UV biomarker genes for classification and risk stratification of cutaneous actinic keratoses and squamous cell carcinoma subtypes

Dawn Queen, MD[#1], **Yao Shen, PhD**[#2], **Megan H. Trager, BA**[#1], **Adriana T. Lopez, MD**[#3], **Faramarz H. Samie, MD, PhD**[4], **Jesse M. Lewin, MD**[4], **George W. Niedt, MD**[5], **Larisa J. Geskin, MD**[4], **Liang Liu, PhD**[4,6,*]

[1]Columbia University Vagelos College of Physicians and Surgeons, New York, NY 10032

[2]Department of Systems Biology, Columbia University Irving Medical Center, New York, NY 10032

[3]Memorial Sloan Kettering Cancer Center, New York, NY 10032

[4]Department of Dermatology, Columbia University Irving Medical Center, New York, NY 10032

[5]Department of Pathology, Columbia University Irving Medical Center, New York, NY 10032

[6]The Hormel Institute, University of Minnesota, Austin, MN 55912

[#] These authors contributed equally to this work.

Abstract

Currently, there is no sensitive molecular test for identifying transformation-prone actinic keratoses (AKs) and aggressive SCC subtypes. Biomarker-based molecular testing represents a promising tool for risk stratifying these lesions. We evaluated the utility of a panel of UV-biomarker genes in distinguishing between benign and transformation-prone AKs and SCCs. The expression of the UV-biomarker genes in 31 SCC and normal skin (NS) pairs and 10 AK/NS pairs was quantified using the NanoString nCounter system. Biomarker testing models were built using logistic regression models with leave one-out cross validation in the training set. The best model to classify AKs versus SCCs (area under curve (AUC) 0.814, precision score 0.833, recall 0.714) was constructed using a top-ranked set of 13 UV-biomarker genes. Another model based on a 15-gene panel was developed to differentiate histologically concerning from less concerning SCCs (AUC 1, precision score 1, recall 0.714). Finally, 12 of the UV-biomarker genes were differentially expressed between AKs and SCCs, while 10 genes were uniquely expressed in the more

concerning SCCs. UV-biomarker gene subsets demonstrate dynamic utilities as molecular tools to classify and risk stratify AK and SCC lesions, which will complement histopathologic diagnosis to guide treatment of high-risk patients.

Keywords

Actinic Keratosis; squamous cell carcinoma; ultraviolet radiation

Introduction

Squamous cell carcinoma (SCC) is the second most frequent nonmelanoma skin cancer (NMSC), with over 700,000 cases diagnosed annually (1-3). SCCs are generally curable apart from a concerning subset with increased tendency for metastasis, recurrence, and disease-related morbidity and mortality (4). Although most SCCs arise from precursor lesions termed actinic keratoses (AK), the estimated rate of progression from AK to SCC is less than 5% (5-8). There is a significant unmet need for reliable and sensitive methods to identify transformation-prone AKs and concerning SCCs which may facilitate early intervention and improve disease outcomes (9).

Ultraviolet radiation (UVR) is a major risk factor for melanoma and NMSC including SCCs and their AK precursors (10). Repeated sunburns cause cumulative genetic and epigenetic lesions that promote malignant transformation in sun-damaged skin cells (11, 12). Imaging methods such as dermoscopy aid in tumor detection (13), while biopsy with histopathological analysis can establish a definitive diagnosis. However, these methods are limited because they rely on visual pattern recognition and cannot detect oncogenic molecular alterations in sun-exposed skin that may persist for decades (13-16). More sensitive biomarker-based molecular methods may facilitate identification of concerning lesions prior to gross histologic alterations, enabling early interventions and improved outcomes.

Studies have attempted to identify biomarkers for monitoring UV damage and skin cancer risk (13, 17-20). However, no consensus UV-biomarker panel is available due to large variations among studies and lack of cross-validation of candidate UV-biomarker genes. We recently reported a panel of highly conserved UV-biomarker genes discovered in human keratinocytes in response to various UVR conditions and among keratinocytes from different donors (21, 22). This UV-biomarker panel consists of UV-responsive genes involved in cancer-related pathways. Through statistical and bioinformatics analyses, we selected a set of 77 UV-biomarker genes that are often dysregulated in human SCCs. We posit that this UV-biomarker subset has potential in skin cancer risk stratification. Here, we performed multiplex gene expression analyses to characterize the ability of subsets of the UV-biomarker genes to identify transformation-prone AKs and SCCs.

Materials and Methods

Patient tissue collection, procurement, and histopathological analysis

The study was approved by the Institutional Review Board at Columbia University Irving Medical Center (CUIMC) (IRB#AAAR3361). Fresh surgically resected AK or SCC with patient-matched adjacent normal skin (NS) were obtained from patients (male and female, > age 18) who presented to CUIMC for treatment of clinically evident AK or SCC on sun-exposed areas (face, neck, upper back, arms). AKs were obtained using shave biopsy, and SCCs were collected during Mohs micrographic surgery (MMS). Prior to each SCC collection, incisional biopsy was collected for histopathological analysis. During MMS, frozen sections were processed by trained histotechnicians, stained with hematoxylin-eosin, and reviewed by the surgeon. Half of each AK was used for RNA extraction and gene expression analysis, and the other half was formalin fixed, paraffin embedded, and used for histopathological grading by hematoxylin-eosin staining based on established diagnostic criteria. For RNA extraction, freshly collected AK, SCC, and matched NS tissues were procured by incubation in 5 U/mL dispase solution (StemCell Technologies) for 1 hour at 37 °C to separate and remove dermal and adipose tissues.

