Oncogene (2007) 26, 6885–6895
© 2007 Nature Publishing Group All rights reserved 0950-9232/07 $30.00
www.nature.com/onc

ORIGINAL ARTICLE

# Expression profile of skin papillomas with high cancer risk displays a unique genetic signature that clusters with squamous cell carcinomas and predicts risk for malignant conversion

N Darwiche[1], A Ryscavage[2], R Perez-Lorenzo[3], L Wright[2], D-S Bae[2,4], H Hennings[2], SH Yuspa[2] and AB Glick[3]

[1]*Department of Biology, American University of Beirut, Beirut, Lebanon;* [2]*Laboratory of Cancer Biology and Genetics, National Cancer Institute, NIH, Bethesda, MD, USA and* [3]*Department of Veterinary and Biomedical Sciences, Center for Molecular Toxicology and Carcinogenesis, The Pennsylvania State University, University Park, PA, USA*

**Chemical induction of squamous tumors in the mouse skin induces multiple benign papillomas: high-frequency terminally benign low-risk papillomas and low-frequency high-risk papillomas, the putative precursor lesions to squamous cell carcinoma (SCC). We have compared the gene expression profile of twenty different early low- and high-risk papillomas with normal skin and SCC. Unsupervised clustering of 514 differentially expressed genes ($P<0.001$) showed that 9/10 high-risk papillomas clustered with SCC, while 1/10 clustered with low-risk papillomas, and this correlated with keratin markers of tumor progression. Prediction analysis for microarrays (PAM) identified 87 genes that distinguished the two papilloma classes, and a majority of these had a similar expression pattern in both high-risk papillomas and SCC. Additional classifier algorithms generated a gene list that correctly classified unknown benign tumors as low- or high-risk concordant with promotion protocol and keratin profiling. Reduced expression of immune function genes characterized the high-risk papillomas and SCC. Immunohistochemistry confirmed reduced T-cell number in high-risk papillomas, suggesting that reduced adaptive immunity defines papillomas that progress to SCC. These results demonstrate that murine premalignant lesions can be segregated into subgroups by gene expression patterns that correlate with risk for malignant conversion, and suggest a paradigm for generating diagnostic biomarkers for human premalignant lesions with unknown individual risk for malignant conversion.**
*Oncogene* (2007) **26,** 6885–6895; doi:10.1038/sj.onc.1210491; published online 21 May 2007

**Keywords:** skin carcinogenesis; microarray; squamous cell carcinoma; tumor immunology

Correspondence: Dr AB Glick, Center for Molecular Toxicology and Carcinogenesis, The Pennsylvania State University, University Park, PA 16801, USA or N Darwiche, Department of Biology, American University of Beirut, Beirut, Lebanon.
E-mails: abg11@psu.edu or nd03@aub.edu.lb
[4]Current address: Department of Pharmacology and Physiology, The George Washington University, Washington, DC 20037, USA.
Received 3 December 2006; revised 19 March 2007; accepted 19 March 2007; published online 21 May 2007

## Introduction

While cancer development in humans is a multi-step process often requiring years for the emergence of a malignant invasive tumor from premalignant lesions, it is not known if all premalignant lesions have an equal probability of progression or if a subset with higher risk for malignant potential exists. To test whether premalignant lesions have distinguishable stable gene expression profiles that define their risk for malignant progression, we have utilized a mouse model of chemically induced skin carcinogenesis where the multiple stages of cancer development are readily accessible and have been well characterized at both the tissue and molecular levels. Studies from multiple laboratories have shown that genetic changes occurring in this tumor model are relevant to pathways altered in human squamous cell carcinoma (SCC) and other cancers of lining epithelia (Yuspa, 1998). In this model the mouse epidermis is treated with the carcinogen dimethyl benz[*a*]-anthracene (DMBA) followed by 20 weekly applications of a tumor promoter, such as 12-*O*-tetradecanoylphorbol-13-acetate (TPA). Benign papillomas with initiating mutations in the c-Ha *ras* gene arise within 10–15 weeks, and a small percentage progress to locally invasive SCC. However, studies over the last decade have identified two subclasses of benign lesions with differing rate of premalignant progression and characterized several biomarkers that distinguish them at early stages (Hennings *et al*., 1985; Glick *et al*., 1993; Tennenbaum *et al*., 1993; Darwiche *et al*., 1996). Low-risk papillomas are terminally benign and do not progress to SCC, while a smaller fraction of benign tumors, termed high-risk papillomas, progress to SCC with high frequency. Differing tumor promotion protocols enrich for the low- or high-risk papilloma subtypes, and thus they can be readily and predictably isolated for further analysis (Hennings *et al*., 1985). To test the hypothesis that high-risk papillomas represent the precursor lesions to SCC at the molecular level, we have used oligo microarrays to compare the gene expression profile of early stages of benign low- and high-risk papillomas in relationship to normal skin, chronic TPA promoted normal skin and SCC.

