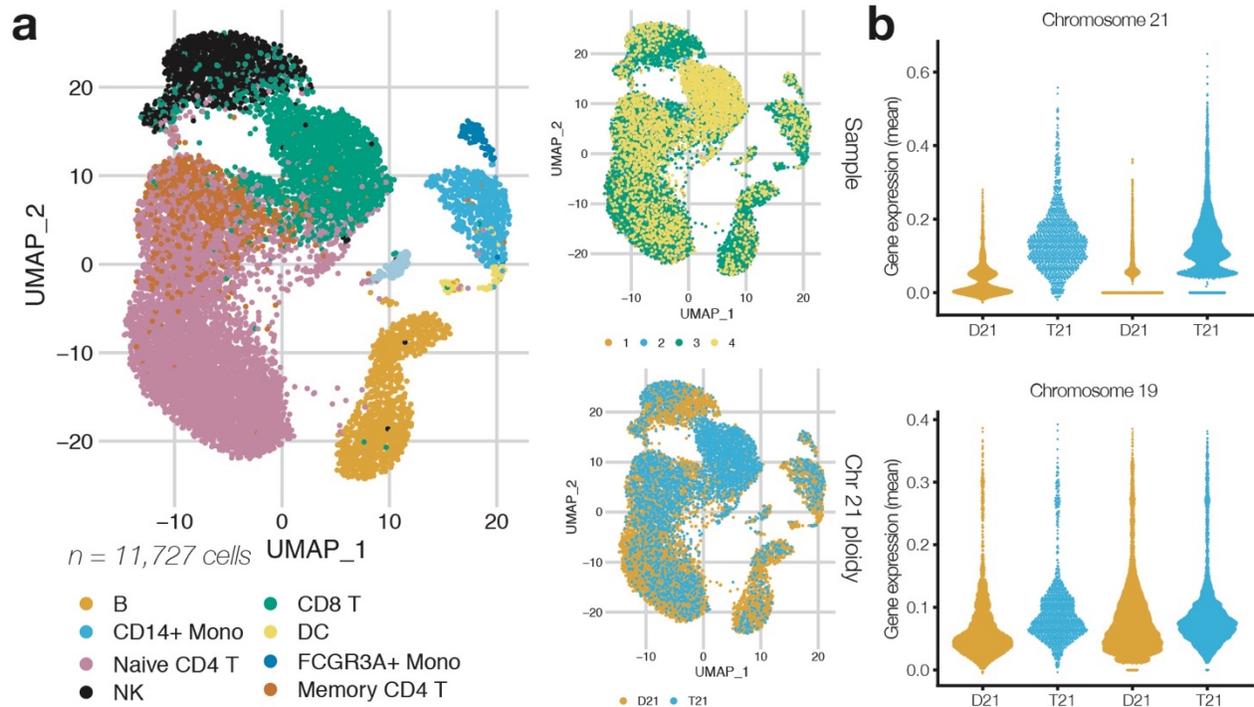


Figure B.1: Single cell analysis of primary peripheral blood mononuclear cells in people with trisomy 21

a. Two-dimensional UMAP projection of variable gene expression across 11,727 human PBMCs captured in a 10x Genomics 3' Gene Expression experiment. Major cell types were determined by marker gene expression (CD19 for B cells; IL7R for T cells; LYZ, FCGR3A, CD14 for monocytes; FCER1A for dendritic cells; GNLY for NK cells). Samples across four individuals, two people with T21 and two D21 individuals, were integrated and the major sources of variation are derived from cell types and not individuals or chromosome 21 ploidy.

b. Mean autosomal gene expression from chromosome 21 (top) and chromosome 19 (bottom) for each individual colored by chromosome 21 ploidy.