RNA extraction and multiplex gene expression analysis of UV-biomarker genes by NanoString

Total RNA was isolated from procured NS epidermis, AK, or SCC tissues as previously reported using the RNeasy Kit (QIAGEN, Gaithersburg, MD) (21). Human SCC and adjacent normal skin tissues were lysed in Qiazol (QIAGEN) and homogenized using TissueLyser II (QIAGEN). Total RNA was isolated from each homogenized tissue using the miRNeasy Kit (QIAGEN). All RNA samples were subsequently analyzed using an RNA 6000 nano chip (Agilent Technologies) to confirm the RNA quantity and integrity. RNA samples from 10 AK/NS pairs and 31 SCC/NS pairs with satisfactory RNA quality and quantity were included in subsequent NanoString experiments. To evaluate the biomarker potential and utility of 77 top-ranked and highly conserved UV-responsive genes (21), simultaneous quantification of these UV-biomarker genes was performed on the NanoString nCounter system (NanoString) using a custom NanoString CodeSet. NanoString experiments were performed using 100 ng of RNA per sample and reagents from NanoString at the Genome Technology Center of New York University. The counts per gene were normalized to the geometric mean of the internal reference genes (GLYR1, NMT1, XPO7) included in the CodeSet. Normalized fold change (FC) for each target gene in each AK-NS or SCC-NS pair was generated and log2-transformed log2FC). log2FC values were used in unsupervised data clustering analysis.

Data analysis and statistical models

Unsupervised hierarchical clustering of samples based on log2FC values of each UV-biomarker gene was performed using average linkage based on their pairwise distances. We computed the Euclidean distances between sample pairs using their normalized log2FC values. Student's t-test was applied and two logistic regression biomarker models were built to compare and classify AKs versus SCCs and concerning versus less concerning SCCs, respectively. 70% of the dataset was used as the training set with leave-one-out cross-

validation (LOOCV), and the remaining 30% was used as the testing set. Within each iteration of the LOOCV, a Student's t-test was used to identify the top genes that were different between the AKs and SCCs or between the concerning and less concerning SCCs to be used as the features of the logistic regression models. Samples were classified as more concerning (invasive or evolving SCC) or less concerning (SCC in situ or keratoacanthoma) based on histopathological evaluation by a dermatopathologist (G.N.). Training performance was evaluated based on the predicted likelihood of each sample to be either an AK or SCC and more or less concerning in the LOOCV, which provides an unbiased measure of the training model performance. Classifier performance was estimated by computing the area under the curve (AUC) scores for the receiver-operator curve (ROC), the precision score, and the recall score. All statistical analyses and visualization were performed using the R software package.

Results

Thirty AK/NS pairs and 41 SCC/NS pairs were collected. Following RNA extraction and quality assessment, 41 matched pairs (10 AK/NS; 31 SCC/NS) with satisfactory RNA quantity and quality were included in the multiplex gene expression analysis using the NanoString nCounter system. Demographic information and histologic diagnosis are summarized in Table 1. Representative histological images are shown in Figure 1. In addition to patient samples, two pairs of UV-irradiated and non-irradiated control human keratinocytes were included in the multiplex gene expression analysis to represent UV-induced gene expression signatures. Gene expression datasets were used to identify molecular classifiers that can distinguish between AKs and SCCs. Based on their clinical and histological features (Table 1), SCC samples were divided into a less concerning SCC subgroup comprising keratoacanthomas (KAs) and in situ SCCs, which generally exhibit a more indolent course, and a more concerning subgroup comprising evolving or invasive SCCs that often progress and exhibit invasive, aggressive behavior (23, 24).

Similarities in Gene Signatures between AK, SCC and UV Signatures

The similarity of AK and SCC gene signatures to the UV signatures was determined using cosine similarity, which is the dot product of the two non-zero vectors divided by the product of the two vectors' lengths (magnitudes) and is defined to equal the cosine of the angle between the vectors (25). Genes with log2FC values greater than 0.9 were considered over-expressed and assigned with a value of 1. Genes with log2FC values less than −0.9 were considered under-expressed and assigned a value of −1. All other genes with log2FC values between −0.9 to 0.9 were assigned a value of zero. Unsupervised hierarchical clustering analysis based on the expression signature of the 77 UV-biomarker genes confirmed greater similarity between the AK signatures and the UV signature than between the SCC and UV signatures (Figure 2). However, among the 31 SCCs, six SCCs demonstrated a higher similarity to the UV signature than the rest of the SCC samples, indicating some molecular heterogeneity among the SCCs (Table 1). Similarly, cosine similarity-based analysis revealed a greater similarity between AK gene signatures and the UV signature than between SCC gene signatures and the UV signature (Supplementary

Figure 1). These findings suggest that UVR plays a more consistent role in the pathogenesis of AKs than SCCs.

Development of a Classifier to Distinguish AKs from SCC

Next, we built a statistical model based on selected UV-biomarker gene subsets to differentiate between gene signatures for AKs and SCCs by randomly sampling 70% of the dataset (n=29) from the 41 samples (10 AKs, 31 SCCs) as the training dataset. The remaining samples (30%, N=12) comprised the testing dataset for assessing the model's accuracy and specificity. Within the 29 training samples, LOOCV was performed and a logistic regression model based on these genes was built to predict the likelihood of each sample being from the SCC group or AK group.

To optimize the number of biomarker genes to be included, we ranked all 77 genes from 1 to 77, and tested each classifier model containing different numbers of genes (up to 77) using LOOCV. We identified an optimal classifier based on a 13-gene panel that yielded an AUC score of 0.846, a precision score of 0.95, and a recall score of 0.792 (Figure 3A, **left panel**). Using this training model, we built a model using all 29 training samples, which was used to determine the model's accuracy for predicting the 12 testing samples as AK or SCC. As shown in the right panel in Figure 3A, this model showed a strong prediction power with an AUC score of 0.814, a precision score of 0.833, and a recall score of 0.714 for the testing samples. Unsupervised hierarchical clustering using this 13-gene panel also divided the 12 testing samples into the AK group and SCC group that is largely consistent with their clinical diagnosis except for one AK and three SCC outliers (Figure 3B).