## Results

### *Identification of differentially expressed genes in multistage skin carcinogenesis*

Cutaneous squamous tumors were generated and analysed by microarray analysis as described in Materials and methods, and in Supplementary methods. To identify differentially expressed genes, expression ratios from filtered arrays were compared using ANOVA and genes selected with a $P<0.001$ between at least one group, or Significance Analysis for Microarrays (SAM) with a shrinkage of 4.5 and a false discovery rate of 0.0023. Both analyses identified 514 genes (Supplementary Table I), with concordance of 87% between these two gene sets. Confirming the cross platform reproducibility of the mouse multistage skin model, this gene set contains highly induced and downregulated genes similar to another cDNA microarray study comparing pooled papillomas, SCC- and TPA-treated normal skin (Hummerich *et al*., 2006), such as S100a9, S100a8, S100a11, small proline-rich proteins Sprr2f, Sprr2k, Slpi, the ras family member Arhh, Fabp5 and several members of the glutathione *S*-transferase family (Supplementary online dataset). Additional highly induced genes across all tumors identified in our dataset include markers of inflammation such as interleukin-1$\beta$ (IL-1$\beta$), Ear5, pregnancy-specific glycoprotein 28 (Psg28) and the cysteine protease inhibitor stefin A1. Among the most highly downregulated genes across all tumors are mesenchymal proteins actin $\alpha$1, $\gamma$2, myosin1A and the cartilage associated homeobox gene Sox5.

We used semi-quantitative reverse transcriptase (RT)–PCR (Figure 1) and quantitative RT–PCR (Supplementary Table II) to validate the expression patterns for seven genes that were overexpressed in tumors (IL-1$\beta$, S100a8, S100a9, 2610528A11Rik, Stfa1, Birc5, Arg1); eight genes downregulated in tumors (2510006C20Rik, 1700013N18Rik, ChChd5, Dctn1, Rtn4, Nibrin, Mapk8ip3, Rbm12); and von Hippel–Lindau syndrome homologue a gene that distinguishes low- and high-risk tumors (Table 1). The trend was similar, but expected variability in the ratios between the two methods was observed due to calculated average values obtained from different numbers of tissues and/or method sensitivity. Taken together these results as well as concordance of expression patterns in this dataset with published expression and array studies (Hummerich *et al*., 2006) show that the expression patterns observed with the oligo microarrays reflect valid differences in gene expression between groups of tumors.

### *High-risk papillomas cluster with squamous cell carcinomas*

We used multidimensional scaling (MDS) to represent in a three-dimensional space the relationship between each tumor based on the expression pattern. This analysis showed close similarity between the high-risk papillomas and SCC with the low-risk papillomas clustering in a tight group distinct from these tumors and normal skin (Figure 2a). Similarly, unsupervised hierarchical clustering using a Pearson correlation of the individual tumor arrays generated a dendrogram that linked low-risk tumors to the normal skin samples, while the high-risk papillomas to that of the SCC (Figure 2b). Three of the malignant tumors clustered with the low-risk papillomas while one low-risk papilloma, 1B-50B, clustered with the SCC, even though histological examination of 1B-50B did not reveal invasion through the basement membrane (data not shown). In both analyses, one tumor generated using a high-risk promotion protocol, 1B-29B, clustered with the low-risk papillomas (Figure 2a and b, arrow). To determine if this gene expression clustering correlated with previously identified markers of progression, the tumors were characterized for expression of keratins (K) 1, 10, 13 and 8 (Larcher *et al*., 1992; Tennenbaum *et al*., 1993). As expected 10/10 low-risk papillomas expressed K1 and K10 throughout while K13 was absent or expressed in small foci (Figure 3). In contrast, 9/10 high-risk papillomas expressed K13 abundantly and exhibited varying degrees of K1 loss. In concordance with its expression profile, tumor 1B-29B generated with a high-risk promotion protocol had a keratin profile similar to low-risk papillomas (Figure 3). None of the benign tumors were K8-positive while all of the SCC expressed this marker, suggesting that expression profiling can identify premalignant lesions before full acquisition of a malignant phenotype.

