Bioinformatic Prediction of Ultraviolet Light Mutagenesis Sensitivity of Human Genes and a Method for Genetically Engineering UVB Resistance

Kevin A. Lease¹ and Chris Papageorgio²

¹Department of Internal Medicine, Education Office, One Hospital Drive, MA419, University of Missouri 65212, USA.
²SSM Cancer Care, 6400 Clayton Rd. Ste. 302, Saint Louis, MO 63117, USA.
Corresponding author email: christos_papageorgiou@ssmhc.com

Abstract: Living on earth, we are exposed to ultraviolet (UV) light as part of the solar radiation. UVB spectrum light exposure contributes to the development of skin cancer by interacting with pyrimidine pairs to create lesions called cyclobutane pyrimidine dimers. If these lesions are not removed by nucleotide excision repair, they often give rise to C to T transition mutations. Based on these observations, a bioinformatics approach was used to predict the vulnerability of human protein coding genes to UVB induced loss of function mutations. This data was used to evaluate in depth those genes associated with malignant melanoma. In addition, we demonstrate a method of genetically engineering genes that significantly improves resistance to UVB loss of function mutations.

Keywords: melanoma, ultraviolet light, mutagenesis, skin cancer
Introduction

Skin cancers are the most common primary type of cancer. The three most common primary types of cutaneous cancer are squamous cell carcinoma, basal cell carcinoma, and malignant melanoma. Melanoma is of particular interest since it tends to occur in younger adults, and has a relatively poor prognosis if metastasis has occurred. Human epidemiological studies as well as lab animal experiments indicate that UV exposure is an important etiological factor in all three types of skin cancer.

The spectrum of solar UV light can be divided by wavelength into UVA, UVB, and UVC. UVB spectrum light exposure is thought to explain 80%–90% of the mutagenic effects of sunlight. DNA sequence analysis has revealed the target and effect of UVB exposure, which has come to be known as the ‘UVB signature’ mutation. The targets of UVB in DNA are pyrimidine pairs containing cytosine, ie, CC, CT, or TC. Following UVB exposure, these pyrimidine pairs give rise to either C-T point mutations or CC-TT dinucleotide changes.

The mechanism by which UVB exposure causes this pattern of mutagenesis is partially understood. UVB energy can be absorbed by pyrimidines and can result in two new intrastrand covalent bonds formed between adjacent pyrimidines. This structure is known as a cyclobutane pyrimidine dimer (CPD). How the CPD gives rise to the observed mutations is unclear. One model supposes that a DNA polymerase introduces the incorrect nucleotide when using the CPD containing strand as the template. Another model supposes that at least one of the cytosines in the CPD are methylated and that the CPD structure increases the basal spontaneous deamination of 5-methylcytosine to thymidine.

However, not every CPD that forms will necessarily give rise to a mutation. CPD’s create a distortion in the DNA that can potentially be recognized by the nucleotide excision repair (NER) enzymes. These enzymes can excise part of the strand flanking a CPD and use the complementary strand as a template to replace the excised sequence. The importance of this repair pathway is highlighted by Xeroderma pigmentosum patients who exhibit increased frequency of skin cancers.

To sum up, when cells are exposed to UVB, UVB signature mutations may occur. If these genetic alterations perturb key regulators of the cell cycle or apoptosis, skin cancer may develop. Several genetic loci have been associated with malignant melanoma including Tumor Protein 53 (TP53), Cyclin-dependent Kinase 4 (CDK4), Cyclin-dependent Kinase Inhibitor 2A (CDKN2A), Melanocortin 1 Receptor (MC1R), Microphthalmia-associated Transcription Factor (MITF), v-Kit Hardy-Zuckerman 4 Feline Sarcoma Viral Oncogene Homolog (KIT), v-Raf Murine Sarcoma Viral Oncogene Homolog B (BRAF), Neuroblastoma RAS viral Oncogene Homolog (NRAS), B-cell Lymphoma 2 (BCL-2), Apoptotic Protease Activating Factor 1 (APAF1), v-AKT Murine Thymoma Viral Oncogene Homolog (AKT), Phosphatase and Tensin Homolog (PTEN), and FK506 Binding Protein 12-Rapamycin Associated Protein 1 (FRAP1).

