Gene Expression Patterns of Normal Human Skin, Actinic Keratosis, and Squamous Cell Carcinoma

A Spectrum of Disease Progression

R. Steven Padilla, MD, MBA; Sheldon Sebastian, MD; Zeyu Jiang, PhD; Ingo Nindl, PhD; Richard Larson, MD, PhD

Objectives: To identify and compare the gene expression profiles of actinic keratosis (AK) and squamous cell carcinoma (SCC) and to further clarify critical genetic alterations in the evolution of SCC from normal sun-damaged human skin.

Design: Observational study.

Setting: University practice.

Patients: Skin biopsy specimens were obtained from 16 patients. The specimens included 14 normal non–sun-exposed skin samples, 14 normal sun-exposed skin samples, 5 AKs, and 15 cutaneous SCCs.

Main Outcome Measures: Gene expression profiles from normal non–sun-exposed skin, normal sun-exposed skin, AKs, and SCCs.

Results: Using a highly astringent shrunken centroid threshold of 6.52 and the prediction analysis of microarrays, we identified 89 unique genes that most likely contribute to the molecular evolution of SCC. Our model was cross-validated using data from a separate study and clearly distinguishes between skin tumors (AK and SCC) and normal skin independent of sun exposure. Genes that were upregulated in AK and SCC were downregulated in normal skin, and genes that were downregulated in AK and SCC were upregulated in normal skin.

Conclusions: The finding of similar differentially expressed genes in AK and SCC confirms that AK is a precursor lesion of SCC and indicates that they are closely related genetically. Clear elucidation of these relationships will be critical to improving therapeutic approaches.

Arch Dermatol. 2010;146(3):288-293
keratinocytes. Studies of metastatic SCC have evaluated changes in proteases, adhesion molecules, and growth regulation pathways. Specific matrix metalloproteinases (MMP-2, MMP-7, and MMP-12) have been detected in SCC, along with the expression of E-cadherens and P-cadherens.

The objectives of our study were to identify and compare the gene expression profiles of AK and SCC and to further clarify critical genetic alterations in the evolution of SCC from normal sun-damaged human skin. Using microarray techniques, we identified progressive alterations in the gene expression patterns of human keratinocytes at different disease stages.

Methods

Patient Demographics and Skin Sampling Methods

Patients included in this study had biopsy-proved SCC at the Department of Dermatology, University of New Mexico. At the time of SCC treatment, patients provided informed consent and were enrolled into the study. Human research review committee approval was obtained for all study participants. Up to 4 samples were collected from 16 patients and consisted of specimens from normal non–sun-exposed skin, normal sun-exposed skin, AK, and SCC. Following surgical removal of each specimen, a "rough" microdissection was performed to remove surrounding normal tissue. The fresh tissue was then snap frozen in liquid nitrogen and transported to the laboratory for RNA isolation using a nontoxic aqueous reagent designed to protect RNA from degradation that does not dissolve or disrupt the structure of tissue (RNA later; Ambion, Austin, Texas). High-density gene microarray studies were then performed on all RNA samples. As summarized in Table 1, most of our patients were male, which is higher than the 2:1 ratio of men to women generally seen for cutaneous SCC. A total of 48 samples were analyzed. Not all samples were suitable for microarray studies; 4 of 48 specimens analyzed included normal non–sun-exposed skin, normal sun-exposed skin, AK, and SCC samples from the same patient.

Oligonucleotide Microarray Hybridization and Data Preprocessing

Three to 4 site-specific RNA samples were obtained from each patient. To standardize the analysis, the same quantity of total RNA (3.0 µg) was labeled from each patient sample obtained. This was based on the amount obtained from the microdissected tissues. Total RNA from each sample was converted to biotinylated complementary RNA (cRNA) per the manufacturer's suggested protocol (One-Cycle Target Labeling Kit; Affymetrix, Santa Clara, California). Biotinylated cRNA was synthesized by in vitro transcription (GeneChip IVT Labeling Kit, Affymetrix).

Biotin-labeled cRNA was purified (GeneChip Sample Cleanup Module, Affymetrix), and the absorbance was measured at 260 nm to determine its yield (NanoDrop Spectrophotometer; NanoDrop Technologies Inc, Wilmington, Delaware). Twenty micrograms of labeled cRNA was fragmented, and the quality was assessed for purified and fragmented cRNA (2100 Bioanalyzer and RNA 6000 Nano LabChip Kit; Agilent Technologies, Santa Clara). The labeled fragmented cRNA was hybridized to microarrays (GeneChip HGU133, version 2.0 Plus; Affymetrix) for 16 hours at 45°C according to the manufacturer's protocol for this microarray type. Washing and staining were performed according to the antibody used (Fluidics Station 450 and GeneChip Scanner 3000, Affymetrix). Microarray data were normalized by robust microarray analysis. One sample that had unusual residuals due to hybridization was excluded from the study.