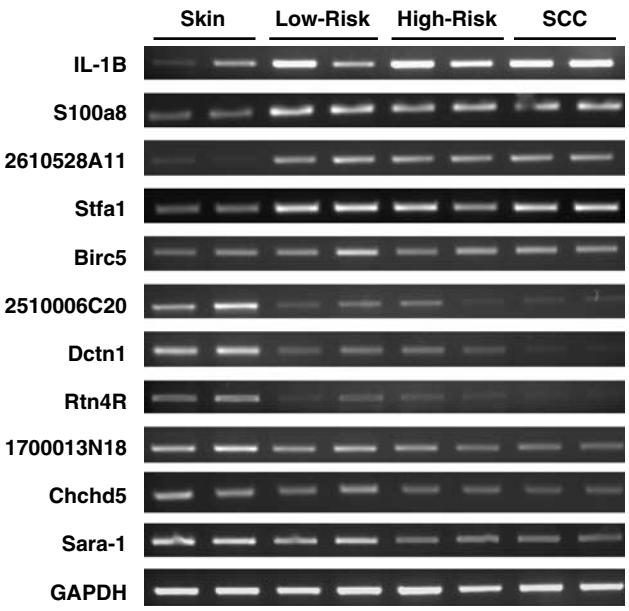



**Figure 1** Validation of microarray data by semi-quantitative reverse transcriptase (RT)–PCR. Two independent tissue samples from normal skin, low- and high-risk tumors and squamous cell carcinoma were selected for microarray data validation by RT–PCR. Semi-quantitative RT–PCR on 11 genes from the microarray analysis was performed as described in Materials and methods.

### *A molecular signature of low-risk papillomas linked to downregulated innate and adaptive immunity*

To identify genes that distinguish benign lesions with differing risk for malignant progression, we used the