TP53 encodes the transcriptional regulator known as p53. Following genetic damage or other cellular stresses, p53 becomes stabilized and can arrest the cell cycle or induce apoptosis. Mutations in p53 are found in up to 5% of primary melanomas.

CDK4 is a protein kinase that phosphorylates Retinoblastoma (Rb) to enable G1-S cell cycle progression. Familial melanoma has been associated with lesions in CDK4. The CDKN2A locus encodes two isoforms involved in melanoma, p16INK4a and p14ARF. p16INK4a binds to CDK4 and inhibits cyclin D binding. Binding of p14ARF with Murine Double Minute 2 (MDM2) results in stabilization of p53. Loss-of-function mutations in CDKN2A cause increased CDK4 activity and a predisposition to malignant melanoma. Mutations in CDK2NA are the most common cause of familial melanoma, but can also occur sporadically.

MC1R is a seven transmembrane G protein coupled cell surface receptor for melanocyte-stimulating hormone (MSH-alpha). Receptor activation results in a signaling cascade resulting in expression of MITF and other genes neccesary for eumelanin synthesis. MC1R polymorphisms are associated with melanoma. MITF is a transcription factor that is a master regulator of melanocyte development, survival and melanogenesis. Amplification of MITF has been found in melanoma.

The tyrosine kinase KIT is required for the developmental migration and proliferation of melanocytes. Upon binding its ligand Stem-cell Factor, KIT
activates a mitogen-activated protein kinase (MAPK) signaling cascade. Activating lesions in KIT have been reported in melanoma. NRAS is a member of the RAS family of plasma membrane bound small G proteins that can be activated by receptor tyrosine kinases. When in its GTP bound form, RAS can activate RAF. RAF is a protein kinase that is part of a MAPK signaling cascade that promotes cellular proliferation. NRAS is mutated in 33% of primary melanomas, and associated with increased activity. BRAF is mutated in about two thirds of melanomas, typically causing increased activity.

BCL-2 is primarily found in the outer membrane of mitochondria where it acts to prevent release of apoptosis mediators from the mitochondrial intermembrane space. The expression of BCL-2 may correlate with melanoma progression and BCL-2 antisense therapy has been shown to increase melanoma chemosensitivity in a mouse model.

APAF1 forms apoptosomes with cytosolic cytochrome c. This complex activates caspase-9, resulting in programmed cell death. APAF1 may undergo loss of heterozygosity as well as decreased expression due to methylation in melanoma.

PTEN encodes a lipid and protein phosphatase that regulates levels of phosphatidylinositol (3,4,5)-triphosphate (PIP3) by dephosphorylation. Thus, decreased PTEN activity causes an increase in PIP3. PIP3 recruits the serine/threonine protein kinase AKT3 to the cell membrane where it is becomes activated. Through phosphorylation of various substrates, AKT3 inhibits apoptosis. AKT3 activity is increased in up to 60% of nonfamilial melanomas due to either increased copies of AKT3 or loss of PTEN.

FRAP1, also known as mammalian Target of Rapamycin (mTOR), is a protein kinase that is found in two different complexes called mTORC1 and mTORC2. mTORC1 promotes protein translation, favoring cell proliferation. mTORC2 activates AKT by phosphorylation. Increased mTOR activity was observed in 73% of melanomas.

Despite the known importance of UVB sites in the etiology of skin cancers, there has been little work characterizing their distribution in the human genome. One goal of this work was to have a genome wide analysis of UVB sites with respect to coding sequences. Furthermore, we investigated the degree to which these UVB sites would be deleterious if mutated. In addition, we explored the relationship between UVB sites and genes implicated in malignant melanoma. One supposition is that these coding sequences would have an overabundance of UVB sites that may predispose them to mutation. Another possibility is that the UVB sites found in these coding sequences are present in a sequence context that is more deleterious than that of the average gene.