Statistical Microarray Analysis

We used order-restricted inference as an exploratory tool to identify gene expression patterns, and we performed prediction analy-
s of microarrays (PAM) for gene selection and supervised class prediction.18 PAM uses a nearest shrunken centroid classifier to develop a predictive signature for 2 or more classes. As an improvement over the standard nearest centroid classification, PAM also “shrinks” each of the class centroids toward its overall mean, termed a threshold, to improve the prediction accuracy. PAM has become the criterion standard for comparison of microarray data among groups. PAM applies k-fold cross-validation in selecting the optimal number of genes by controlling overall and individual class cross-validation error rates. The maximal k is set to the least number of samples in all classes. In addition, we performed multiple analyses using PAM by setting different random number seeds. Although the resultant ranking of genes stayed the same, the “optimal” number of genes based on cross-validation usually changes with different runs, and we averaged the results from 6 runs. Hierarchical clustering was performed using the R package gplots (http://cran.rproject.org/src/contrib/Descriptions/gplots.html). Principal component analysis (PCA) was accomplished in R, for which a 3-dimensional plot was generated by projecting data points to the first 3 principal components; the image was developed using the R package rgl. To search for gene network pathways, we searched BioCarta, KEGG, and Reactome pathways (http://www.ingenuity.com) and available software programs (https://www.affymetrix.com/products/software/compatible/pathway.affx).

RESULTS

Cross-validation by independent data set

By searching PubMed (http://www.ncbi.nlm.nih.gov/sites/entrez) and microarray databases, we identified a study by Nindl et al19 that used 15 gene microarrays (U133A, Affymetrix) to identify differentially expressed genes in cutaneous SCC obtained from immunosuppressed organ transplant recipients. The data set was retrieved from the Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE2503. Although the microarray platform used by Nindl et al is several years old, it is compatible with the platform that was used in our experiments (U133A, version 2.0 Plus; Affymetrix). In addition, the publicly released data from their study were generated by an algorithm (MAS, version 5.0; Affymetrix), which is an older method of data standardization. Because of difficulty in obtaining the raw .CEL files from the authors, we elected to standardize and perform cross-validation. The GSE2503 data set contained 15 chips (4 AK, 5 SCC, and 6 normal). The U133A platform contains roughly 22 000 probe sets covering approximately 14 500 genes, whereas the U133A 2.0 Plus platform contains roughly 54 000 probe sets covering 38 500 well-characterized human genes. Extensive normalization procedures were undertaken in the earlier study, which we replicated. These included setting the expression values that were less than 300 to 300, normalizing samples by their respective median, and then normalizing each gene by its median. Finally, the authors used PAM after normalization of the data. We performed the same steps of data normalization except that expression values less than 35.3 were set to 35.3 based on rank order difference of the genes in 2 platforms (Affymetrix). Using 2277 probe sets common to the 2 platforms and the study by Nindl et al, we performed independent analyses by PAM on normalized data using robust microarray analysis, which included 2-class and 3-class sample classifications.

Our findings indicate that AK and SCC are genetically related. Their similar differentially expressed genes confirm that AK is a precursor lesion of SCC.

Overall change in gene expression

We initially set out to ascertain whether AK and SCC samples showed different gene expression overall. We used order-restricted inference to analyze 48 patient samples obtained from normal non–sun-exposed skin, normal sun-exposed skin, AKs, and SCCs. Our data revealed direct correlation of abnormal gene expression in the progression of normal skin to AK to SCC (Figure 1). A total of 186 genes were statistically significant at a threshold of $P = .0001$. Among them, 101 genes were expressed progressively higher along the spectrum of normal non–sun-exposed skin to SCC, and 85 genes were expressed progressively lower along this spectrum of disease. The degree of gene expression was most evident in AK and SCC, whereas normal skin appeared to have minimal alteration of gene expression. Because order-restricted inference is not a multivariate approach, validation of gene sets as a genetic classifier is difficult using this procedure alone. These findings support the concept that AK and SCC are genetically similar.
GENETIC COMPARISON OF SKIN LESIONS

To determine whether AK and SCC could be distinguished by microarray analysis, we used high-stringency standards to identify 258 “signature genes” (false discovery rate, <1 gene), which could then be used to measure the overall pattern of gene expression in the progression from normal skin to actinically damaged skin and ultimately to SCC. We performed PAM to select genes that can predict the clinical outcome of “normal skin” vs AK or SCC based on the U133A 2.0 Plus platform. As shown in Figure 2, the lowest overall cross-validation error rate was 0.43% when choosing a centroid threshold of 6.52, which corresponded to 89 unique genes (eTable, http://www.archdermatol.com). Based on cross-validation, all normal non–sun-exposed or normal sun-exposed cases were correctly classified, and only 1 case of AK or SCC was erroneously classified as a normal non–sun-exposed or normal sun-exposed case.