**Table 1** Expression ratios of genes significantly different between low- and high-risk papillomas

| *Gene* | *Description* | *UniGene* | *Skin* | *Chronic TPA* | *Low-risk* | *High-risk* | *SCC* |
|---|---|---|---|---|---|---|---|
| $\beta$-Globin | Mus musculus $\beta$ globin | | 1.0 | 13.9 | 1.4 | 9.7 | 11.0 |
| 5.7SRNA | 5.7S U4 small nuclear RNA type a | | 1.0 | 38.1 | 0.8 | 5.8 | 4.5 |
| 4.5SRNA | 4.5S RNA | | 1.0 | 1.1 | 5.1 | 0.8 | |
| 7SRNA | 7S small cytoplasmic RNA | | 1.0 | 3.3 | 0.9 | 4.0 | 2.1 |
| 1810064A23RIKEN | | | 1.0 | 3.5 | 0.8 | 3.3 | 1.8 |
| 2210019G11RIKEN | | | 1.0 | 5.0 | 0.8 | 2.7 | 1.9 |
| Galk1 | Galactokinase 1 | Mm.213873 | 1.0 | 0.5 | 7.6 | 2.4 | 1.2 |
| 2610303G11RiKEN | | Mm.53530 | 1.0 | 1.2 | 9.0 | 2.4 | 2.7 |
| Icam1 | Intercellular adhesion molecule | Mm.90364 | 1.0 | 1.2 | 0.7 | 2.2 | 1.4 |
| Col5a3 | Procollagen, type V, $\alpha$ 3 | Mm.334994 | 1.0 | | 8.0 | 2.2 | 3.2 |
| Rnf3 | Polycomb group ring finger 3 | Mm.261218 | 1.0 | 1.7 | 0.8 | 2.1 | 0.9 |
| TSAP2 | p53-induced apoptosis differentially expressed | Mm.366224 | 1.0 | 1.2 | 0.4 | 2.0 | 0.7 |
| Glt8d1 | Glycosyltransferase 8 domain containing 1 | Mm.8766 | 1.0 | 0.9 | 0.4 | 1.9 | 1.6 |
| 4931423N10RIKEN | | Mm.314026 | 1.0 | 1.7 | 0.5 | 1.8 | 1.1 |
| Vhlh | Von Hippel–Lindau syndrome homolog | Mm.29407 | 1.0 | 0.9 | 0.6 | 1.7 | 1.1 |
| Eml4 | Echinoderm microtubule-associated protein like 4 | Mm.295565 | 1.0 | 1.3 | 0.6 | 1.7 | 0.9 |
| Erf1 | Mus musculus ETS-related transcription factor ERF | | 1.0 | 2.1 | 0.4 | 1.6 | 1.2 |
| Nek4 | NIMA-related expressed kinase 4. | Mm.251494 | 1.0 | 0.6 | 8.7 | 1.1 | 1.1 |
| Prkacb | Protein kinase, cAMP-dependent, catalytic, $\beta$ | Mm.16766 | 1.0 | 5.1 | 3.3 | 1.1 | 1.1 |
| D19Ertd386e DNA segment, Chr 19, ERATO Doi 386 | | Mm.254169 | 1.0 | 0.3 | 4.1 | 1.0 | 1.2 |
| Uck2 | Uridine-cytidine kinase 2 | Mm.280895 | 1.0 | 0.5 | 4.4 | 0.9 | 1.2 |
| Armc9 | Armadillo repeat containing 9 (Armc9), mRNA | Mm.379264 | 1.0 | 0.6 | 1.9 | 0.7 | 0.3 |
| Rexo1 | RNA exonuclease 1 homolog | Mm.1116 | 1.0 | 0.4 | 3.1 | 0.6 | 1.6 |
| Oaz1 | Ornithine decarboxylase antizyme | Mm.683 | 1.0 | 0.5 | 1.5 | 0.6 | 0.7 |
| Zscan2 | Zinc finger and SCAN domain containing 2 | Mm.358645 | 1.0 | 0.5 | 1.9 | 0.6 | 0.5 |
| Plekhm2 | Pleckstrin homology domain containing | Mm.33220 | 1.0 | 0.5 | 2.1 | 0.6 | 0.9 |
| Lgals2 | Galectin2 | Mm.379288 | 1.0 | 0.2 | 2.6 | 0.6 | 0.4 |
| Trpc6 | Transient receptor potential cation channelC6 | Mm.325086 | 1.0 | 0.3 | 2.6 | 0.5 | 0.4 |
| Ttn | Titin transcript variant N2-B | Mm.373672 | 1.0 | 0.8 | 2.3 | 0.5 | 1.0 |
| Calcoco1 | Calcium binding and coiled-coil domain 1 | Mm.71682 | 1.0 | 0.2 | 2.0 | 0.5 | 0.4 |
| Sgta | Small glutamine-rich TPR-containing, $\alpha$ | Mm.30068 | 1.0 | 1.2 | 1.3 | 0.4 | 0.4 |
| Car3 | Carbonic anhydrase 3 | Mm.300 | 1.0 | 1.9 | 1.4 | 0.4 | 0.5 |
| Dnajc17DnaJ (Hsp40) homolog | | Mm.222241 | 1.0 | | 1.3 | 0.3 | 0.5 |
| Psmd2 | Proteasome 26 S subunit | Mm.243234 | 1.0 | 0.8 | 1.1 | 0.3 | 0.6 |
| Gnao1 | Guanine nucleotide binding protein, $\alpha$ o | Mm.354720 | 1.0 | 1.1 | 0.9 | 0.3 | 0.4 |
| Rspo1 | R-spondin homolog | Mm.42202 | 1.0 | 0.7 | 0.8 | 0.3 | 0.3 |
| Scn1b | Sodium channel, voltage-gated, type I, $\beta$ | Mm.1418 | 1.0 | 0.4 | 1.2 | 0.3 | 0.7 |
| Ggtl3 | $\gamma$- Glutamyltransferase-like 3 | Mm.41757 | 1.0 | 0.4 | 1.1 | 0.3 | 0.6 |
| col20a1 | Collagen 20$\alpha$1 | Mm.156855 | 1.0 | 0.2 | 1.2 | 0.2 | 0.2 |
| Mns1 | Meiosis-specific nuclear structural protein 1 | Mm.14485 | 1.0 | 0.4 | 1.3 | 0.2 | 0.5 |
| Tcte1 | t-Complex-associated testis expressed 1 | Mm.1290 | 1.0 | 0.6 | 1.0 | 0.2 | 0.5 |
| *Immune function genes* | | | | | | | |
| IgK chain mRNA V4J2' | | Mm.333130 | 1.0 | 1.0 | 9.2 | 2.4 | 4.7 |
| NC41 | | Mm.305097 | 1.0 | 0.5 | 2.4 | 0.5 | 0.4 |
| LOC56304 | Single-chain Ig VH and VL domains | Mm.358673 | 1.0 | 1.1 | 0.2 | 0.3 | |
| Ig $\kappa$ chain gene | | | 1.0 | 0.4 | 1.1 | 0.2 | 0.3 |
| LOC545850 similar to Ig light chain variable region | | Mm.380074 | 1.0 | 0.3 | 1.1 | 0.3 | 0.3 |
| Igk-V21 | | Mm.372601 | 1.0 | 0.4 | 1.0 | 0.2 | 0.2 |
| Igh-6 | IgM immunoglobulin heavy chain 6 | Mm.342177 | 1.0 | 0.5 | 0.9 | 0.3 | 0.2 |
| IgG light chain gene, V region | | Mm.360535 | 1.0 | 0.3 | 0.8 | 0.2 | 0.2 |
| Igk-V8 | Immunoglobulin $\gamma$ chain, mAb 667 | Mm.333124 | 1.0 | 0.3 | 0.8 | 0.1 | 0.1 |
| H26 immunoglobulin $\kappa$ chain variable region 12/13 | | Mm.304150 | 1.0 | 0.7 | 0.5 | 0.2 | 0.3 |