Finally, we propose a novel approach for genetically modifying skin cancer susceptibility. Given that UVB sites are the target of UVB mutagenesis, decreasing the frequency of these sites should decrease the susceptibility of a coding sequence to UVB mutation. We propose a method for decreasing the frequency of UVB sites while maintaining the wildtype primary amino acid sequence by the rational use of synonymous codon substitution. In this work, we bioinformatically predict the extent to which this method can improve UVB resistance of coding sequences.

Methods

An algorithm was developed which first identifies potential UVB sites in human protein coding sequences, individually introduces UVB signature mutations at these sites, and then predicts the consequence to the protein product. The UVB signature target site is defined as any of the dinucleotides CC, CT, and TC (and GG, AG, and GA considering the change may occur on the complementary strand). The algorithm identifies every UVB site in translated exonic genomic sequences. In addition, UVB sites may potentially be found at the junctions of exons, therefore, when extracting an exon from genomic sequence, we also took one additional nucleotide from each side of an exon in the analysis to not miss those instances where the UVB site straddled the junction. These flanking nucleotides were trimmed after screening for UVB sites in the coding sequence.

The algorithm then individually creates a series of UVB single point mutants of the coding sequence by introducing single transition mutations (C->T or G->A) at each site. For CC and GG dinucleotide sites, there would be two separate single transition mutant sequences generated. In addition, UVB doublet mutants of coding sequences were created at dinucleotides CC (and GG considering the change may...
occur on the complementary strand) by introducing two transitions mutations (CC->TT or GG->AA). Doublet mutants at CC and GG dinucleotides were not created if they straddled two codons (corresponding to phase 3 position of the first codon and phase one position of the subsequent codon) because this information was already captured in the two single mutants created.

For each mutant, the BLOSUM62 amino acid substitution matrix was then used to quantitatively assess the consequence of the altered codon when translated.31 BLOSUM62 uses a log odds ratio score and awards negative scores to unfavorable amino acid substitutions.32 The sum score for all unfavorable mutations for a gene was computed to get an overall amino acid change score representing the UV sensitivity. At left, boxes are drawn around UVB sites. Below the double stranded DNA is the single letter amino acid sequence for each codon of the upper strand. In the middle, nucleotides that have undergone mutation are shown in bold. The translated sequence is shown below the DNA. The BLOSUM62 substitution matrix score for each new amino acid compared to its original amino acid is shown at the bottom. At right, the sum of the negative BLOSUM62 scores is given as the total amino acid change score for this sequence.

Table 1. UVB characteristics of human coding sequences.

| Description                                      | Mean ± SD       |
|--------------------------------------------------|-----------------|
| Mean % of dinucleotides that are UVB sites ±SD   | 44.5% ± 5.2%    |
| Mean % of UVB sites with single nt changes that were deleterious ±SD | 29.8 ± 5.8 |
| Mean % of UVB sites with double nt changes that were deleterious ±SD | 67.9 ± 8.5 |
| Mean amino acid change score per single nt deleterious mutation ±SD | -2.16 ± 0.14 |
| Mean amino acid change score per double nt deleterious mutation ±SD | -2.69 ± 0.21 |

K = > AAA, D = > GAT, E = > GAA, C = > TGT, W = > TGG, R = > CGT, G = > GGT.

GC percentage, length, and dinucleotide composition were calculated for each coding sequence. The output from the perl script was parsed into an Excel spreadsheet for ease of viewing. The data was imported into R for statistical analysis using the Bioconductor package multtest.33,34 P values were corrected for multiple measures using the Benjamini and Hochberg method.35 Mean GC percentage was calculated as the arithmetic mean of each CDS GC percentage.

