

SHORT REPORT

OPEN ACCESS

Full open access to this and thousands of other papers at <http://www.la-press.com>.

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

Cancer Informatics 2011:10 121–131

doi: [10.4137/CIN.S6670](https://doi.org/10.4137/CIN.S6670)

This article is available from <http://www.la-press.com>.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article. Unrestricted non-commercial use is permitted provided the original work is properly cited.



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.¹ 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.²

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.² DNA sequence analysis has revealed the target and effect of UVB exposure, which has come to be known as the ‘UVB signature’ mutation.³ 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.²

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.⁴ 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.²

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.⁴ 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.²

To sum up, when cells are exposed to UVB, UVB signature mutations may occur.³ 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).¹

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.⁵ Mutations in p53 are found in up to 5% of primary melanomas.⁶

CDK4 is a protein kinase that phosphorylates Retinoblastoma (Rb) to enable G1-S cell cycle progression.⁷ Familial melanoma has been associated with lesions in CDK4.⁸ The *CDKN2A* locus encodes two isoforms involved in melanoma, p16INK4a and p14ARF.¹ p16INK4a binds to CDK4 and inhibits cyclin D binding.⁹ Binding of p14ARF with Murine Double Minute 2 (MDM2) results in stabilization of p53.⁵ Loss-of-function mutations in *CDKN2A* cause increased CDK4 activity and a predisposition to malignant melanoma.^{1,8} Mutations in *CDKN2A* are the most common cause of familial melanoma, but can also occur sporadically.¹

MC1R is a seven transmembrane G protein coupled cell surface receptor for melanocyte-stimulating hormone (MSH-alpha).¹⁰ Receptor activation results in a signaling cascade resulting in expression of MITF and other genes necessary for eumelanin synthesis.¹¹ MC1R polymorphisms are associated with melanoma.¹² MITF is a transcription factor that is a master regulator of melanocyte development, survival and melanogenesis.¹³ Amplification of *MITF* has been found in melanoma.¹⁴

The tyrosine kinase KIT is required for the developmental migration and proliferation of melanocytes.¹⁵ Upon binding its ligand Stem-cell Factor, KIT



activates a mitogen-activated protein kinase (MAPK) signaling cascade.¹⁶ Activating lesions in *KIT* have been reported in melanoma.¹⁷

NRAS is a member of the RAS family of plasma membrane bound small G proteins that can be activated by receptor tyrosine kinases.¹⁸ When in its GTP bound form, RAS can activate RAF.¹⁹ RAF is a protein kinase that is part of a MAPK signaling cascade that promotes cellular proliferation.²⁰ NRAS is mutated in 33% of primary melanomas, and associated with increased activity.¹⁸ BRAF is mutated in about two thirds of melanomas, typically causing increased activity.¹

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.²¹ 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.^{22,23}

APAF1 forms apoptosomes with cytosolic cytochrome c.²⁴ 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.²⁵

PTEN encodes a lipid and protein phosphatase that regulates levels of phosphatidylinositol (3,4,5)-triphosphate (PIP3) by dephosphorylation.²⁶ Thus, decreased PTEN activity causes an increase in PIP3. PIP3 recruits the serine/threonine protein kinase AKT3 to the cell membrane where it becomes activated.²⁷ Through phosphorylation of various substrates, AKT3 inhibits apoptosis.²⁷ AKT3 activity is increased in up to 60% of nonfamilial melanomas due to either increased copies of AKT3 or loss of PTEN.²⁸

FRAP1, also known as mammalian Target of Rapamycin (mTOR), is a protein kinase that is found in two different complexes called mTORC1 and mTORC2.²⁹ mTORC1 promotes protein translation, favoring cell proliferation.²⁹ mTORC2 activates AKT by phosphorylation.²⁹ Increased mTOR activity was observed in 73% of melanomas.³⁰

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 susceptibility of the protein coding gene (Fig. 1). A perl script was written to carry out this algorithm.

Version 37 of the human genome chromosomal genbank flat files were obtained for analysis (<http://www.ncbi.nlm.nih.gov/Ftp/>). A total of 32,357 coding sequences (CDS) were analyzed. 95 coding sequences were not analyzed because they were incomplete fragments. Due to our inclusion of flanking nucleotides in screening for UVB sites, the calculation of length of a coding sequence for the purpose of estimating UVB frequency, was defined as the sum of all coding sequence nucleotides including the stop codon + 2 · (number of exons that contained coding sequence).

