AD_____

Award Number: W81XWH-12-1-0333

TITLE: Determining the Location of DNA Modification and Mutation Caused by UVB Light in Skin Cancer

PRINCIPAL INVESTIGATOR: Monica Ransom, PhD

CONTRACTING ORGANIZATION: The Regents of the University of Colorado
Aurora, CO 80045

REPORT DATE: September 2014

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
September 2014	Annual Summary	15 Aug 2013 – 14 Aug 2014

4. TITLE AND SUBTITLE

Determining the Location of DNA Modification and Mutation Caused by UVB Light in Skin Cancer

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-12-1-0333

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Monica Ransom, PhD

E-Mail: Monica.Ransom@ucdenver.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Regents of the University of Colorado
Aurora, CO 80045

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Extrinsic modifications to the DNA bases such as pyrimidine dimers can arise from a variety of exposures and can lead to aberrant cell growth or death. A detailed view of this base modification is necessary for a more complete view of genetic and epigenetic regulation but the process is poorly understood. We are developing a method to look at the precise genomic position of these modifications using a next generation sequencing approach. During this research period we have been performing preliminary experiments to aid in streamlining the processes that will be used in this research. We have purified or obtained the enzymes needed for aim 1 of this research. Using these enzymes we have shown that we obtain cleavage patterns consistent with the administered UV dosage and that sequencing libraries generated for both yeast and human cells show pyrimidine bias on the 5' end, indicating that we are sequencing the dimers. Understanding where these modifications occur is a critical first step to understanding the mutations they cause.

15. SUBJECT TERMS
UV DNA modification, HeLa cells, Skin Cancer

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	16	19b. TELEPHONE NUMBER *(include area code)*

Table of contents

Introduction

Ultraviolet (UV) light damages skin cells by causing the formation of dimers on adjacent pyrimidines in DNA. The two main forms of damage caused by UV light are cyclobutane pyrimidine dimers (CPDs) and 6-4 photoproducts (6-4pp). These nucleobase dimers prevent proper replication and transcription, and can lead to mutation if they are not properly repaired. Mutations caused by UV damage in tumor suppressor genes such as p53 have been found in the majority of skin cancers. Many studies have focused on these and other mutations, found in tumor suppressor genes, or have assessed bulk levels of modification. In this study we wish to elucidate a genome-wide picture of the primary UV-induced DNA modifications, and to determine which of these modifications go on to produce mutations by sequencing exomes of UV-exposed cells.

We have recently developed a general method for identifying modified DNA nucleobases in genomic DNA by coupling *in vitro* base excision with next-generation DNA sequencing. Using this novel approach in *S. cerevisiae* we have determined the precise genome-wide positions of pyrimidine dimers in heavily irradiated cells. We have shown that the sequences acquired showed a strong tendency to derive from genomic positions with pyrimidine dimers and distinguish between sites that have either CPDs or 6-4pp. Together these data show that mapping pyrimidine dimer modification is feasible and may yield tremendous insight into the genome-wide distribution UV-associated DNA dipyrimidine lesions.

Keywords: Pyrimidine dimers, UV light, modification mapping, excision-seq, UVDE, cyclobutane pyrimidine dimers, 6-4 photoproducts

Overall Project Summary

The first task tackled in the statement of work dealt with the generation of sequencing libraries in yeast and human cells. Previously we have shown that libraries could be generated in yeast exposed to UVC light using a commercial glycosylase and photolyases from a collaborating lab. In the first year of work we generated our own glycosylase to replace the commercial one that went off the market. This year we improved this protein preparation by replacing yeast with E. coli expression. We used gateway to clone the S. pombe UVDE glycosylase with the delta 288 mutation (1) into a pet-53-His vector under the T7 promoter. We transformed this construct into E. coli that are competent for protein expression and induced them overnight in .4mM IPTG. The cells were harvested, frozen and lysed by sonication. The lysate was clarified by centrifugation and the supernatant was purified over a nickel column as compared to the initial yeast protein purification (Fig 1A). This protein was concentrated and compared to our previous yeast purification and found to yield 10-15 times as much protein as the previous technique. The enzyme still sheared efficiently as shown in (FIG 1B).