Chi squared tests were used to estimate P values for observed number of UVB sites per CDS, as well as for observed frequency of UVB sites that were deleterious. The expected number of UVB signature sites was calculated as the product of (0.3918 or 0.445) (length of CDS in nucleotides(nt) –1). The number 0.3918

K = > AAA, D = > GAT, E = > GAA, C = > TGT, W = > TGG, R = > CGT, G = > GGT.

GC percentage, length, and dinucleotide composition were calculated for each coding sequence. The output from the perl script was parsed into an Excel spreadsheet for ease of viewing. The data was imported into R for statistical analysis using the Bioconductor package multtest.33,34 P values were corrected for multiple measures using the Benjamini and Hochberg method.35 Mean GC percentage was calculated as the arithmetic mean of each CDS GC percentage.

Chi squared tests were used to estimate P values for observed number of UVB sites per CDS, as well as for observed frequency of UVB sites that were deleterious. The expected number of UVB signature sites was calculated as the product of (0.3918 or 0.445) (length of CDS in nucleotides(nt) –1). The number 0.3918
UV sensitivity of human genes in cancer

To analyze average severity of single deleterious mutations, a control sample of the first 65000 negative BLOSUM62 single amino acid scores were collected from chromosome 5, which had a mean of $-2.14$. This set of scores was used for comparison of the single mutation negative scores from each melanoma related coding sequence two tailed Student’s t-test with equal variances.

Results

We investigated the frequency of UVB sites in human coding sequences. At the outset, we expected UVB sites to be found $3/8$ (or 38.75%) of the time for any dinucleotide of sequence considered. This assumption was based on probability with equal

Table 2. Analysis of UVB site frequencies in the coding sequences of genes implicated in malignant melanoma.

| Gene       | Protein       | Observed | Expected | P value |
|------------|---------------|----------|----------|---------|
| FRAP1      | NP_004949.1   | 3344     | 3403.8   | 0.236 |
| NRAS       | NP_002515.1   | 204      | 253.21   | 1.22E-04 |
| AKT3       | NP_859029.1   | 535      | 621.66   | 1.38E-05 |
| AKT3       | NP_005456.1   | 559      | 640.36   | 6.22E-05 |
| MITF       | NP_937802.1   | 647      | 695.09   | 2.80E-02 |
| MITF       | NP_006713.1   | 643      | 693.75   | 1.98E-02 |
| MITF       | NP_937820.1   | 623      | 673.73   | 1.81E-02 |
| MITF       | NP_937821.1   | 605      | 661.72   | 7.20E-03 |
| MITF       | NP_000239.1   | 504      | 560.25   | 3.62E-03 |
| MITF       | NP_937801.1   | 500      | 552.25   | 6.70E-03 |
| KIT        | NP_000213.1   | 1122     | 1303.85  | 1.53E-10 |
| KIT        | NP_001087241.1| 1120     | 1298.51  | 3.11E-10 |
| BRAF       | NP_004324.2   | 924      | 1023.5   | 1.11E-04 |
| CDKN2A     | NP_000681.1   | 217      | 209.15   | 0.550 |
| CDKN2A     | NP_478102.1   | 242      | 231.84   | 0.456 |
| CDKN2A     | NP_478104.2   | 186      | 155.75   | 2.96E-03 |
| PTEN       | NP_00305.3    | 417      | 538.89   | 2.32E-11 |
| CDK4       | NP_000066.1   | 421      | 405.39   | 0.381 |
| APAF1      | NP_863651.1   | 1425     | 1666.97  | 3.56E-14 |
| APAF1      | NP_863658.1   | 1378     | 1609.57  | 1.70E-13 |
| APAF1      | NP_037361.1   | 1415     | 1652.29  | 8.82E-14 |
| APAF1      | NP_001151.1   | 1368     | 1594.88  | 4.13E-13 |
| APAF1      | NP_863659.1   | 404      | 452.12   | 5.73E-03 |
| TP53       | NP_001119584.1| 562      | 525.54   | 0.0576 |
| TP53       | NP_000537.3   | 562      | 525.54   | 0.0576 |
| TP53       | NP_001119587.1| 365      | 349.32   | 0.340 |
| TP53       | NP_001119585.1| 476      | 462.8    | 0.495 |
| TP53       | NP_001119589.1| 279      | 286.58   | 0.626 |
| TP53       | NP_001119586.1| 471      | 456.12   | 0.435 |
| TP53       | NP_001119588.1| 274      | 279.91   | 0.707 |
| BCL2       | NP_000624.2   | 345      | 319.95   | 0.0978 |
| BCL2       | NP_000648.2   | 296      | 274.56   | 0.128 |
frequency (25%) of the four nucleotides. This expectation was modified based on the average GC composition of the coding sequences of 53.48% to arrive at a 39.18% frequency of a UVB site at any given dinucleotide to be expected by chance. Interestingly, the observed mean frequency of UVB dinucleotides among all coding sequences was 44.5% ± 5.2%, which was significantly greater than expected ($P < 2.2E-16$).