Genes were optimized for UVB resistance by selective synonymous codon substitution. The wildtype primary amino acid sequence was maintained but used the following codons substituted for the original codons: F => TTT, L => TTA, I => ATT, M => ATG, V => GTA, S => TCA, P => CCA, T => ACA, A => GCA, Y => TAT, STOP => TAA, H => CAT, Q => CAA, N => AAT,

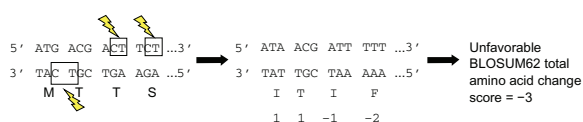


Figure 1. Schematic representation of algorithm to assess UVB sensitivity. At left, boxes are drawn around UVB sites. Below the double stranded DNA is the single letter amino acid sequence for each codon of the upper strand. In the middle, nucleotides that have undergone mutation are shown in bold. The 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.

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

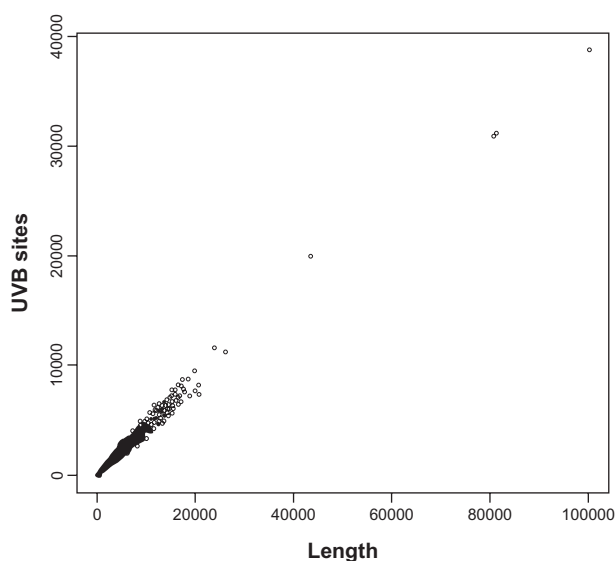


Figure 2. Scatterplot of the number of UVB sites as a function of length of the coding sequence.



is the probability of a dinucleotide being a UVB site based on a 53.48% GC composition and random distribution. The number 0.445 is the probability of a dinucleotide being a UVB site based on the average UVB frequency among all coding sequences. The number of dinucleotides in a single stranded sequence is equal to the length in nt minus one.

The expected number of deleterious single mutation sites in a CDS was calculated by multiplying the number of UVB sites in each CDS by 0.298. The number 0.298 is the mean frequency of a single mutation being deleterious among all coding sequences.

The expected number of deleterious double mutation sites in a CDS was calculated by multiplying the number of CC and GG UVB sites by 0.679. The number 0.679 is the mean frequency of a

double mutation being deleterious among all coding sequences.