Development of a Classifier to Distinguish More Concerning SCCs from Less Concerning SCCs

Next, we attempted to build a molecular classifier based on other UV-biomarker gene subsets for differentiating less concerning SCCs from more concerning SCCs that exhibit higher risk (23, 24). We divided the 31 SCC samples into 23 more concerning SCCs and 8 less concerning SCCs based on the histopathological report (Table 1). We trained a model by randomly sampling 70% of the SCCs (n=23) from both groups as the training dataset. The remaining 30% of samples (n=8) were utilized as the testing dataset. We first performed LOOCV on the training samples by including all possible subsets of the 77 UV-biomarker genes to build a logistic regression model, which was then applied to predict the likelihood of a sample either belonging to the more concerning group or less concerning group. Following the LOOCV tests, the best training model was identified using a 15-gene panel that produced an AUC score of 0.885, a precision score of 0.933, and a recall score of 0.875 (Figure 4A, **left panel**). A logistic regression model using all 22 training samples was then constructed based on these 15 biomarker genes that were the most differentially expressed between the more and less concerning SCC groups. Using the nine testing SCC samples, we demonstrated that the prediction power of this risk classifier model was high with an AUC score of 1, a precision score of 1, and a recall score of 0.714 for the testing set 792 (Figure 4A, **right panel**). Similarly, unsupervised hierarchical clustering using this 15-gene panel clustered these 31 SCC samples into two distinct groups of more and less concerning SCCs,

respectively, which was consistent with their clinical diagnosis except for SCC65 and SCC66 (Figure 4B).

Differentially Expressed Genes Among AKs, More and Less Concerning SCCs

To identify individual UV-biomarker genes differentially expressed between SCCs and AKs, we performed Student's t-tests comparing SCC versus AK signatures for each UV-biomarker gene. We identified six genes with significantly lower log2FC levels in the AK group compared to the SCC group (SPRR3, TTC7A, HDAC9, TNFAIP2, SPNS2, and DPP4) with FDR-corrected p-values < 0.1 (Figure 5A). Six genes had significantly higher log2FC levels in the AK group compared to the SCC group (HSPB8, CLDN4, HIST2H2BE, DEFB1, PVRL4, and HR) with FDR-corrected p-values < 0.1 (Figure 5B). We performed similar comparative gene expression analysis to identify individual genes that were most consistently differentially expressed between the more concerning SCCs and less concerning SCCs using the Student's t-test. Eight genes showed significantly higher log2FC levels in the more concerning SCC samples compared to less concerning SCC samples (HPGD, SPNS2, USP2, HDAC5, LCE6A, RNF220, KRT23, and LCE1A) (Figure 5C). In contrast, only two genes had significantly lower log2FC levels among the more concerning SCCs compared to less concerning SCCs (NMT1 and PRSS22) (Figure 5D). Several genes identified as strong UV-biomarker candidates in previous studies, including PTGS2 (COX-2) (26), were inconsistently dysregulated among SCCs or AKs, suggesting a lesser potential for clinical utility.

Discussion

AKs are common, especially among people of European origin, and are the second leading reason for dermatologic visits (27). Advancing, aggressive SCCs pose health concerns given their invasive nature associated with recurrences, metastasis, and dismal prognosis. Clinical diagnosis of AKs and SCCs relies on gross clinical appearance and histopathological features, which are often subjective and can lead to inconsistent diagnosis (28, 29). There is growing interest in biomarker-based molecular tests for more objective and accurate diagnosis of transformation-prone AKs and concerning SCCs. Building upon our previous efforts to identify a panel of highly-conserved UV-biomarker genes, we performed multiplex gene expression analysis of a 77 UV gene signature panel to characterize their biomarker potential in AK and SCC risk-stratifications. The expression pattern of the UV-biomarker genes demonstrates a gradient distribution of UV signatures among AKs and SCCs. The similarity between AK signatures and the UV signature supports the critical role of UVR in AK pathogenesis. It is intriguing that the unsupervised hierarchical clustering based on the expression of UV-biomarker genes identified two SCC subgroups, with one group more closely related to UV exposure. While UVR is believed to be the major cause of SCC development, accumulating evidence suggests that SCC pathogenesis is multifactorial and dependent on both intrinsic (age, scars, skin pigmentation, and photosensitive dermatoses) (30) and extrinsic factors (UVR, industrial carcinogens, immunosuppressive medications, and human papillomavirus) (31). Our findings of two distinct SCC subgroups corroborate the theory of different molecular pathways leading to SCC development including UV and non-UV carcinogens.

Using selected subsets of UV-biomarker genes, we found differential clustering of AK and SCC groups. While the differentiation of AK and SCC subgroups based on our molecular classifiers (Figures 3 and 4) is highly consistent with the histopathologic diagnosis, some outliers exist. This is attributable to inherent differences in histopathology-based and biomarker-based classification. Although further optimization and validation of the biomarker classifiers will likely improve their diagnostic accuracy, other factors contributing to this inconsistency include tumor heterogeneity, atypical histopathologic features, and inability of histopathology to predict disease progression (28, 29). Therefore, discrepant results between histopathology-based diagnosis and biomarker-based classification are expected. In our samples, all of the AKs that clustered more closely with SCCs were from patients who had a history of melanoma or nonmelanoma skin cancer in the past. Additionally, two of these patients were SOT recipients and predisposed to development of NMSCs, perhaps indicating a transformation-prone nature of these AKs. Further clinical observations and laboratory characterizations of atypical and outlier samples are warranted to resolve the discrepancy and evaluate the reliability of each method. Given the versatility of the UV-biomarker genes, it is possible to develop multiple panels based on different subsets of the UV-biomarker genes for cross validation to improve diagnostic accuracy.