On average, only about 30% of the potential single transitions would result in an unfavorable amino acid substitution (Table 1). Considering doublet UVB mutations at positions 1–2 or 2–3 of codons, about 68% of these would give rise to unfavorable amino acid changes. Unfavorable mutations at doublet sites also tended to be more deleterious substitutions with a mean BLOSUM62 score of −2.69 per residue change compared to −2.16 at single unfavorable sites.

The number of UVB sites in human coding sequences could be important for susceptibility to UVB mutagenesis. We counted the number of UVB sites in human coding sequences and observed that the number increases linearly as a function of the length of the coding sequence ($R^2 = 0.988$) (Fig. 2). In addition, based on our model, the calculated amino acid change score grows more negative in a linear fashion with respect to length ($R^2 = −0.939$) (data not shown). Thus, longer open reading frames are predicted to be more susceptible to UVB. However, neither the number of UVB sites nor the amino acid change score correlated with the GC percentage of

Table 3. Frequency of single site UVB mutations being deleterious in the coding sequences of genes implicated in malignant melanoma.

| Gene | Protein | Observed | Expected | P value |
|------|---------|----------|----------|---------|
| FRAP1| NP_004949.1 | 812 | 978.63 | 4.69E-09 |
| NRAS | NP_002515.1 | 47 | 59.3 | 0.124 |
| AKT3 | NP_859029.1 | 147 | 133.8 | 0.292 |
| AKT3 | NP_005456.1 | 150 | 139.17 | 0.410 |
| MITF | NP_937802.1 | 175 | 185.36 | 0.505 |
| MITF | NP_00671.1 | 171 | 182.38 | 0.454 |
| MITF | NP_937820.1 | 165 | 177.31 | 0.406 |
| MITF | NP_937821.1 | 164 | 173.44 | 0.533 |
| MITF | NP_000239.1 | 135 | 143.34 | 0.547 |
| MITF | NP_937801.1 | 133 | 142.15 | 0.501 |
| KIT | NP_000213.1 | 317 | 302.77 | 0.470 |
| KIT | NP_001087241.1 | 316 | 302.17 | 0.484 |
| BRAF | NP_004324.2 | 311 | 254.49 | 1.93E-04 |
| CDKN2A | NP_000068.1 | 60 | 75.99 | 0.0719 |
| CDKN2A | NP_478102.1 | 105 | 81.95 | 9.62E-03 |
| CDKN2A | NP_478104.2 | 62 | 57.22 | 0.588 |
| PTEN | NP_000305.3 | 112 | 108.17 | 0.764 |
| CDK4 | NP_000066.1 | 123 | 122.48 | 0.972 |
| APAF1 | NP_863651.1 | 368 | 365.35 | 0.915 |
| APAF1 | NP_863658.1 | 356 | 350.45 | 0.813 |
| APAF1 | NP_037361.1 | 362 | 362.37 | 0.988 |
| APAF1 | NP_001151.1 | 350 | 347.47 | 0.916 |
| APAF1 | NP_863659.1 | 106 | 101.02 | 0.678 |
| TP53 | NP_00119584.1 | 193 | 163.01 | 1.79E-02 |
| TP53 | NP_000537.3 | 193 | 163.01 | 1.79E-02 |
| TP53 | NP_00119587.1 | 112 | 106.98 | 0.686 |
| TP53 | NP_00119585.1 | 171 | 138.57 | 4.73E-03 |
| TP53 | NP_00119589.1 | 90 | 82.55 | 0.468 |
| TP53 | NP_00119586.1 | 170 | 137.38 | 4.28E-03 |
| TP53 | NP_00119588.1 | 89 | 81.35 | 0.451 |
| BCL2 | NP_000624.2 | 115 | 111.45 | 0.785 |
| BCL2 | NP_000648.2 | 103 | 97.15 | 0.613 |
the coding sequence ($R^2 = -0.005$ and $R^2 = -0.063$) (data not shown).