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
<i>FRAP1</i>	NP_004949.1	3344	3403.8	0.236
<i>NRAS</i>	NP_002515.1	204	253.21	1.22E-04
<i>AKT3</i>	NP_859029.1	535	621.66	1.38E-05
<i>AKT3</i>	NP_005456.1	559	640.36	6.22E-05
<i>MITF</i>	NP_937802.1	647	695.09	2.80E-02
<i>MITF</i>	NP_006713.1	643	693.75	1.98E-02
<i>MITF</i>	NP_937820.1	623	673.73	1.81E-02
<i>MITF</i>	NP_937821.1	605	661.72	7.20E-03
<i>MITF</i>	NP_000239.1	504	560.25	3.62E-03
<i>MITF</i>	NP_937801.1	500	552.25	6.70E-03
<i>KIT</i>	NP_000213.1	1122	1303.85	1.53E-10
<i>KIT</i>	NP_001087241.1	1120	1298.51	3.11E-10
<i>BRAF</i>	NP_004324.2	924	1023.5	1.11E-04
<i>CDKN2A</i>	NP_000068.1	217	209.15	0.550
<i>CDKN2A</i>	NP_478102.1	242	231.84	0.456
<i>CDKN2A</i>	NP_478104.2	186	155.75	2.96E-03
<i>PTEN</i>	NP_000305.3	417	538.89	2.32E-11
<i>CDK4</i>	NP_000066.1	421	405.39	0.381
<i>APAF1</i>	NP_863651.1	1425	1666.97	3.56E-14
<i>APAF1</i>	NP_863658.1	1378	1609.57	1.70E-13
<i>APAF1</i>	NP_037361.1	1415	1652.29	8.82E-14
<i>APAF1</i>	NP_001151.1	1368	1594.88	4.13E-13
<i>APAF1</i>	NP_863659.1	404	452.12	5.73E-03
<i>TP53</i>	NP_001119584.1	562	525.54	0.0576
<i>TP53</i>	NP_000537.3	562	525.54	0.0576
<i>TP53</i>	NP_001119587.1	365	349.32	0.340
<i>TP53</i>	NP_001119585.1	476	462.8	0.495
<i>TP53</i>	NP_001119589.1	279	286.58	0.626
<i>TP53</i>	NP_001119586.1	471	456.12	0.435
<i>TP53</i>	NP_001119588.1	274	279.91	0.707
<i>BCL2</i>	NP_000624.2	345	319.95	0.0978
<i>BCL2</i>	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 $44.5\% \pm 5.2\%$, 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
<i>FRAP1</i>	NP_004949.1	812	978.63	4.69E-09
<i>NRAS</i>	NP_002515.1	47	59.3	0.124
<i>AKT3</i>	NP_859029.1	147	133.8	0.292
<i>AKT3</i>	NP_005456.1	150	139.17	0.410
<i>MITF</i>	NP_937802.1	175	185.36	0.505
<i>MITF</i>	NP_006713.1	171	182.38	0.454
<i>MITF</i>	NP_937820.1	165	177.31	0.406
<i>MITF</i>	NP_937821.1	164	173.44	0.533
<i>MITF</i>	NP_000239.1	135	143.34	0.547
<i>MITF</i>	NP_937801.1	133	142.15	0.501
<i>KIT</i>	NP_000213.1	317	302.77	0.470
<i>KIT</i>	NP_001087241.1	316	302.17	0.484
<i>BRAF</i>	NP_004324.2	311	254.49	1.93E-04
<i>CDKN2A</i>	NP_000068.1	60	75.99	0.0719
<i>CDKN2A</i>	NP_478102.1	105	81.95	9.62E-03
<i>CDKN2A</i>	NP_478104.2	62	57.22	0.588
<i>PTEN</i>	NP_000305.3	112	108.17	0.764
<i>CDK4</i>	NP_000066.1	123	122.48	0.972
<i>APAF1</i>	NP_863651.1	368	365.35	0.915
<i>APAF1</i>	NP_863658.1	356	350.45	0.813
<i>APAF1</i>	NP_037361.1	362	362.37	0.988
<i>APAF1</i>	NP_001151.1	350	347.47	0.916
<i>APAF1</i>	NP_863659.1	106	101.02	0.678
<i>TP53</i>	NP_001119584.1	193	163.01	1.79E-02
<i>TP53</i>	NP_000537.3	193	163.01	1.79E-02
<i>TP53</i>	NP_001119587.1	112	106.98	0.686
<i>TP53</i>	NP_001119585.1	171	138.57	4.73E-03
<i>TP53</i>	NP_001119589.1	90	82.55	0.468
<i>TP53</i>	NP_001119586.1	170	137.38	4.28E-03
<i>TP53</i>	NP_001119588.1	89	81.35	0.451
<i>BCL2</i>	NP_000624.2	115	111.45	0.785