**Table 1** (*continued*)

| *Gene* | *Description* | *UniGene* | *Skin* | *Chronic TPA* | *Low-risk* | *High-risk* | *SCC* |
|---|---|---|---|---|---|---|---|
| T-cell receptor $\beta$ chain V$\beta$5 repertoire | | | 1.0 | | 1.6 | 0.4 | 0.5 |
| T-cell receptor $\beta$ chain mRNA, partial cds | | | 1.0 | | 1.5 | 0.3 | 0.3 |
| Tcrb-V13 | T-cell receptor $\beta$, variable 13, mRNA | Mm.333026 | 1.0 | 0.4 | 1.4 | 0.5 | 0.6 |
| T-cell receptor $\beta$ chain mRNA | | | 1.0 | | 1.4 | 0.3 | 0.3 |
| T-cell receptor $\beta$ chain mRNA, partial cds | | | 1.0 | 0.5 | 1.1 | 0.4 | 0.4 |
| T-cell receptor $\beta$ chain mRNA, partial cds | | | 1.0 | 0.5 | 0.9 | 0.4 | 0.4 |
| T-cell receptor $\alpha$ chain | | | 1.0 | 0.3 | 0.9 | 0.4 | 0.2 |
| T cell receptor $\beta$ chain mRNA, partial cds | | | 1.0 | 0.5 | 0.8 | 0.3 | 0.4 |
| Tcrb T-cell receptor $\beta$ | | Mm.380431 | 1.0 | 0.1 | 0.6 | 0.1 | 0.2 |
| Cx3cl1 | Chemokine (C-X3-C motif) ligand 1 | Mm.103711 | 1.0 | 1.5 | 9.6 | 1.6 | 2.3 |
| Traf2 | Tnf receptor-associated factor 2 | Mm.3399 | 1.0 | | 2.1 | 8.6 | 1.8 |
| Hamp1 | Hepcidin antimicrobial peptide 1 | Mm.23995 | 1.0 | 1.1 | 6.2 | 1.5 | 0.9 |
| Hemgn | Hemogen (Hemgn), mRNA | Mm.25793 | 1.0 | 0.4 | 1.8 | 0.2 | 0.8 |
| Laf4 | AF4/FMR2 family, member 3 | Mm.336679 | 1.0 | 0.4 | 1.2 | 0.2 | 0.2 |
| Amphoterin | | Mm.366196 | 1.0 | 0.7 | 1.1 | 0.3 | 0.5 |
| IL15 | Shorter isoform of interleukin 15 | | 1.0 | | 1.0 | 4.4 | 7.7 |