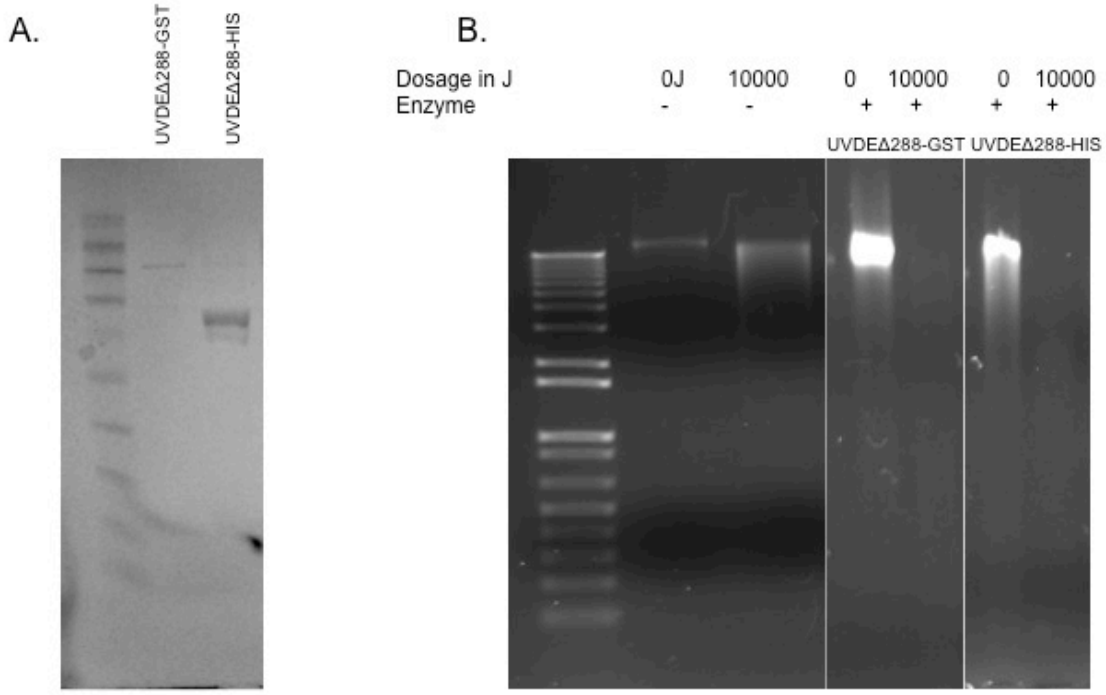

Figure 1. UVDE made in E. coli works similarly to UVDE purified from yeast. Protein was purified from E. coli containing T7 driven UVDEΔ288-HIS after induction. Protein expression was several fold increased from previous yeast GST expression (Fig 1A). The His-tagged enzyme was compared to the yeast GST enzyme on genomic DNA either untreated or dosed with 10000J/m2 of UV irradiation. When used at the same concentration we obtained similar shearing patterns (compare lane 3&4 to 4&5) from the two protein preps (Fig. 1B).

During year one of this funding we have shown that homemade UVDE yields a library similar to what we had obtained previously with commercial enzyme. During the last year we looked more in depth at the original data we received to determine the sensitivity of the assay as well as looking for patterns in the data we might have missed. We determined that the sensitivity of this assay in yeast was quite high with more than 85% of the aligned sequences acquired deriving from genomic positions with pyrimidine dimers. In total we saw that 38% of the total genomic dipyrimidines were hit in the CPD library with 72% of the TT dipyrimidines in the genome having reads (Fig 2A). The 6-4pp library hit only 5% of the total dipyrimidines indicating more specificity of the damage itself or of the repair enzyme used to generate the libraries. We also went on to look and the average number of hits in the two libraries and subsequently saw a increase in the average number of times each hit occurred in 6-4 photoproduct libraries, again indicating an increased specificity (Fig 2B.). We went on to further look at the local base content surrounding the modified dipyrimidines and saw that in CPD libraries the bases up and downstream of the modified base reflected the same percentages as the yeast genome (Fig. 2C), whereas in the 6-4pp library the base 3' to the dipyrimidine shows a bias to being a A residue (Fig 2D) (2). This may indicate an otherwise unknown specificity for the damage to occur within these trinucleotides or for the repair enzymes to be less efficient at repair of these sites. We also further analyzed the genomic positions of this data and showed that the coverage of modifications was generally uniform across the genome in yeast and the location of the dipyrimidines couldn't be associated with chromatin context or several other DNA features tested (data not shown).