<i>BCL2</i>	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
<i>FRAP1</i>	NP_004949.1	374	406.04	0.0483
<i>NRAS</i>	NP_002515.1	20	23.77	0.41952
<i>AKT3</i>	NP_859029.1	57	53.64	0.66931
<i>AKT3</i>	NP_005456.1	60	56.36	0.6467
<i>MITF</i>	NP_937802.1	97	86.23	0.18206
<i>MITF</i>	NP_006713.1	90	82.16	0.35182
<i>MITF</i>	NP_937820.1	89	80.8	0.31838
<i>MITF</i>	NP_937821.1	89	80.8	0.31838
<i>MITF</i>	NP_000239.1	70	63.83	0.41952
<i>MITF</i>	NP_937801.1	68	62.47	0.47403
<i>KIT</i>	NP_000213.1	143	136.48	0.58688
<i>KIT</i>	NP_001087241.1	143	135.8	0.5377
<i>BRAF</i>	NP_004324.2	130	124.94	0.67396
<i>CDKN2A</i>	NP_000068.1	30	40.74	0.03377
<i>CDKN2A</i>	NP_478102.1	51	55.68	0.5311
<i>CDKN2A</i>	NP_478104.2	34	35.99	0.77353
<i>PTEN</i>	NP_000305.3	46	40.74	0.38021
<i>CDK4</i>	NP_000066.1	59	57.72	0.89452
<i>APAF1</i>	NP_863651.1	136	126.97	0.39808
<i>APAF1</i>	NP_863658.1	131	122.22	0.40319
<i>APAF1</i>	NP_037361.1	135	125.62	0.37138
<i>APAF1</i>	NP_001151.1	130	120.86	0.37571
<i>APAF1</i>	NP_863659.1	46	40.74	0.38021
<i>TP53</i>	NP_001119584.1	107	92.34	0.05998
<i>TP53</i>	NP_000537.3	107	92.34	0.05998
<i>TP53</i>	NP_001119587.1	64	57.04	0.31334
<i>TP53</i>	NP_001119585.1	92	79.44	0.08774
<i>TP53</i>	NP_001119589.1	49	44.14	0.44857
<i>TP53</i>	NP_001119586.1	92	79.44	0.08774
<i>TP53</i>	NP_001119588.1	49	44.14	0.44857
<i>BCL2</i>	NP_000624.2	65	68.58	0.69252
<i>BCL2</i>	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 *APAF1* and *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
<i>FRAP1</i>	NP_004949.1	-2.33	8.96E-05
<i>NRAS</i>	NP_002515.1	-1.98	0.682
<i>AKT3</i>	NP_859029.1	-2.23	0.738
<i>AKT3</i>	NP_005456.1	-2.21	0.831
<i>MITF</i>	NP_937802.1	-2.05	0.682
<i>MITF</i>	NP_006713.1	-2.02	0.549
<i>MITF</i>	NP_937820.1	-2.05	0.682
<i>MITF</i>	NP_937821.1	-2.07	0.750
<i>MITF</i>	NP_000239.1	-2.04	0.682
<i>MITF</i>	NP_937801.1	-2.04	0.682
<i>KIT</i>	NP_000213.1	-2.1	0.831
<i>KIT</i>	NP_001087241.1	-2.1	0.831
<i>BRAF</i>	NP_004324.2	-2.16	0.969
<i>CDKN2A</i>	NP_000068.1	-2.15	0.979
<i>CDKN2A</i>	NP_478102.1	-2.03	0.682
<i>CDKN2A</i>	NP_478104.2	-2.19	0.960
<i>PTEN</i>	NP_000305.3	-2.16	0.979
<i>CDK4</i>	NP_000066.1	-2.2	0.831
<i>APAF1</i>	NP_863651.1	-2.34	3.49E-03
<i>APAF1</i>	NP_863658.1	-2.35	2.56E-03
<i>APAF1</i>	NP_037361.1	-2.35	2.56E-03
<i>APAF1</i>	NP_001151.1	-2.36	2.56E-03
<i>APAF1</i>	NP_863659.1	-2.32	0.386
<i>TP53</i>	NP_001119584.1	-2.13	0.979
<i>TP53</i>	NP_000537.3	-2.13	0.979
<i>TP53</i>	NP_001119587.1	-2.11	0.965
<i>TP53</i>	NP_001119585.1	-2.14	0.995
<i>TP53</i>	NP_001119589.1	-2.11	0.969
<i>TP53</i>	NP_001119586.1	-2.14	0.979
<i>TP53</i>	NP_001119588.1	-2.1	0.965
<i>BCL2</i>	NP_000624.2	-2.2	0.831
<i>BCL2</i>	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%.³⁶ In addition, the frequency of melanoma is increasing and over 60,000 new cases occur each year in the United States.³⁷ 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