PAM analysis was performed on the expression ratios of low-risk and low-risk papillomas, and the resulting gene list obtained at a $\delta$ of 2.8 and minimum misclassification error of 0.048. The resulting gene list was used to filter all tissue types and the expression ratios within each tissue class were averaged, and normalized to the ratio of normal mouse skin samples. cAMP, cyclic adenosine monophosphate; NIMA, never in mitosis gene a; PAM, prediction analysis for microarray; SCC, squamous cell carcinoma; TPA, 12-*O*-tetradecanoylphorbol-13-acetate.

prediction analysis for microarrays (PAM) (Tibshirani *et al*., 2002) tool present on the mAdB analysis program to identify genes that were significantly different between these two groups. For this analysis tumor 1B-29B was grouped with the low-risk papillomas. At the minimum misclassification error of 0.048, 87 genes were identified that classified the tumors into two groups. Of these genes, 54/87 (62%) had a similar expression pattern in both high-risk papillomas and SCC (Figure 4, Table 1). The most highly induced genes in high-risk papillomas were $\beta$-globin, three small nuclear RNAs, U4, U3A and 7S and several RIKEN sequences. In addition, the adhesion molecule intercellular adhesion molecule 1, the Ets-related transcription factor Erf1 and the tumor suppressor VHL were elevated in high-risk papillomas relative to low-risk papillomas. Surprisingly, genes directly involved in cell proliferation that were different between these two classes such as the mitotic regulator NIMA 4, uridine cytidine kinase 2, and ornithine decarboxylase antizyme Oaz1 were reduced in the high-risk papillomas. Reduced expression of signal transducers such as the guanine nucleotide binding protein $\alpha$ o, and cAMP-dependent protein kinase $\beta$-catalytic subunit also distinguished high-risk papillomas from low-risk tumors.

Twenty-eight percent (24/84) of the genes whose expression ratio was distinct between low and high-risk papillomas were associated with the immune response, and most were T-cell $\alpha\beta$ receptor (TCR) genes or genes encoding immunoglobulins that were downregulated in the high-risk papillomas. While $\gamma\delta$ TCR genes, characteristic of the most prominent epidermal T-cell class, were not present on the PAM list that segregated benign papillomas, within the entire 514 differentially expressed gene list there was a threefold reduction in expression of TCR$\delta$ (T-cell receptor $\delta$-chain mRNA) from low-risk papillomas to high-risk papilloma and SCC. Other genes associated with inflammation and inflammatory pathways such as Traf2, and PMSD2 (TRAP2) involved in the TNF receptor signaling pathway, the pro-inflammatory chemokine CX3CL1 (fractalkine) and Hamp1 (Peyssonnaux *et al*., 2006) were expressed at lower levels or were downregulated in the high-risk papillomas and SCC compared to the low-risk papillomas. Additionally, genes such as Laf4, a B-cell transcription factor (Ma and Staudt, 1996) and amphoterin (HMG-1B), a chromatin protein with proinflammatory properties (Dumitriu *et al*., 2005), and Lglas2 (galectin2), a $\beta$-galactosidase-binding protein that mediates lymphotoxin secretion (Ozaki *et al*., 2004), were significantly downregulated in both the high-risk papillomas and SCC. The alternatively spliced form of IL15, a cytoplasmic isoform of the cytokine that impairs $\gamma$-IFN production and T- and NK-cell responses (Nishimura *et al*., 1998, 2000), was also highly upregulated in the high-risk papillomas and SCC. These data suggest that one early difference between low-risk and high-risk benign papillomas is the degree of anti-tumor immune response. To test this, both types of papillomas were assayed for the level of infiltrating lymphocytes using immunohistochemistry for the T-cell marker CD3$\varepsilon$. Figure 5 shows that