A

dipyrimidines	Number of occurances in the genome	% of CPD dipyrimidines hit 1 or more times	% of CPD dipyrimidines hit 10 or more times	% 6-4pp dipyrimidines hit 1 or more times	% 6-4pp dipyrimidines hit 10 or more times
TT	1313442	73%	6%	3%	2%
TC	754111	17%	1%	6%	4%
CT	705190	18%	1%	1%	.5%
CC	473637	8%	1%	3%	2%
Sum	**3246380**	**38%**	**5%**	**3%**	**2%**
Non-dipyrimidine	755030	1%	.5%	.2%	0%

B.

Dipyrimidine	Average number of hits CPD	Average number of hits 6-4
TT	4.2	15.4
TC	2.7	15.4
CT	2.7	12
CC	2.5	13.4

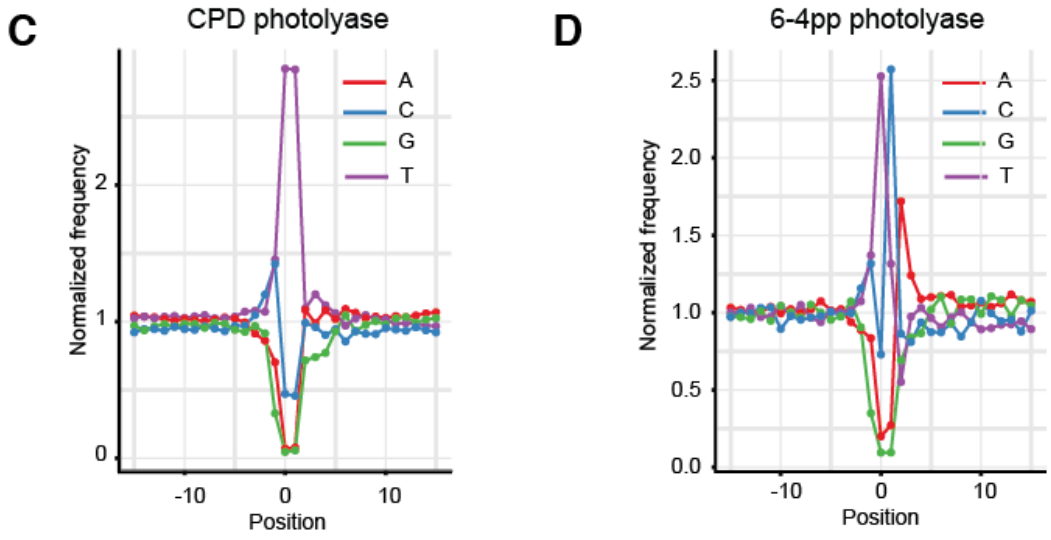

Figure 2. Additional data analysis on libraries obtained from yeast cells treated with a high dose of UVC light. Data for the number of hits in the genome are broken down into percentages of total dipyrimidines hit, and the percentage hit more than 10 times for CPD and 6-4 libraries (Fig. 2A). Data for the average number of times each position was hit is indicated for CPD and 6-4 libraries (Fig. 2B). Frequency of nucleotides relative to mapped positions of sequences from pre-digestion Excision-seq libraries for mapping cyclobutane dimers in S. cerevisiae. Position 0 corresponds to the mapped position of the 5' end for CPD (Fig. 2C) and 6-4 libraries (Fig. 2D).

We then used this method to look at dipyrimidines in human HeLa cells. We started with UVC to obtain a high level of damage to initially look at. We measured the UVC lethality of both yeast and HeLa cells and found HeLa cells to be 10 times more sensitive than yeast. We collected damaged HeLa cells following irradiation with low and high dosages of UVC (500 and 10000J/m^2) respectively. We collected genomic DNA for each sample and digested 8µg with 7.5µg of UVDE. We then treated these samples with either CPD or 6-4 photolyase for 2 hours under UVA light. Samples were then run through the standard Illumina protocol of polishing, a-tailing, adapter ligation and PCR. Libraries were obtained in low abundance for both libraries as stated for Task 2. Libraries were pooled and sequenced and as a test of quality the presence of a dipyrimidine on the 5' end of the read was established. The percentage of each dinucleotide combination for the whole genome was then determined and used as a control for base bias in the genome. Dinucleotide bias was observed in the +1 register (indicating the first base of the read and the one directly previous to it) as expected (Fig3A). Data from irradiated yeast treated in a similar fashion is shown for comparison (Fig3B). This bias was significantly reduced from what we had previously seen for yeast, which we hypothesize to be due to a variety of causes mainly the large size of the genome compared to the small sized fragments needed to generate libraries.