Importantly, the UV-biomarker gene subsets can provide molecular diagnostic information that complements standard histologic evaluations. Molecular classifiers derived from the UV-biomarker panel can be used to develop affordable molecular tests to distinguish transformation-prone AKs from more benign AKs. Given the vast prevalence of AKs among the older Caucasian population, identification of indolent, low-malignant potential AKs will allow clinicians to avoid unnecessary treatment and limit healthcare costs. Notably, the classifier in Figure 3B reveals that AK1 exhibits a stronger SCC-like signature than other AKs. This AK was from a patient who was immunosuppressed, which increases risk for SCC. It is realistic that this AK was progressing towards SCC despite its clinical appearance as an AK. Similar biomarker-tests are needed to risk-stratify SCC subtypes and provide additional information in atypical cases. Tumor-associated UV-biomarkers have the potential to detect cancer early and enable risk stratify lesions, enabling a targeted approach for aggressive modalities.

Based on the SCC risk classifier in Figure 4B, one invasive SCC sample (SCC66) by histologic analysis clustered molecularly with the less concerning SCC group. This patient had a solid organ transplant and had many prior BCCs, AKs, a melanoma, and a SCC. The patient was over 70 years-old and had no evidence of metastatic or perineural invasion. We postulate that despite a pathological diagnosis of invasive SCC, this SCC may have demonstrated lower risk behavior, never going on to metastasize or recur. Another patient diagnosed with invasive SCC (SCC65) clustered with the less concerning group. This patient's pathology showed a predominantly in situ and focally invasive SCC. Histologically, this was characterized as an invasive SCC with the tumor extending to the margins, but the biomarker-based test classified it with the less concerning SCCs, perhaps highlighting the mixed nature of this tumor with invasive and in situ components. Thus, biomarker-based tests can provide clinically-relevant, sensitive molecular information that may complement histologic assessment.

A limitation of the current study is the relatively small number of AK samples. Instead of using whole skin biopsy or bulk tumor tissue to extract RNA for gene expression analysis, all clinical samples were carefully procured to obtain epidermal tissue with minimal contamination of other cell types from the dermal and adipose tissue to minimize tissue heterogeneity. While this stringent tissue procurement protocol greatly increased our data quality, it inevitably rendered losses of the limited clinical specimens. Although we collected AKs from over 30 patients, many AK specimens were too small as half of each was saved for histopathologic analyses. In the end, we obtained sufficient RNA from only 10 AK specimens, which were used in subsequent gene expression analyses. Future studies using a larger cohort of AK samples are warranted.

In conclusion, our analyses demonstrate that the UV-biomarker panel identified in our previous studies contains a dynamic set of genes with versatile implications in clinical diagnosis and risk stratification of cutaneous lesions related to UVR. We have demonstrated that subsets of the UV-biomarker genes can be used to develop molecular classifiers to distinguish AKs from SCC subtypes. While classification of AK and SCC subtypes by these novel molecular classifiers is largely consistent with clinical and histopathological diagnosis, the molecular classifiers provide distinctive molecular information on atypical samples that correlate better with their clinical course than histopathology. Biomarker-based diagnosis may be especially sensitive for identifying patients prone to biologically more aggressive skin cancers who would benefit from more proactive treatment. Upon further optimization and clinical validations, these biomarker-based classifiers can provide complementary information to enable accurate clinical classification and risk stratification of cutaneous lesions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

Funding:

Data Availability Statement:

All data generated or analyzed during this study are included in this article and its supplementary information files.

Abbreviations:

AK	Actinic Keratosis
SCC	Squamous cell carcinoma

NS	normal skin
UV	Ultraviolet
NMSC	nonmelanoma skin cancer
log2FC	log2-transformed FC
LOOCV	leave one out cross validation
AUC	area under the curve
ROC	receiver-operator curve

References

1. Karia PS, Han J, and Schmults CD (2013) Cutaneous squamous cell carcinoma: estimated incidence of disease, nodal metastasis, and deaths from disease in the United States, 2012. Journal of the American Academy of Dermatology 68, 957–966 [PubMed: 23375456]

2. Rogers HW, Weinstock MA, Feldman SR, and Coldiron BM (2015) Incidence Estimate of Nonmelanoma Skin Cancer (Keratinocyte Carcinomas) in the U.S. Population, 2012. JAMA dermatology 151, 1081–1086 [PubMed: 25928283]

3. Karia PS, Jambusaria-Pahlajani A, Harrington DP, Murphy GF, Qureshi AA, and Schmults CD (2014) Evaluation of American Joint Committee on Cancer, International Union Against Cancer, and Brigham and Women's Hospital tumor staging for cutaneous squamous cell carcinoma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 32, 327–334 [PubMed: 24366933]

4. Schmults CD, Karia PS, Carter JB, Han J, and Qureshi AA (2013) Factors predictive of recurrence and death from cutaneous squamous cell carcinoma: a 10-year, single-institution cohort study. JAMA dermatology 149, 541–547 [PubMed: 23677079]

5. Criscione VD, Weinstock MA, Naylor MF, Luque C, Eide MJ, and Bingham SF (2009) Actinic keratoses: Natural history and risk of malignant transformation in the Veterans Affairs Topical Tretinoin Chemoprevention Trial. Cancer 115, 2523–2530 [PubMed: 19382202]

6. Lebwohl M (2003) Actinic keratosis: epidemiology and progression to squamous cell carcinoma. The British journal of dermatology 149 Suppl 66, 31–33 [PubMed: 14616345]

7. Feldman SR, and Fleischer AB Jr. (2011) Progression of actinic keratosis to squamous cell carcinoma revisited: clinical and treatment implications. Cutis 87, 201–207 [PubMed: 21644496]

8. Glogau RG (2000) The risk of progression to invasive disease. Journal of the American Academy of Dermatology 42, 23–24 [PubMed: 10607353]

9. Chitsazzadeh V, Coarfa C, Drummond JA, Nguyen T, Joseph A, Chilukuri S, Charpiot E, Adelmann CH, Ching G, Nguyen TN, Nicholas C, Thomas VD, Migden M, MacFarlane D, Thompson E, Shen J, Takata Y, McNiece K, Polansky MA, Abbas HA, Rajapakshe K, Gower A, Spira A, Covington KR, Xiao W, Gunaratne P, Pickering C, Frederick M, Myers JN, Shen L, Yao H, Su X, Rapini RP, Wheeler DA, Hawk ET, Flores ER, and Tsai KY (2016) Cross-species identification of genomic drivers of squamous cell carcinoma development across preneoplastic intermediates. Nat Commun 7, 12601 [PubMed: 27574101]