We investigated whether genes associated with malignant melanoma would have a significantly different number of UVB sites than would be expected from the average frequency of UVB sites. Seven of the genes had significantly fewer than expected UVB sites, while four were not significantly different from the average gene. One isoform of CDK2NA was found to have a significantly increased number of UVB sites (Table 2).

It could be that a greater proportion of the UVB sites in genes implicated in malignant melanoma could give rise to deleterious substitutions if mutated. The predicted proportion of UVB site mutations causing a deleterious change was evaluated in the genes linked to malignant melanoma (Tables 3 and 4). A subset of those coding sequences did have a significantly increased proportion of sites predicted to produce unfavorable changes, including four of seven coding sequences from TP53, one of the CDK2NA isoforms and BRAF. FRAP1 had a smaller proportion of sites that were unfavorable.

We studied whether the UVB sites in coding sequences of genes implicated in melanoma would on average cause more deleterious changes than those found in the average coding sequence. This would be reflected by more negative BLOSUM62 scores. The predicted mean severity of UVB deleterious mutations was compared in genes linked

| Table 4. Frequency of deleterious double site UVB mutations in the coding sequences of genes implicated in malignant melanoma. |
|---|---|---|---|
| Gene | Protein | Observed | Expected | $P$ value |
| FRAP1 | NP_004949.1 | 374 | 406.04 | 0.0483 |
| NRAS | NP_002515.1 | 20 | 23.77 | 0.41952 |
| AKT3 | NP_859029.1 | 57 | 53.64 | 0.66931 |
| AKT3 | NP_005456.1 | 60 | 56.36 | 0.6467 |
| MITF | NP_937820.1 | 97 | 86.23 | 0.18206 |
| MITF | NP_006713.1 | 90 | 82.16 | 0.35182 |
| MITF | NP_937820.1 | 89 | 80.8 | 0.31838 |
| MITF | NP_937821.1 | 89 | 80.8 | 0.31838 |
| MITF | NP_000239.1 | 70 | 63.83 | 0.41952 |
| MITF | NP_937801.1 | 68 | 62.47 | 0.47403 |
| KIT | NP_000213.1 | 143 | 136.48 | 0.58688 |
| KIT | NP_00187241.1 | 143 | 135.8 | 0.5377 |
| BRAF | NP_004324.2 | 130 | 124.94 | 0.67396 |
| CDKN2A | NP_000668.1 | 30 | 40.74 | 0.03377 |
| CDKN2A | NP_478102.1 | 51 | 55.68 | 0.5311 |
| CDKN2A | NP_478104.2 | 34 | 35.99 | 0.77353 |
| PTEN | NP_000305.3 | 46 | 40.74 | 0.38021 |
| CDK4 | NP_000666.1 | 59 | 57.72 | 0.89452 |
| APAF1 | NP_863651.1 | 136 | 126.97 | 0.39808 |
| APAF1 | NP_863658.1 | 131 | 122.22 | 0.40319 |
| APAF1 | NP_037361.1 | 135 | 125.62 | 0.37138 |
| APAF1 | NP_001151.1 | 130 | 120.86 | 0.37571 |
| APAF1 | NP_863659.1 | 46 | 40.74 | 0.38021 |
| TP53 | NP_00119584.1 | 107 | 92.34 | 0.05998 |
| TP53 | NP_000537.3 | 107 | 92.34 | 0.05998 |
| TP53 | NP_00119587.1 | 64 | 57.04 | 0.31334 |
| TP53 | NP_00119585.1 | 92 | 79.44 | 0.08774 |
| TP53 | NP_00119589.1 | 49 | 44.14 | 0.44857 |
| TP53 | NP_00119586.1 | 92 | 79.44 | 0.08774 |
| TP53 | NP_00119588.1 | 49 | 44.14 | 0.44857 |
| BCL2 | NP_000624.2 | 65 | 68.58 | 0.69252 |
| BCL2 | NP_000648.2 | 58 | 61.11 | 0.72115 |
to melanoma and a large control group (Table 5). Mutations at UVB sites in the coding sequences of \textit{APAF1} and \textit{FRAP1} were predicted to be significantly more deleterious.