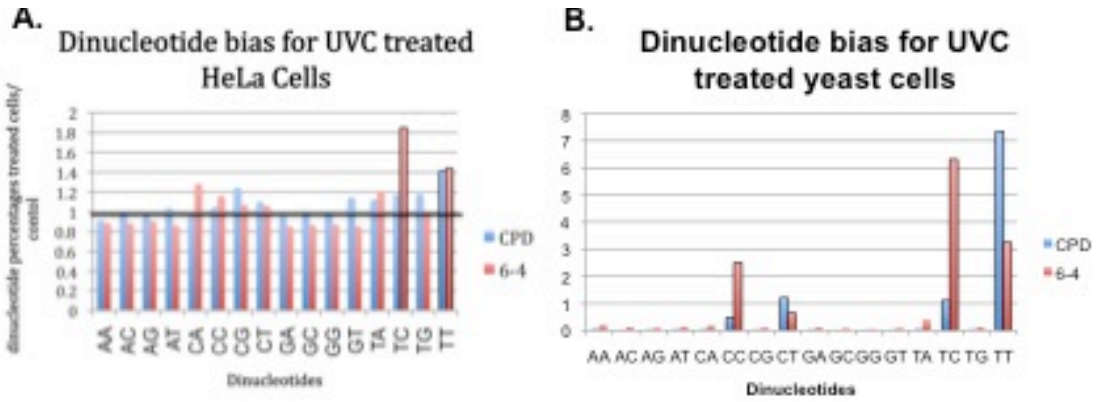

Figure 3. Illumina sequencing libraries were obtained from yeast and HeLa cells treated with a high dose of UVC light. Cells were treated with 10000J/m^2 of UVC light and genomic DNA was prepared and analyzed for cleavage with UVDE. Samples were treated with either CPD or 6-4 photolyase and run through standard Illumina preparation. Dinucleotide bias on the 5' end of sample reads as compared to control dinucleotide bias is shown for yeast (Fig. 3A) and HeLa cells (Fig. 3B). The biased dipyrimidines are outlined in black or comparison.

We decided to troubleshoot our protocol using low doses of UVC light in yeast cells. When the UV dosage is lowered we see a decrease in the percentage of 5' biased ends in our sample libraries below 5000J/m^2. This is due to the lack of sufficiently small double stranded DNA fragments that have dimers on either end. This also leads to an increasing level of background noise from other DNA breaks that are occurring in the cells or during the processing of the DNA. To work around this we developed a circular ligation approach that allows us to map single modifications as well as to remove the bias generated during the PCR step (Fig. 4).