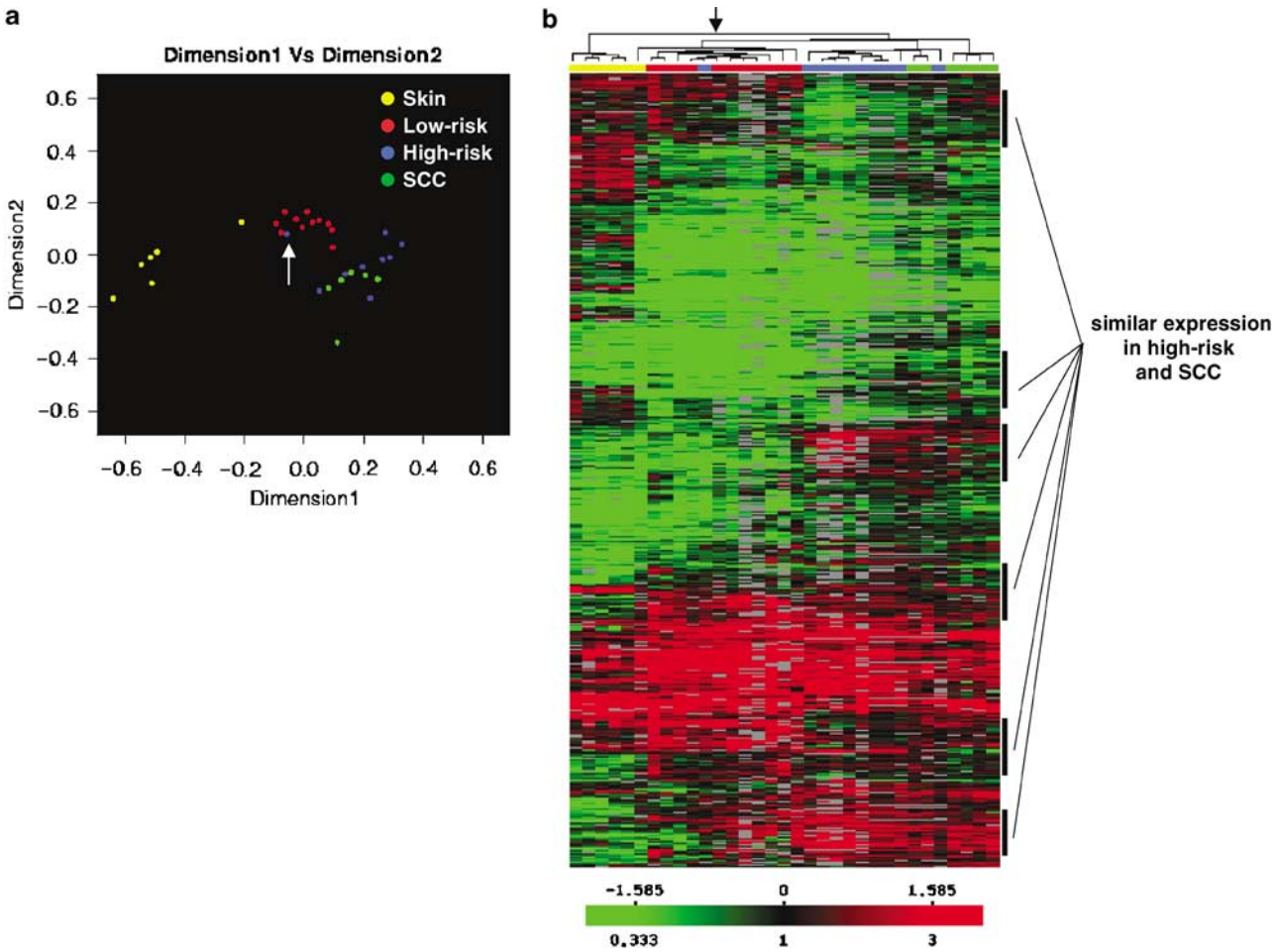



**Figure 2** Clustering of differentially expressed genes links high-risk papillomas to squamous cell carcinomas. (**a**) Clustering of arrays using multidimensional scaling after ANOVA filtering of $P<0.001$. (**b**) Heat map and dendrogram of unsupervised hierarchical clustering (Pearson correlation) of arrays with the differentially expressed gene set. Lines represent clusters of genes whose expression pattern is similar between low-risk papillomas and squamous cell carcinoma (SCC) and distinct from other tissues. Normal skin ($n=6$), low-risk papillomas ($n=10$), high-risk papillomas ($n=10$) and SCC ($n=6$). In both methods, high-risk papillomas and SCC arrays show a high degree of similarity except for tumor 1B-29B, a high-risk tumor that clusters with the low-risk papilloma group (arrow). For both the multidimensional scaling (MDS) and hierarchical clustering, tissues are coded as yellow, normal skin; red, low-risk papilloma; blue, low-risk papilloma; green SCC.

low-risk papillomas (6/6) have focal areas with high CD3ε+ infiltration in the tumor epithelia and abundant CD3ε+ cells in the tumor stroma, while within the high-risk papillomas tumor cell compartment and tumor stroma there are few CD3ε+ cells in the high-risk tumor or tumor stroma. Thus, these results validate the microarray data showing reduced levels of TCR expression in the high-risk papillomas. Taken together these results suggest that although tumor development is associated with elevated expression of proinflammatory genes, high-risk benign lesions and SCC are associated with reduced T-cell-associated anti-tumor adaptive immunity.

*A distinct profile for chronic tumor promotion linked to high-risk papillomas*