10. Stern RS (2010) Prevalence of a history of skin cancer in 2007: results of an incidence-based model. Archives of dermatology 146, 279–282 [PubMed: 20231498]

11. Ziegler A, Jonason AS, Leffell DJ, Simon JA, Sharma HW, Kimmelman J, Remington L, Jacks T, and Brash DE (1994) Sunburn and p53 in the onset of skin cancer. Nature 372, 773–776 [PubMed: 7997263]

12. Bolshakov S, Walker CM, Strom SS, Selvan MS, Clayman GL, El-Naggar A, Lippman SM, Kripke ML, and Ananthaswamy HN (2003) p53 mutations in human aggressive and nonaggressive basal and squamous cell carcinomas. Clinical cancer research : an official journal of the American Association for Cancer Research 9, 228–234 [PubMed: 12538474]

13. Skvara H, Teban L, Fiebiger M, Binder M, and Kittler H (2005) Limitations of dermoscopy in the recognition of melanoma. Arch Dermatol 141, 155–160 [PubMed: 15724011]

14. Osborne JE, Chave TA, and Hutchinson PE (2003) Comparison of diagnostic accuracy for cutaneous malignant melanoma between general dermatology, plastic surgery and pigmented lesion clinics. Br J Dermatol 148, 252–258 [PubMed: 12588376]

15. Carli P, Nardini P, Crocetti E, De Giorgi V, and Giannotti B (2004) Frequency and characteristics of melanomas missed at a pigmented lesion clinic: a registry-based study. Melanoma Res 14, 403–407 [PubMed: 15457097]

16. Marks R, Jolley D, McCormack C, and Dorevitch AP (1997) Who removes pigmented skin lesions? J Am Acad Dermatol 36, 721–726 [PubMed: 9146533]

17. Dawes JM, Antunes-Martins A, Perkins JR, Paterson KJ, Sisignano M, Schmid R, Rust W, Hildebrandt T, Geisslinger G, Orengo C, Bennett DL, and McMahon SB (2014) Genome-wide transcriptional profiling of skin and dorsal root ganglia after ultraviolet-B-induced inflammation. PloS one 9, e93338 [PubMed: 24732968]

18. Yang G, Zhang G, Pittelkow MR, Ramoni M, and Tsao H (2006) Expression profiling of UVB response in melanocytes identifies a set of p53-target genes. The Journal of investigative dermatology 126, 2490–2506 [PubMed: 16888633]

19. Rieger KE, and Chu G (2004) Portrait of transcriptional responses to ultraviolet and ionizing radiation in human cells. Nucleic Acids Res 32, 4786–4803 [PubMed: 15356296]

20. Dazard JE, Gal H, Amariglio N, Rechavi G, Domany E, and Givol D (2003) Genome-wide comparison of human keratinocyte and squamous cell carcinoma responses to UVB irradiation: implications for skin and epithelial cancer. Oncogene 22, 2993–3006 [PubMed: 12771951]

21. Shen Y, Kim AL, Du R, and Liu L (2016) Transcriptome Analysis Identifies the Dysregulation of Ultraviolet Target Genes in Human Skin Cancers. PloS one 11, e0163054 [PubMed: 27643989]

22. Shen Y, Chan G, Xie M, Zeng W, and Liu L (2019) Identification of master regulator genes of UV response and their implications for skin carcinogenesis. Carcinogenesis 40, 687–694 [PubMed: 30452757]

23. Mullen JT, Feng L, Xing Y, Mansfield PF, Gershenwald JE, Lee JE, Ross MI, and Cormier JN (2006) Invasive squamous cell carcinoma of the skin: defining a high-risk group. Annals of surgical oncology 13, 902–909 [PubMed: 16788750]

24. Karaa A, and Khachemoune A (2007) Keratoacanthoma: a tumor in search of a classification. International journal of dermatology 46, 671–678 [PubMed: 17614793]

25. Li M, Chen X, Li X, Ma B, and Vitanyi MB (2004) The similarity metric. IEEE Transactions on Information Theory 50, 3250–3264

26. Elmets CA, Ledet JJ, and Athar M (2014) Cyclooxygenases: mediators of UV-induced skin cancer and potential targets for prevention. The Journal of investigative dermatology 134, 2497–2502 [PubMed: 24804836]

27. Holmes C, Foley P, Freeman M, and Chong AH (2007) Solar keratosis: epidemiology, pathogenesis, presentation and treatment. The Australasian journal of dermatology 48, 67–74; quiz 75-66 [PubMed: 17535191]

28. Rossi R, Mori M, and Lotti T (2007) Actinic keratosis. International journal of dermatology 46, 895–904 [PubMed: 17822489]

29. Rowert-Huber J, Patel MJ, Forschner T, Ulrich C, Eberle J, Kerl H, Sterry W, and Stockfleth E (2007) Actinic keratosis is an early in situ squamous cell carcinoma: a proposal for reclassification. The British journal of dermatology 156 Suppl 3, 8–12 [PubMed: 17488400]

30. Lohmann CM, and Solomon AR (2001) Clinicopathologic variants of cutaneous squamous cell carcinoma. Adv Anat Pathol 8, 27–36 [PubMed: 11152092]