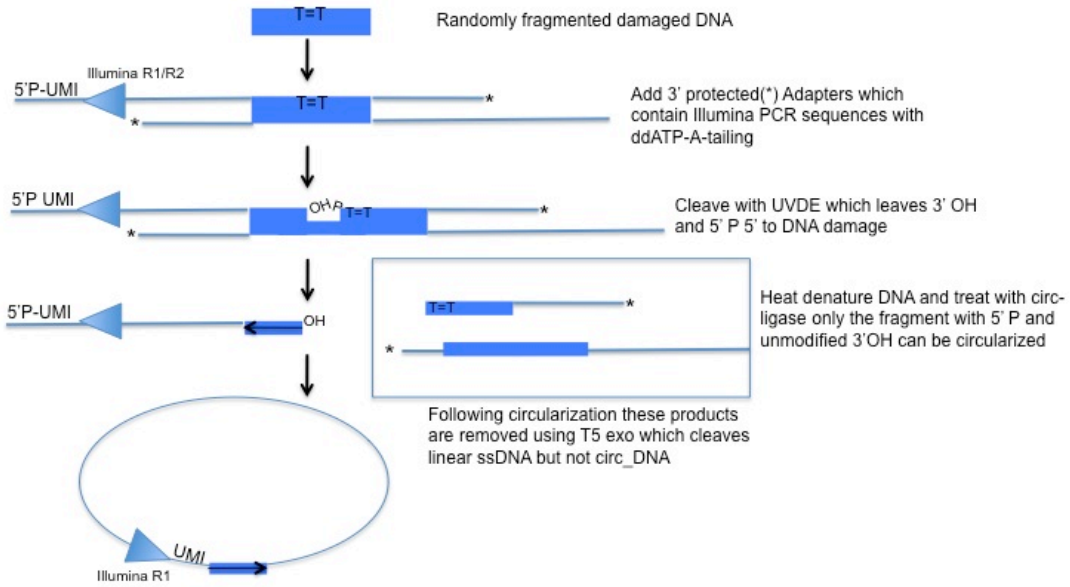

Figure 4.In this protocol damaged DNA is randomly fragmented by bioruption and then modified adapters containing a 3' protection moiety and a 5' unique molecular identifier (UMI) sequence (3) are ligated on with ddATP. The pyrimidine dimer is cleaved with UVDE that creates a new 3' end that is competent for circularization. The DNA is then heat denatured, circularized with circ-ligase, and single stranded DNA is removed with T5 exonuclease. The only strand competent for circularization is the strand with the 5' UMI sequence and the 3'OH generated by UVDE cleavage of the pyrimidine dimer. Once circularized the DNA can be PCR amplified and sequenced with standard Illumina procedures. This protocol generates libraries with a 5'UMI sequence followed by the sites of the pyrimidine dimer. The UMI tag allows for sequences that were replicated during the PCR step to be removed before the data analysis. This is an important step when working with libraries that are in a low abundance because PCR bias is high.

Using this approach we generated libraries for UVC treated yeast cells at dosages of 1000J/m^2 and 20J/m^2 (Fig 5A and B). These libraries showed bias at a lower UV dosage indicating that achieving low dose UVB libraries from human cells would be possible.

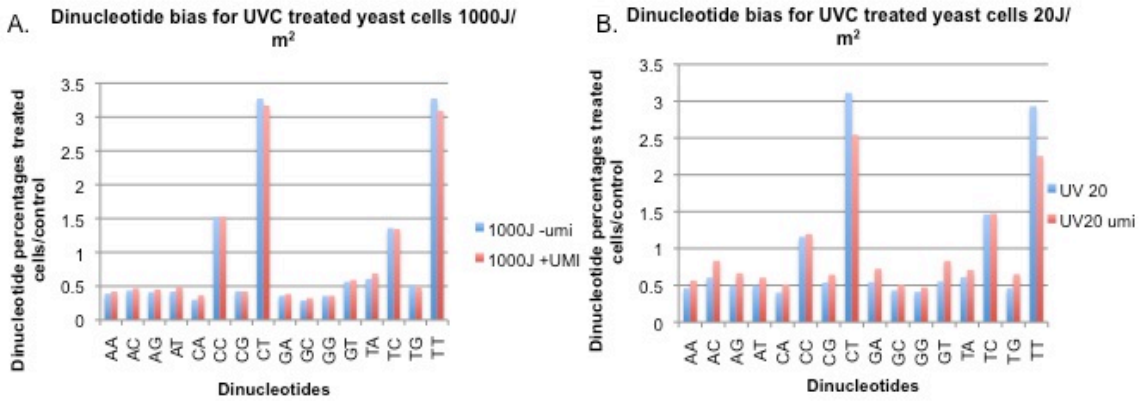

Figure 5. Pyrimidine dimers are enriched at the 5' ends in low dose UV damaged libraries. Yeast cells were irradiated with either 1000J/m^2 or 20J/m^2 of UVC light and DNA was isolated and prepared using the protocol described previously. In all samples we determined the percentage of the dinucleotides at the 5' of libraries between a UV damaged library and the dinucleotides present in genomic DNA. All 4 dinucleotides show enrichment in the UV treated sequencing library. The blue bars indicate the data prior to accounting for the UMI derived PCR bias the red following it.

To start working towards this goal we obtained a UVB light from Coleman and began performing experiments but upon measuring the UV wavelength with a dosimeter we determined that the UV spectrum was quite broad and all 3 wavelengths of UV light were being administered. To address this we obtained an LED bulb from Qphotonics that emits light at 315nm ± 10 nm (4) and incorporated it into a light source that emits UVB at 20J/m^2s. Using this light source with primary keratinocyte cells we were able to show low levels of DNA damage as measured by UVDE cleavage (Fig 6). This mild shearing pattern is obtained because the DNA damage is not saturated enough to yield smaller molecular weight fragments.