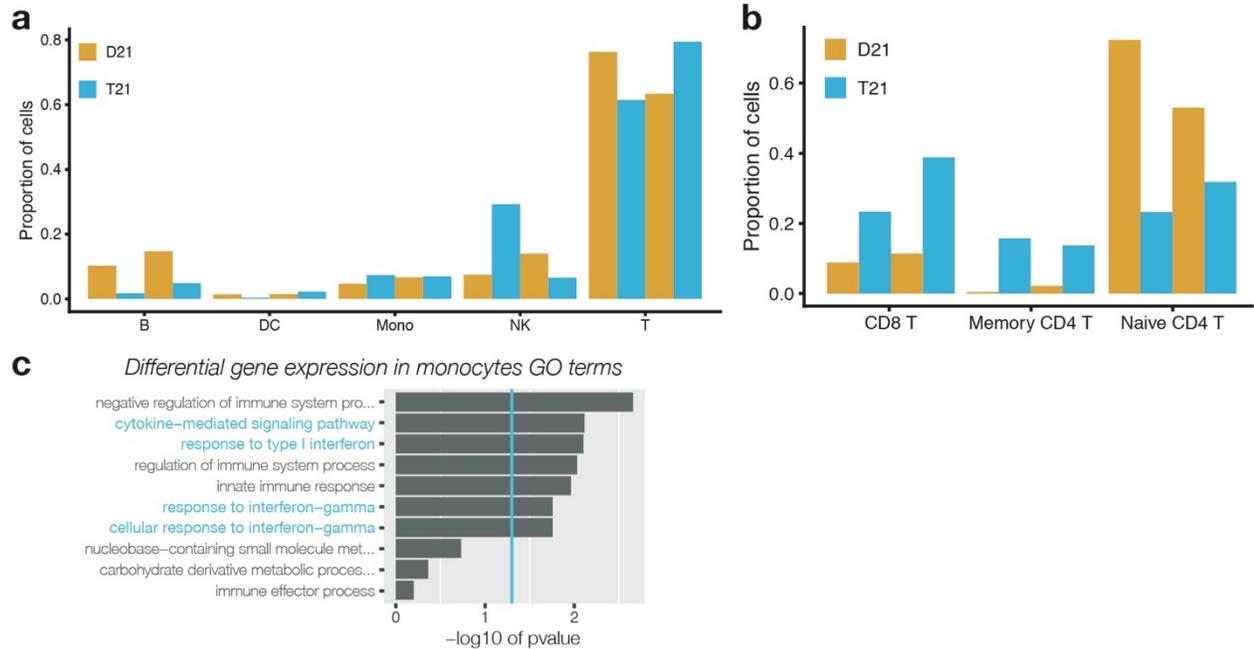


Figure B.2: Confirming known differences in cell populations and gene expression in individuals with T21 using single-cell RNA seq

a. Proportion of cells identified in four individuals with T21 or D21 by immune cell type, colored by chromosome 21 ploidy.

b. Proportion of cells identified in four individuals with T21 or D21 by T cell subtype, colored by chromosome 21 ploidy.

c. Gene ontology terms identified from differential gene expression in monocytes from individuals with T21 compared to individuals with D21 arranged by descending \log_{10} of p values. Blue line represents $P = 0.05$. GO terms highlighted in blue have been previously identified.

DNA repair in cells from people with T21 does not differ from D21 cells

We also measured single-cell DNA repair on DNA repair substrates for base excision repair and ribonucleotide excision repair enzymes. Major signals in Haircut come from the excision of damaged bases by DNA glycosylases and strand incision by AP endonucleases or RNaseH2 (**Figure B.3**, left). While some studies found differences in Pol β expression and activity in cells from people with T21 (Patterson and Cabelof 2012), we only measured moderate and mostly insignificant differences in repair intermediates from any cell type in our cohort (**Figure B.3**, middle and right). Again, our cohort is small and the sequencing depth per sample varies greatly. These factors make it difficult to make firm conclusions from these data. However, the lack of a large difference in DNA repair in any cell type between individuals with T21 and D21, indicate that any differences in functional DNA repair in healthy cells may be quite small.

More samples and sequencing depth are needed

Since we had such a small cohort, it is difficult to make firm conclusions from these data. Moreover, any differences that we saw in DNA repair capacity was driven by a single individual with T21. There is natural variability in DNA repair capacity between healthy individuals and the differences in repair intermediates measured in our data across all people is within the natural variation in DNA repair measured previously (Chaim et al. 2017c).

To fully answer questions about DNA repair in people with T21, we need to expand our cohort. Previous studies that found differences in immune cell populations between people with D21 and people with T21 have collected samples from ~10-30 individuals (Sullivan et al. 2016; Araya et al. 2019). However, the natural heterogeneity in DNA repair between individuals can range up to three standard deviations from the mean (Chaim et al. 2017c), so we may need to

measure a much larger cohort to understand any functional differences in DNA repair between individuals with T21 compared to individuals with D21.

To increase our cohort size, it would be very helpful to test Haircut using viability frozen samples. Single-cell techniques are plagued with batch effects and attempting to measure single-cell gene expression and DNA repair from dozens of individuals would be logistically challenging if we were to collect and run all the samples in a single day. However, the Human Trisome Project has viably frozen samples from hundreds of individuals. If we confirm that viably frozen samples can be used to collect reproducible single-cell DNA repair results, then we would be able to expand our study to understand interindividual differences in DNA repair in immune cells from individuals with T21 and D21.

The Human Trisome Project and people from the Espinosa lab have also created several datasets using samples from individuals with T21. If we integrated some of our single-cell gene expression data and DNA repair measurements with other data sets (e.g. CY-TOF, bulk RNA-seq, etc.), we may be able to further understand heterogeneity of DNA repair and cell functions in individuals with Down syndrome.