31. (2010) Squamous Cell Carcinoma: A Review of Etiology, Pathogenesis, Treatment, and Variants. Journal of the Dermatology Nurses' Association 2, 17–18
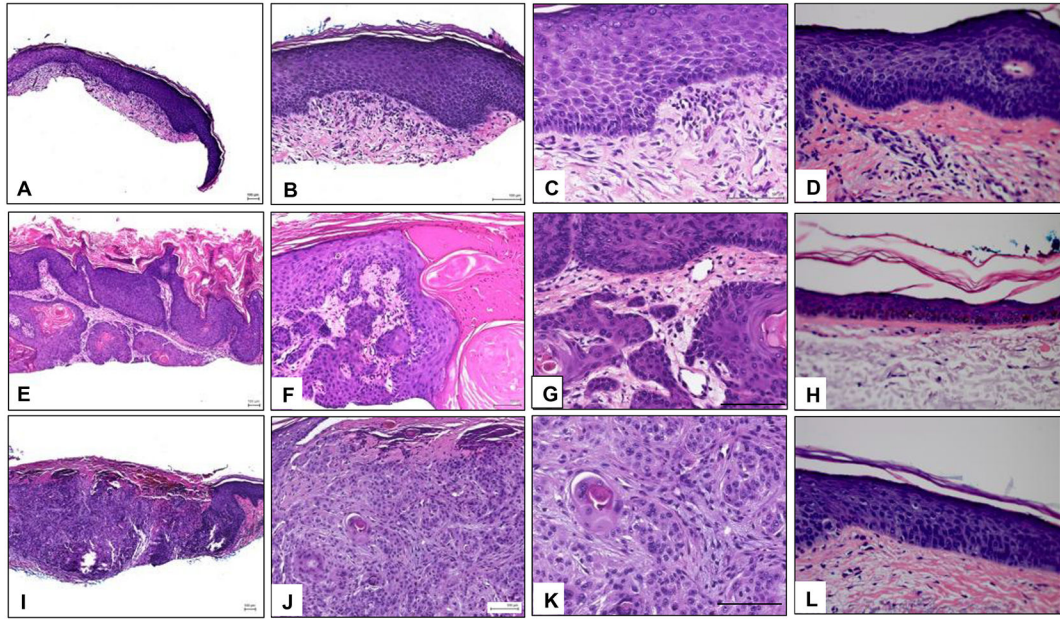
Figure 1.

Histopathology of representative samples. **A-C)** Images of a representative AK sample at 4x, 10x, and 20x, respectively, showing epidermal hyperplasia and dysplasia and marked hyper and parakeratosis with alteration of the ortho and parakeratotic keratin. The dermis shows solar elastosis and an infiltrate of mononuclear cells. **D)** Matched normal skin adjacent to the AK sample. **E-G)** Images of a representative KA sample at 4x, 10x, and 20x, respectively. Epidermis shows a crater-like invagination filled with ortho and parakeratotic horn. The lining of the invagination is formed by proliferating dysplastic squamous epithelium. **H)** Matched normal skin adjacent to the KA sample. **I-K)** Images of a representative SCC sample at 4x, 10x, and 20x, respectively. Arising from the epidermis and extending into the dermis there are aggregates of dysplastic keratinocytes. **L)** Matched normal skin adjacent to the SCC sample. AK: Actinic keratosis, KA: keratoacanthoma, SCC: squamous cell carcinoma.
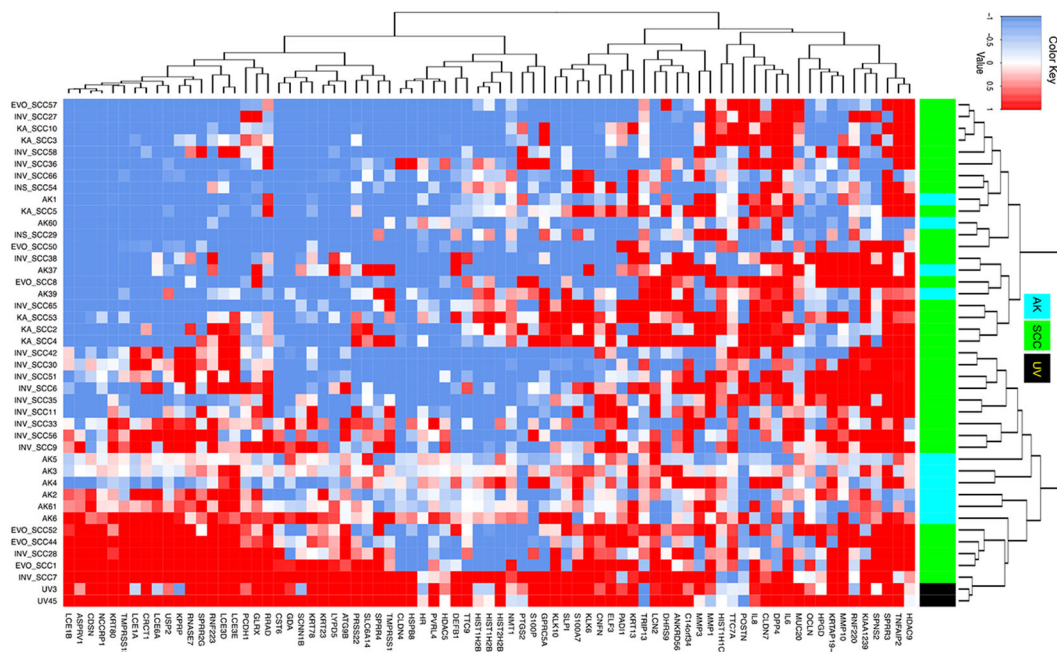
Figure 2.
Heatmap showing the gene expression signatures (log 2-fold change) of 77 genes in each of the UV, SCC, or AK samples versus their matched normal skin control. Red, white, and blue colors indicate over-expressed, not differentially expressed, or under-expressed genes, respectively. Black, cyan, and green colors in the sidebar represent UV, AK, and SCC samples, respectively.