**Table 6.** Synonymous codon substitution can decrease predicted UVB vulnerability of genes implicated in malignant melanoma.

Gene	Protein	WT UVB sites	WT score	Optimized UVB sites	Optimized score
FRAP1	NP_004949.1	3284	-2874	1464	-2096
NRAS	NP_002515.1	199	-137	108	-105
AKT3	NP_859029.1	449	-487	261	-355
AKT3	NP_005456.1	467	-497	274	-361
MITF	NP_937802.1	622	-612	305	-466
MITF	NP_006713.1	612	-582	295	-449
MITF	NP_937820.1	595	-574	288	-444
MITF	NP_937821.1	582	-575	284	-449
MITF	NP_000239.1	481	-460	239	-358
MITF	NP_937801.1	477	-449	236	-351
KIT	NP_000213.1	1016	-1058	558	-789
KIT	NP_001087241.1	1014	-1057	556	-788
BRAF	NP_004324.2	854	-1029	511	-795
CDKN2A	NP_000068.1	255	-217	120	-151
CDKN2A	NP_478102.1	275	-349	157	-258
CDKN2A	NP_478104.2	192	-226	108	-159
PTEN	NP_000305.3	363	-372	225	-311
CDK4	NP_000066.1	411	-445	204	-302
APAF1	NP_863651.1	1226	-1217	694	-931
APAF1	NP_863658.1	1176	-1183	668	-909
APAF1	NP_037361.1	1216	-1205	685	-923
APAF1	NP_001151.1	1166	-1171	659	-901
APAF1	NP_863659.1	339	-364	200	-250
TP53	NP_001119584.1	547	-715	293	-516
TP53	NP_000537.3	547	-715	293	-516
TP53	NP_001119587.1	359	-402	174	-268
TP53	NP_001119585.1	465	-639	258	-485
TP53	NP_001119589.1	277	-326	139	-237
TP53	NP_001119586.1	461	-633	258	-479
TP53	NP_001119588.1	273	-320	139	-231
BCL2	NP_000624.2	374	-427	175	-295
BCL2	NP_000648.2	326	-381	160	-262

gene therapy has promise.³⁸ 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. *Annu 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. Nospikel 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.