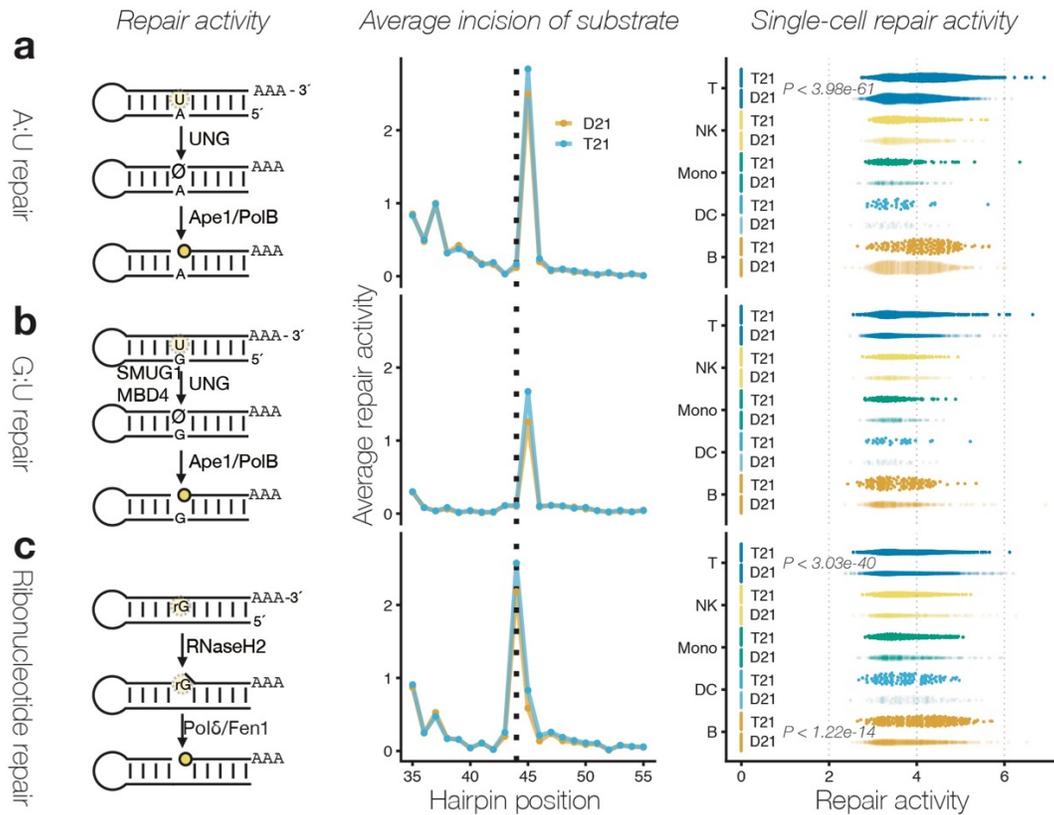


Figure B.3: Single-cell DNA repair measurements from PBMCs from individuals with T21

a. Repair of a substrate containing a uracil:adenine (U:A) base-pair initiates with UNG-mediated removal of the uracil nucleobase followed by processing of the abasic site by Ape-1 and Pol β (left). Chromosome 21 ploidy-specific counts of incision and processing (mean) are plotted against the position of the hairpin (middle). Single-cell repair activities (natural logarithm of counts at the incision site divided by total counts for that cell multiplied by a scaling factor of 10^4) are plotted for each cell from each cell type (right). T cells from individuals with T21 have higher uracil incision relative to T cells from individuals with D21 ($P < 10^{-60}$).

b. Repair of a substrates containing a uracil:guanine base-pair initiates with UNG, SMUG, or MBD4-mediated removal of the uracil nucleobase followed by processing of the abasic site by Ape-1 and Pol β (left). Chromosome 21 ploidy-specific counts of incision and processing (mean) are plotted against the position of the hairpin (middle). Single-cell repair activities (natural logarithm of counts at the incision site divided by total counts for that cell multiplied by a scaling factor of 10^4) are plotted for each cell from each cell type (right).

c. Repair of a substrate containing a riboG:C base-pair initiates with RNASEH2-mediated incision 5' of the ribonucleotide followed by processing by Pol δ and Fen1 (left). Chromosome 21 ploidy-specific counts of incision and processing (mean) are plotted against the position of the hairpin (middle). Single-cell repair activities (natural logarithm of counts at the incision site divided by total counts for that cell multiplied by a scaling factor of 10^4) are plotted for each cell from each cell type with significant comparisons noted (right).