We then used 89 unique genes identified by PAM and shrunken centroid analysis in a hierarchical cluster analysis (Figure 3). Hierarchical clustering of samples and genes delineates distinct clustering of AK or SCC skin...
and normal non–sun-exposed or normal sun-exposed skin. This analysis demonstrates the genetic relationship between the lesions studied. We identified the following 2 gene families, without any outliers: (1) an SCC family consisting of 21 AK and SCC clustered samples and (2) a normal skin family consisting of 27 clustered normal skin samples (non–sun exposed and sun exposed). In addition, our molecular classifier revealed the following 2 distinct gene groups: (1) genes with upregulated expression in AK and SCC and (2) genes with downregulated expression in AK and SCC. The opposite expression profile was found in the normal skin samples. This provides strong evidence that the genetic alterations leading to AK and SCC are similar. It is notable that 5 of 6 AKs clustered in a subgroup of the SCC family, although a subset of genes that statistically distinguished AK and SCC could not be identified.

### CROSS-VALIDATION STUDY

To assess independent verification of the molecular classifier obtained in our study, we used data by Nindl et al19 from 15 U133A gene microarrays to identify differentially expressed genes in AK and SCC. This platform was compatible with the U133A 2.0 Plus platform used for our analysis. Their data set included 4 AK, 5 SCC, and 6 normal skin specimens. A set of 118 genes was presented in their study, with none showing a statistically significant difference between AK and SCC. After normalizing their data, we found an overlap of 99 common genes with our 89-member gene classifier. Those 99 common genes were used to ascertain how well we could predict the clinical samples from their study. Our 99-gene classifier was able to predict with 100% accuracy the origin of their samples (ie, normal skin vs AK or SCC) (Table 2). In addition, the heat map (Figure 4) shows hierarchical clustering of the samples and genes and again outlines adjacent clustering of AK or SCC samples and normal skin samples. This provides further evidence that our gene classifier is a legitimate tool for distinguishing AK and SCC.

### COMMENT

The case for a close genetic relationship between AK and SCC continues to grow, as does evidence that UV radiation acts as a key initiating carcinogenic event. We set out to evaluate the entire spectrum of disease leading to SCC. Based on human samples obtained from non–sun-exposed and sun-exposed skin, AK and SCC in our analysis demonstrate a continuum and an expansion of gene expression along this disease spectrum. The upregulation and downregulation of genes shown in Figure 1 were continual among common genes in normal skin and became dramatically more expressed in AK and SCC. Furthermore, hierarchical clustering demonstrated that AK and SCC are closely related genetically and can be distinguished from benign lesions. Whether these genes are causal and are the critical carcinogenic events that trigger the evolution of UV-damaged skin to SCC is unclear, but our results identify potential novel genes to investigate for their role in skin carcinogenesis.

To our knowledge, this work is the first to identify genetic signatures that clearly distinguish genetic predictors of AK or SCC. Using a highly stringent shrunked centroid threshold of 6.52 (Figure 2) and PAM, we identified 89 unique genes that most likely contribute to the molecular evolution of SCC. Our model clearly distinguishes between skin tumors (AK and SCC) and

---

**Table 2. Results From Our Prediction Analysis of Microarrays Model for 15 Patient Samples From the Study by Nindl et al**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Diagnosis by Nindl et al<strong>19</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted diagnosis using our 99-member gene classifier</td>
<td>Normal Skin</td>
</tr>
<tr>
<td>Actual diagnosis</td>
<td>6</td>
</tr>
</tbody>
</table>

Abbreviations: AK, actinic keratosis; SCC, squamous cell carcinoma.

*a For further information, see the hierarchical clustering in Figure 4.
mal skin independent of sun exposure. As illustrated by the heat map (Figure 3), 2 distinct gene groups were apparent. Genes that were upregulated in AK and SCC were inherently downregulated in normal skin and vice versa. We validated these findings using data from a previous study\textsuperscript{19} that was unsuccessful in identifying genetic signatures for AK or SCC. The 89-member gene classifier herein clearly distinguished AK or SCC lesions from normal skin but not from each other. Although some of the gene classifiers are not entirely specific and it is possible that some gene classifiers may demonstrate overlapping results, they may overlap with several other related genes. As a result, it is impossible to determine whether a specific hybridization and a less specific gene target are indeed identical. Although the initial induction, progression, and sequential relationships are unclear, the gene predictors can be explored to determine their role in disease progression. Our data suggest a close and likely interdependence of gene families between AK and SCC.

This study reveals gene family correlation and critical commonalities and differences among AK, SCC, and clinically benign skin. Further studies should focus on the interrelationship of the key carcinogenic genes that lead to SCC and how their relative expression or sequence of expression influences clinical behavior. Clear elucidation of these relationships will be critical to improving therapeutic approaches.

Accepted for Publication: July 8, 2009.
Correspondence: Sheldon Sebastian, MD, Dermatology and Skin Cancer Center, 11550 Granada, Leawood, KS 66211 (sheldon.sebastian@gmail.com).

Author Contributions: Drs Padilla, Sebastian, and Nindl had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: Padilla. Acquisition of data: Padilla and Nindl. Analysis and interpretation of data: Padilla, Sebastian, Jiang, and Larson. Drafting of the manuscript: Padilla and Sebastian. Critical revision of the manuscript for important intellectual content: Nindl and Larson. Obtained funding: Padilla. Financial Disclosure: None reported.

Online-Only Material: An eTable is available at http://www.archdermatol.com.

REFERENCES