Figure 3.
A) Receiver-operating curves (ROC) curves showing the training and the testing performance of the Logistic regression model using the 29 training samples with LOOCV (left) and the 12 testing samples (right), respectively. B) Heatmap showing clustering of AK samples in orange and SCC samples in green based on the 13 features selected by the model.
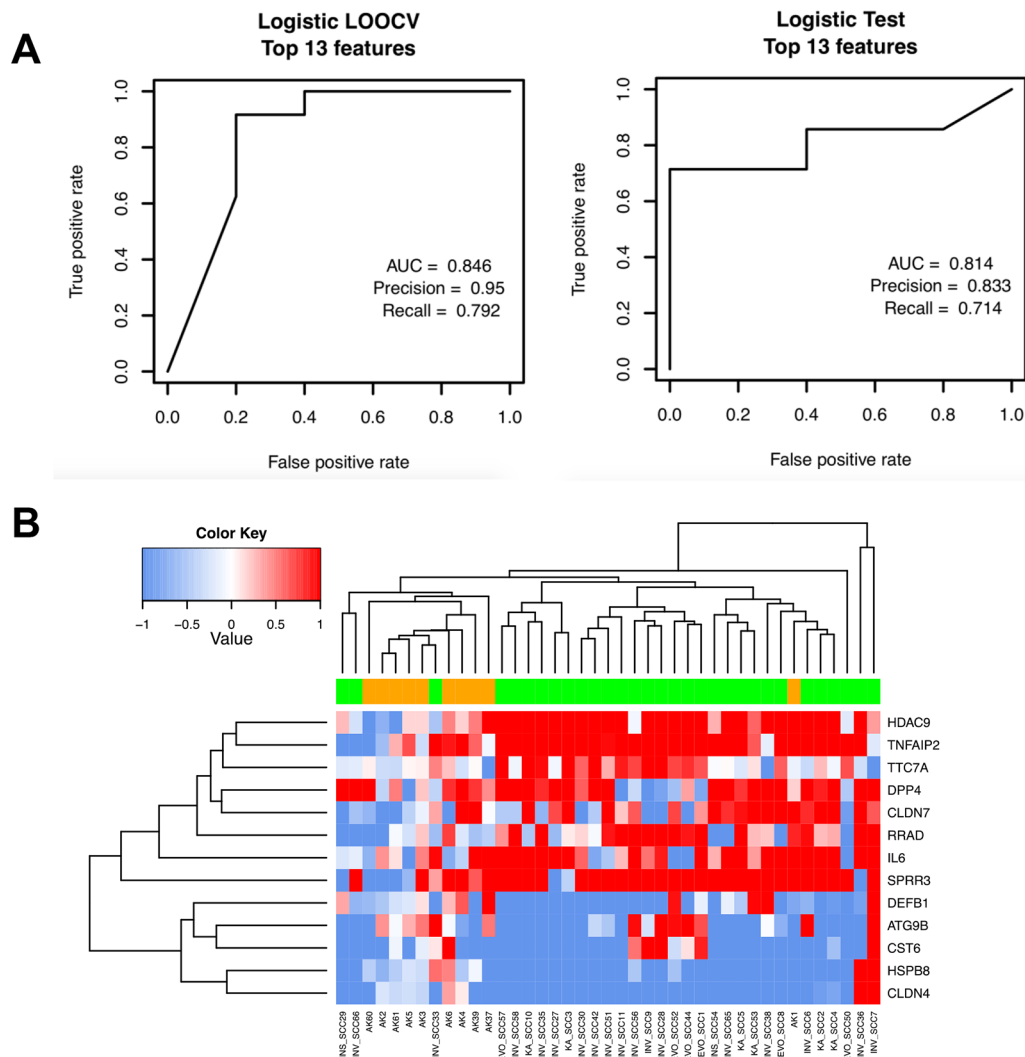
Figure 4.
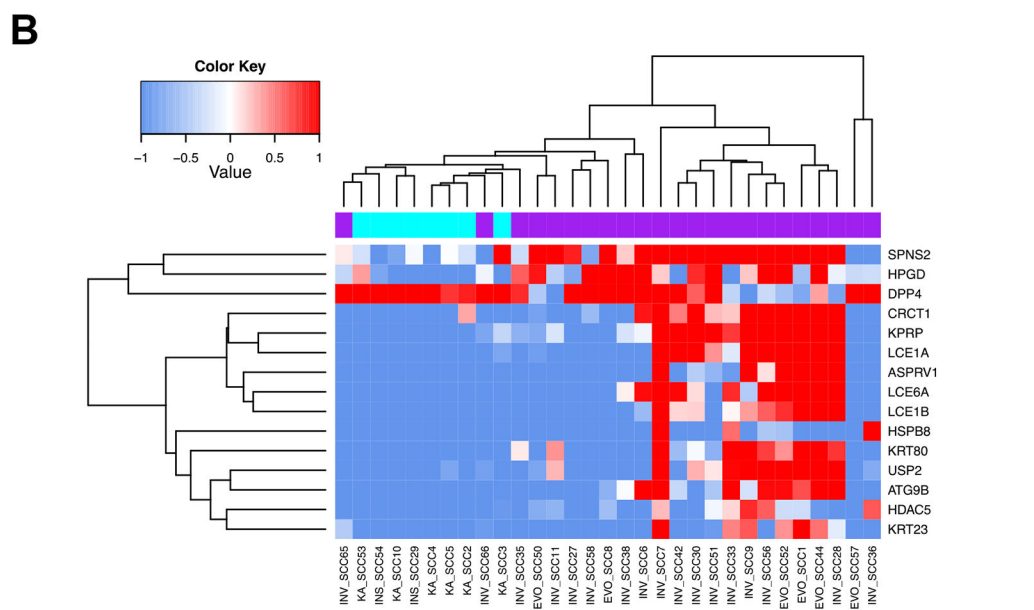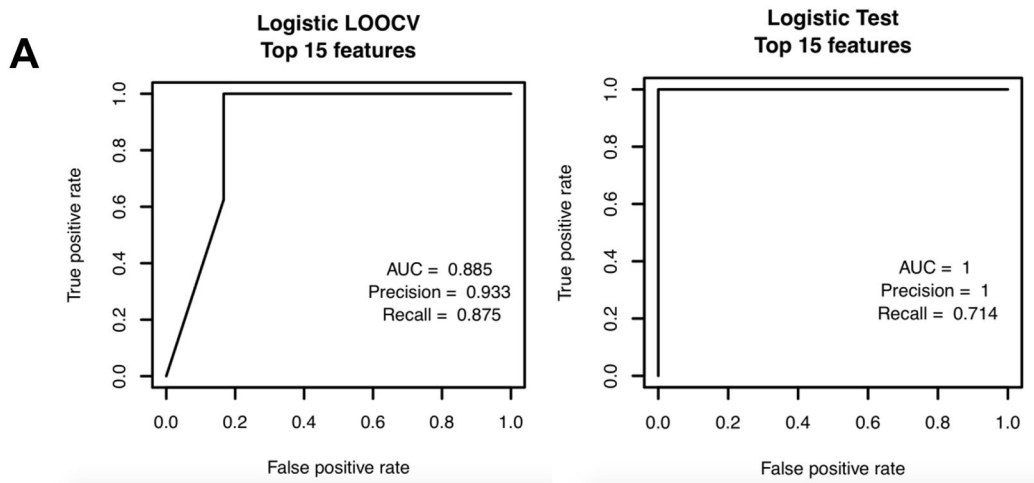A) Receiver-operating curves (ROC) curves showing the training and the testing performance of the logistic regression model using the 22 training SCC samples with LOOCV (left) and in the 9 testing SCC samples (right), respectively. B) Heatmap showing the clustering of more concerning SCC (purple) and less concerning SCC (cyan) signatures based on the 15 UV biomarker genes selected by this model.