Finally, we explored the use of rational synonymous codon substitution to alter the coding sequence without changing the primary amino acid sequence in order to decrease the predicted vulnerability to UVB (Table 6). This method was found to decrease the number of UVB sites by almost 50% and improve the amino acid change score for the coding sequences of all genes associated with malignant melanoma by about 25%.

Table 5. Mean severity of deleterious single mutations in the coding sequences of genes implicated in malignant melanoma.

| Gene       | Protein     | Mean severity | P value     |
|------------|-------------|---------------|-------------|
| \textit{FRAP1} | NP_004949.1 | -2.33         | 8.96E-05    |
| \textit{NRAS} | NP_002515.1 | -1.98         | 0.682       |
| \textit{AKT3} | NP_859029.1 | -2.23         | 0.738       |
| \textit{AKT3} | NP_005456.1 | -2.21         | 0.831       |
| \textit{MITF} | NP_937802.1 | -2.05         | 0.682       |
| \textit{MITF} | NP_006713.1 | -2.02         | 0.549       |
| \textit{MITF} | NP_937820.1 | -2.05         | 0.682       |
| \textit{MITF} | NP_937821.1 | -2.07         | 0.750       |
| \textit{MITF} | NP_000239.1 | -2.04         | 0.682       |
| \textit{MITF} | NP_937801.1 | -2.04         | 0.682       |
| \textit{KIT} | NP_000213.1 | -2.1          | 0.831       |
| \textit{KIT} | NP_001087241.1 | -2.1 | 0.831 |
| \textit{BRAF} | NP_004324.2 | -2.16         | 0.969       |
| \textit{CDKN2A} | NP_000068.1 | -2.15         | 0.979       |
| \textit{CDKN2A} | NP_478102.1 | -2.03         | 0.682       |
| \textit{CDKN2A} | NP_478104.2 | -2.19         | 0.960       |
| \textit{PTEN} | NP_000305.3 | -2.16         | 0.979       |
| \textit{CDK4} | NP_000066.1 | -2.2          | 0.831       |
| \textit{APAF1} | NP_863651.1 | -2.34         | 3.49E-03    |
| \textit{APAF1} | NP_863651.1 | -2.35         | 2.56E-03    |
| \textit{APAF1} | NP_037361.1 | -2.35         | 2.56E-03    |
| \textit{APAF1} | NP_001151.1 | -2.36         | 2.56E-03    |
| \textit{APAF1} | NP_863659.1 | -2.32         | 0.386       |
| \textit{TP53} | NP_001119584.1 | -2.13 | 0.979 |
| \textit{TP53} | NP_000537.3 | -2.13         | 0.979       |
| \textit{TP53} | NP_001119587.1 | -2.11 | 0.965 |
| \textit{TP53} | NP_001119585.1 | -2.14 | 0.995 |
| \textit{TP53} | NP_001119589.1 | -2.11 | 0.969 |
| \textit{TP53} | NP_001119586.1 | -2.14 | 0.979 |
| \textit{TP53} | NP_001119588.1 | -2.1 | 0.965 |
| \textit{BCL2} | NP_000624.2 | -2.2          | 0.831       |
| \textit{BCL2} | NP_000648.2 | -2.2          | 0.831       |