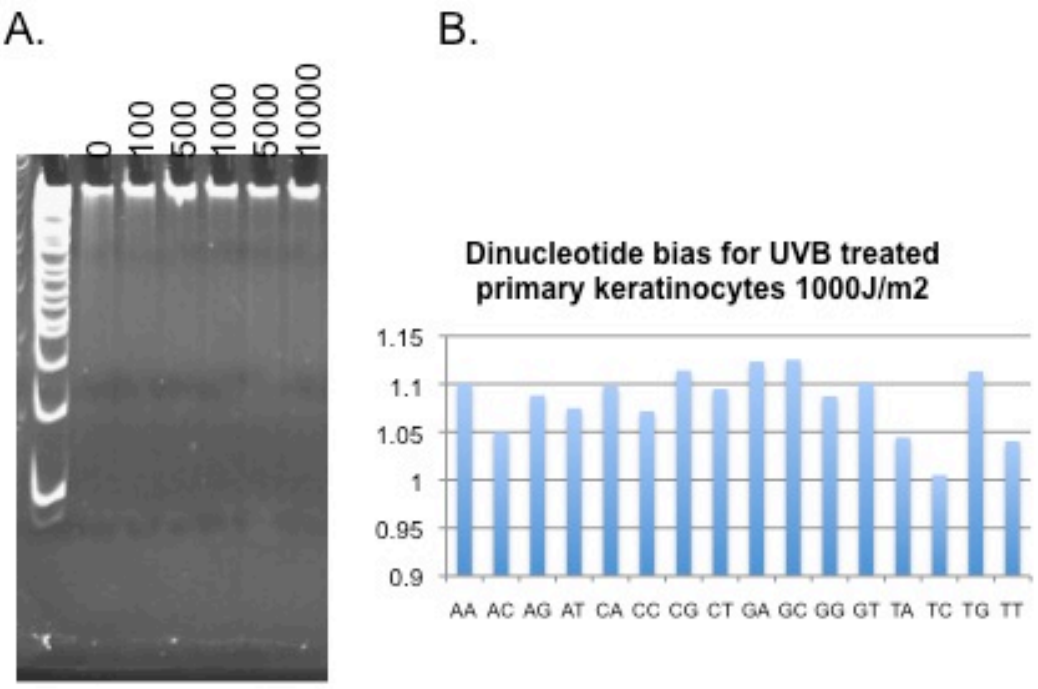

Figure 6. UVB light generates DNA damage that is visible following UVDE cleavage but is unable to form biased pyrimidine libraries. Primary keritinocytes were treated with the various dosages of damage indicated above the gel in $J/m^{2.}$ As the dosage increased the DNA fragmentation was increased in the smaller molecular weight ranges (Fig.6A). This shearing is significantly less than seen with UVC dosages as UVB is 100 fold less damaging (5). When this DNA is made into an Illumina library using the circularization protocol there is no dipyrimidine bias seen (Fig. 6B)

When these samples were used to generate sequencing libraries we were unable to see clear DNA bias in several samples (Fig 6B). We believe this may be due to several causes such as background levels of single stranded breaks present in the DNA, mild shearing during the preparation of the DNA, or inefficient circular ligation. To try to address these issues we can try to enrich our DNA utilizing antibodies specific to pyrimidine dimers using a DNA pull-down approach. Once the modifications sites in our libraries are enriched we will try the circular ligation method to generate libraries for additional Illumina sequencing. Additionally there are more efficient ligation enzymes that can be used with small modifications to the protocol. Using these techniques we hope to be able to map the position of pyrimidine dimers across the genome in human keritinocytes damaged by UVB light. Once this is accomplished we will take these cells and start targeted genome sequencing to determine the mutation frequency in these samples and start more in depth data analysis.

Key research accomplishments:

- Generation of high yield UVDE enzyme
- Generation of low dosage circ-ligase protocol
- Sequencing of low dosage UVC yeast cells
- Obtaining and troubleshooting UVB lamp
- Library preparation using UVB and human keratinocytes

Conclusion

We have obtained large quantities of working enzymes we need to perform future experiments to determine the localization of UV modifications genomewide. Having a source of this enzyme in the lab will allow us the freedom to troubleshoot any problems that we have with this protocol along the way. We are still modifying and changing the protocol as we go to allow us to obtain the biologically relevant libraries we are interested in.