7. Connell-Crowley L, Harper JW, Goodrich DW. Cyclin D1/Cdk4 regulates retinoblastoma protein-mediated cell cycle arrest by site-specific phosphorylation. *Mol Biol Cell*. 1997;8:287–301.
8. Goldstein AM, Struewing JP, Chidambaram A, Fraser MC, Tucker MA. Genotype-phenotype relationships in US. melanoma-prone families with CDKN2A and CDK4 mutations. *JNCI J Natl Cancer Inst*. 2000;92:1006–10.
9. Ragione FD, Russo GL, Oliva A, et al. Biochemical characterization of p16INK4- and p18-containing complexes in human cell lines. *J Biol Chem*. 1996;271:15942–9.
10. Schaffer JV, Bologna JL. The melanocortin-1 receptor. Red hair and beyond. *Archives of Dermatology*. 2001;137:1477–85.
11. Rouzad F, Costin G, Yamaguchi Y, Valencia JC, Berens WF, Chen KG, et al. Regulation of constitutive and UVR-induced skin pigmentation by melanocortin 1 receptor isoforms. *Faseb J*. 2006;20:1927–9.
12. Kennedy C, ter Huurne J, Berkhout M, Gruis N, Bastiaens M, Bergman W, et al. Melanocortin 1 receptor (MC1R) gene variants are associated with an increased risk for cutaneous melanoma which is largely independent of skin type and hair color. *J Invest Dermatol*. 2001;117:294–300.
13. Vachtenheim J, Borovansky J. “Transcription physiology” of pigment formation in melanocytes: central role of MITF. *Exp Dermatol*. 2010;19:617–27.
14. Garraway LA, Widlund HR, Rubin MA, Getz G, Berger AJ, Ramaswamy S, et al. Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature*. 2005;436:117–22.
15. Carlson JA, Linette GP, Aplin A, Ng B, Slominski A. Melanocyte receptors: clinical implications and therapeutic relevance. *Dermatol Clin*. 2007;25:541–57.
16. Funasaka Y, Boulton T, Cobb M, Yarden Y, Fan B, Lyman SD, et al. c-Kit-kinase induces a cascade of protein tyrosine phosphorylation in normal human melanocytes in response to mast cell growth factor and stimulates mitogen-activated protein kinase but is down-regulated in melanomas. *Mol Biol Cell*. 1992;3:197–209.
17. Curtin JA, Busam K, Pinkel D, Bastian BC. Somatic activation of KIT in distinct subtypes of melanoma. *J Clin Oncol*. 2006;24:4340–6.
18. Ghosh P, Chin L. Genetics and genomics of melanoma. *Expert Rev Dermatol*. 2009;4:131–43.
19. Wimmer R, Baccharini M. Partner exchange: protein-protein interactions in the Raf pathway. *Trends Bio Sci*. 2010;35:660–8.
20. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, et al. Mutations of the BRAF gene in human cancer. *Nature*. 2002;417:949–54.
21. Chipuk JE, Green DR. How do BCL-2 proteins induce mitochondrial outer membrane permeabilization? *Trends in cell biology*. 2008;18:157–64.
22. Bush JA, Li G. The role of Bcl-2 family members in the progression of cutaneous melanoma. *Clin Exp Metastasis*. 2003;20:531–9.
23. Jansen B, Schlagbauer-Wadl H, Brown BD, Bryan RN, van Elsas A, Muller M, et al. Bcl-2 antisense therapy chemosensitizes human melanoma in SCID mice. *Nat Med*. 1998;4:232–4.
24. Anichini A, Mortarini R, Sensi M, Zanon M. APAF-1 signaling in human melanoma. *Cancer Lett*. 2006;238:168–79.
25. Soengas MS, Capodieci P, Polsky D, Mora J, Esteller M, Opitz-Araya X. Inactivation of the apoptosis effector Apaf-1 in malignant melanoma. *Nature*. 2001;409:207–11.
26. Waite KA, Eng C. Protean PTEN: form and function. *Am J Hum Genet*. 2002;70:829–44.
27. Nicholson KM, Anderson NG. The protein kinase B/Akt signaling pathway in human malignancy. *Cell Signal*. 2002;14:381–95.
28. Stahl JM, Sharma A, Cheung M, Zimmerman M, Cheng JQ, Bosenberg MW, et al. Deregulated Akt3 activity promotes development of malignant melanoma. *Cancer Res*. 2004;64:7002–10.
29. Guertin D, Sabatini DM. Defining the role of mTOR in cancer. *Cancer Cell*. 2007;12:9–22.
30. Karbowniczek M, Spittle CS, Morrison T, Wu H, Henske EP. mTOR is activated in the majority of malignant melanomas. *Journal of Investigative Dermatology*. 2007;128:980–7.
31. Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, et al. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature Genetics*. 1999;22:231–8.
32. Eddy SR. Where did the BLOSUM62 alignment score matrix come from? *Nature Biotechnology*. 2004;22:1035–6.
33. Pollard KS, Gilbert HN, Ge Y, Taylor S, Dudoit S. Multtest: Resampling-based multiple hypothesis testing. R package version 2.1.3; 2009. Available at: <http://CRAN.R-project.org/package=multtest>.
34. Gentleman R, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biology*. 2004;5:R80.
35. Benjamini Y, Hochberg Y. On the adaptive control of the false discovery rate in multiple testing with independent statistics. *J Behav Educ Statist*. 2000; 25:60–83.
36. Bhatia S, Tykodi S, Thompson JA. Treatment of metastatic melanoma: an overview. *Oncology (Williston Park)*. 2009;23:488–96.
37. Wolff T, Tai E, Miller T. Screening for skin cancer: an update of the evidence for the US. Preventive Services Task Force; 2009. *Evidence Syntheses, No. 67*. Available at: <http://www.ncbi.nlm.nih.gov/books/NBK34051/>.
38. Khavari PA, Rollman O, Vahlquist A. Cutaneous gene transfer for skin and systemic diseases. *J Intern Med*. 2002;252:1–10.



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.

Publish with Libertas Academica and every scientist working in your field can read your article

"I would like to say that this is the most author-friendly editing process I have experienced in over 150 publications. Thank you most sincerely."

"The communication between your staff and me has been terrific. Whenever progress is made with the manuscript, I receive notice. Quite honestly, I've never had such complete communication with a journal."

"LA is different, and hopefully represents a kind of scientific publication machinery that removes the hurdles from free flow of scientific thought."

Your paper will be:

- Available to your entire community free of charge
- Fairly and quickly peer reviewed
- Yours! You retain copyright

<http://www.la-press.com>