**Discussion**

Despite the importance of UVB sites in skin cancer formation, the extent of these targets and their potential for deleterious mutations has not been previously investigated on a genome-wide basis. We found that these sites are quite prevalent in human coding sequences and are predicted to have significant deleterious potential. However, it was surprising that the majority of the time, the predicted consequence of any one single transition mutation at a UVB site is not unfavorable. It is interesting to note that the average coding sequence has more UVB sites than would be expected at random. This may be a consequence of codon usage biases or it may reflect an adaptive mechanism to facilitate UVB molecular mutation.

As UVB is a major causative factor in skin cancer and leads to UVB signature mutations, we hypothesized that the sequence properties of genes may influence their likelihood to suffer from UVB induced loss of function mutations. That is, certain genes would have sequences that either contain more UVB targets or the targets tend to occur in a more sensitive context that would lead to unfavorable mutations in either greater frequency or severity or some combination thereof. Surprisingly, most genes associated with melanoma have a fewer than expected number of UVB sites. This may reflect a selective process for these genes to avoid UVB mutagenesis. However, our data suggests that all three mechanisms may be at play depending on the coding sequence. Some coding sequences have an increased number of UVB sites, some have an increased frequency of UVB sites that are deleterious and some tend to have more deleterious changes on average when mutated. These qualities may be useful in screening for additional genes important in melanoma, to identify additional diagnostic and therapeutic molecular targets.

Metastatic melanoma is associated with a 10 year survival rate of less than 10%. In addition, the frequency of melanoma is increasing and over 60,000 new cases occur each year in the United States. Therefore, the need for new methods for primary prevention of melanoma or in its treatment are greater than ever. We have described a new method for decreasing the susceptibility of coding sequences to UV, much like sunscreen for genes. Although currently preclinical, cutaneous
gene therapy has promise. As gene therapy advances, genetically engineering decreased vulnerability of a gene to UVB may be useful therapeutically in the treatment of this devastating disease.

Acknowledgements
The authors would like to thank the anonymous reviewers for their comments and suggestions.

Disclosures
This manuscript has been read and approved by all authors. This paper is unique and not under consideration by any other publication and has not been published elsewhere. The authors and peer reviewers report no conflicts of interest. The authors confirm that they have permission to reproduce any copyrighted material.

References
1. Ibrahim N, Haluska FG. Molecular pathogenesis of cutaneous melanocytic neoplasms. Ann Rev Pathol Mech Dis. 2009;4:551–79.
2. Pfeifer GP, You Y, Besaratinia A. Mutations induced by ultraviolet light. Mutation Research. 2005;571:19–31.
3. Wikondahl NM, Brash DE. Ultraviolet radiation induced signature mutations in photocarcinogenesis. J Invest Dermatol Symp Proc. 1999;4:6–10.
4. Nouspikel T. DNA repair in mammalian cells: nucleotide excision repair: variations on versatility. Cell Mol Life Sci. 2009;66:994–1009.
5. Vogelstein B, Lane D, Levine AJ. Surfing the p53 network. Nature. 2000;408:307–10.
6. Akslen LA, Monstad SE, Larsen B, Straume O, Ogreid D. Frequent mutations of the p53 gene in cutaneous melanomas of the nodular type. Int J Cancer. 1998;79:91–5.
Supplementary Data
Collection of output files from perl script are available at http://bengal.missouri.edu/kale0a.
Perl script: Mel15.pl.
Excel file of data in a tabular format: mel_table.xls.