We further analyzed the preliminary data we generated to try to find additional patterns and information we may have missed in the general analysis we performed. We began by using a segmentation approach to look for correlation to known chromatin properties. We were unable to find any strong global correlation to any of the datasets we tried. Upon doing some statistics on our data set we were able to show that we had a relatively low false positive rate of less than 5% of our reads as well as to see that we hit the majority of dipyrimidines in our CPD sample most just a few times. In our 6-4 libraries we hit a much smaller proportion of the total dipyrimidines but hit most of them many times. This may indicate that 6-4 photoproducts might be influenced by local properties of sequence or chromatin context although there are significantly fewer sites so it is possible that these sites are just being amplified in the library upon PCR. We went on to study the sequence content of the bases up and downstream of the dipyrimidines. For the CPD library we saw coverage of all the bases up and downstream of the dipyrimidine at levels expected from the genomic content of the yeast. For the 6-4pp library we saw the base levels as expected for every base except for at the +1 position were we saw a bias for the A residue. This may suggest that 6-4 photoproducts are preferentially created at A containing trinucleotides, or that the X. laevis 6-4 photolyase enzyme preferentially repairs these sites.

We took further steps to obtain the libraries we want as our end product. The first thing we needed was the ability to generate libraries from less damaged DNA. To that end we generated a new circular ligase protocol that allowed us to get down to a low dosage of UVC light that is minimally lethal to yeast cells. This protocol has several benefits including allowing us to map a single modification site instead of relying on multiple nearby sites for cleavage and it contains a UMI tag that allows us to accurately count events as well as compare different samples. We then switched to UVB light that is more biologically relevant since it is the wavelength of light that we receive through the ozone. We obtained a UVB LED bulb that has a very narrow spectrum of light emitted and incorporated that into a system that generates a relatively high dosage output so we can irradiate our cells over a few minutes. This was important to try to decrease the amount of time to try to reduce the background events that are going on in the cell. Finally we started working with a primary keratinocyte cell line that we feel most accurately represents the human skin. We were able to show that these cells showed damage when treated with UVB but have been unable to generate a sequencing library from these cells. We hope that with more troubleshooting

we will be able to achieve these libraries and begin to address our questions about the localization of these modifications genomewide.

We have generated a method to study the DNA modifications caused by exposure to UV light. We have shown that these libraries from high doses can discern between CPD and 6-4 photoproducts and have found a novel preference for 6-4 photoproducts to either form at YYA sequences or for light repair to happen preferentially there. These new methods for studying genomewide distribution of UV modification may bring clarity to the relationship between UV DNA modification and mutation. We hope that with this new knowledge will come advancements in the prevention and treatment of skin cancer.

Publications, Abstracts, and Presentations

- **Manuscripts, abstracts, presentations**;

  D. Bryan*, M. Ransom*, B. Adane, K. York, J. Hesselberth, High resolution mapping of modified DNA nucleobases using excision repair enzymes, Genome Res. 24:1534.

  Ransom, M and Hesselberth, J. Poster session presented at: Cell Biology and Molecular Oncology Retreat; 2013. Oct. 21; Aurora CO.

  Ransom, M and Hesselberth, J. Poster session presented at: 5th annual postdoctoral research day; 2014 Mar. 14; Aurora, CO.

Licenses applied for and/or issued; nothing to report

Reportable outcomes: nothing to report

Other Achievements:

- **Degrees obtained that are supported by this award**; none
  - Coursera courses
    - Learn to Program; The fundamentals
    - The Data Scientist's Toolbox
    - Programming for Everyone (Python)
    - R programming
    - Getting and Cleaning Data
- **Development of cell lines, tissue or serum repositories**; none
- **Informatics such as databases and animal models**, etc;
  - GSE51361-NCBI Gene Expression Omnibus
- **Funding applied for based on work supported by this award**; none
- **Employment or research opportunities applied for and/or received based on experience/training supported by this award**; none

References:

1. A. M. Avery et al., Nucleic Acids Res 27, 2256 (1999).
2. D.S. Bryan et al., Genome Res in press (2014).
3. T. Kivioja et al., Nat Methods 9, 72 (2011).
4. QPhotonics, LLC, 2000. Web. 6 Aug. 2014.
5. A. Besaratinia et al., FASEB J 9, 3079 (2011)

Appendices: Nothing to Report