Figure 5.
Differential expression of SCCs and AKs and more concerning versus less concerning SCCs. A) Boxplot of genes with lower expression values in AK signatures than in SCC signatures. B) Boxplot of genes with higher expression values in AK signatures than in SCC signatures. C) Boxplot of genes with lower expression values in more concerning SCC signatures than in less concerning SCC signatures. D) Boxplot of genes with higher expression values in more concerning SCC signatures than in less concerning SCC signatures.

Table 1.

Patient and tumor characteristics*

Sample ID	Age	Sex	Race/Ethnicity	Patient History	Specimen Location	Histopathologic Diagnosis
AK-1	67	M	White/NHL	OT (liver), B	Forehead	AK
AK-2	67	M	White/NHL	OT (liver), B	Cheek	AK
AK-3	79	M	White/NHL	B, P	Shoulder	AK
AK-4	79	M	Not specified	B, P	Arm	AK
AK-5	79	M	Not specified	B, P	Chest	AK
AK-6	86	M	White/NHL	B, P	Dorsal hand	AK
AK-37	73	M	White/NHL	OT (renal), B, M, P	L shoulder	AK
AK-39	85	M	White/NHL	P	Mid parietal scalp	AK
AK-60	76	M	White/NHL	N	R forearm	AK
AK-61	76	M	Not specified	N	L scalp	AK
SCC-1	93	M	White/NHL	L	Posterior scalp	EVO
SCC-2	80	M	White/NHL	P	Scalp vertex	KA
SCC-3	70	M	Unknown	OT (lung), P	Forehead	KA
SCC-4	67	F	White/NHL	OT (lung), P	R lower leg	KA
SCC-5	53	F	Unknown	-	L lower leg	KA
SCC-6	77	M	White/NHL	-	Scalp	INV
SCC-7	74	F	Not specified	N	Scalp	INV
SCC-8	87	M	Not specified	P	Cheek	EVO
SCC-9	78	M	White/NHL	P	Temple	INV
SCC-10	70	M	White/NHL	OT (lung), P	Mandible	KA
SCC-11	70	M	White/NHL	OT (renal), B, P	Thumb base	INV
SCC-27	71	M	White/NHL	OT (lung), P	L parietal scalp	INV
SCC-28	83	M	Not specified	-	Crown	INV
SCC-29	78	M	White/NHL	L, M, N	Scalp	INS
SCC-30	90	M	White/NHL	B, P	R forehead	INV
SCC-33	84	M	White/NHL	B, O	Mid frontal hairline	INV
SCC-35	57	M	Native Hawaiian or Pacific Islander	L, P, S	L crown	INV
SCC-36	75	M	Not specified	-	R temple	INV
SCC-38	63	M	White/NHL	OT (renal), B, M, S	L mid helix	INV
SCC-42	89	M	Not specified	-	L vertex scalp	INV
SCC-44	67	M	White/NHL	OT (renal), B	L nasal bridge	EVO
SCC-50	94	F	Not specified	B	L cheek	DIG
SCC-51	88	M	Not specified	P	L cheek	INV
SCC-52	58	M	White/Unknown	P	Nasal tip	EVO
SCC-53	65	M	White/NHL	OT (lung)	Upper chest	KA
SCC-54	85	M	White/NHL	-	R upper chest	INV

Sample ID	Age	Sex	Race/Ethnicity	Patient History	Specimen Location	Histopathologic Diagnosis
SCC-56	62	M	White/NHL	OT (liver)	L temporal	INS
SCC-57	90	M	White/Unknown	B, P	L dorsal wrist	EVO
SCC-58	69	F	White/NHL	OT (lung), P	R brow	INV
SCC-65	84	F	White/NHL	B, M, P	Central chest	INV
SCC-66	73	M	White/NHL	OT (renal), B, M, P	R dorsal Hand	INV

*
Abbreviations INV: invasive; INS: in situ; EVO evolving; KA: keratoacanthoma type; AK: actinic keratosis; DIG: digitated; OT: organ transplant patient; P: past history of SCC(s); B: history of BCC(s); N: history of NMSC (unknown type); M: history of melanoma; L: history of lymphoma; S: history of splenectomy; O: other; L: left; R: right; NHL: Not Hispanic or Latino; White/Unknown: white race/unknown ethnicity.