To rule out that the difference in gene expression profiles between low- and high-risk papillomas and SCC was due to persistent application of a tumor promoter to the low-risk papillomas, we examined expression of the 514 differentially expressed genes obtained from the initial analysis in skin treated chronically with TPA for 5 weeks. Similar to published studies of short-term TPA treatment (Schlingemann *et al*., 2003; Hummerich *et al*., 2006), a number of genes were induced or down-regulated in both TPA-treated skin and all tumors such as S100A8, several Sprr2 family members, PSG28, Esd, dynactin (dctn1) and myosin 1a Myo1A (Supplementary Table I). Many genes were distinctly regulated between tumor-promoted normal skin and all tumors including arginase and Stefin A1, which were induced 3–20-fold in tumors but were either unaffected or downregulated in chronically tumor-promoted skin. MDS analysis of arrays from all tissues showed that the overall gene expression pattern of chronically phorbol ester-promoted skin was distinct from both normal skin and from all tumors (data not shown). These results demonstrate

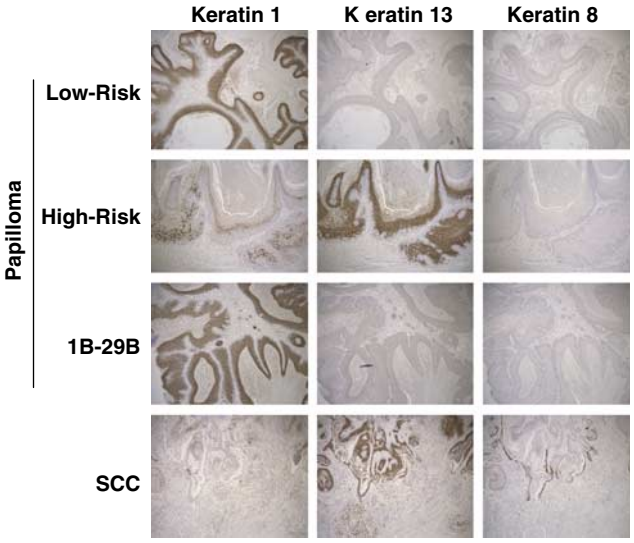



**Figure 3** Keratin immunophenotyping of low- and high-risk tumors correlates with expression clustering. Ethanol-fixed tissue sections from all low- and high-risk papillomas and SCC were immunostained for keratins 1, 13 and 8 as described in Materials and methods. The pattern shown for the low-risk and high-risk papillomas is representative of 10/10 high-risk papillomas and 9/10 high-risk tumors. The tumor 1B-29B was generated with the high-risk tumor promotion protocol, but clustered with the low-risk group based on expression analysis and histochemical phenotyping.

that the difference in gene expression between the low-risk and high-risk papillomas is not due to a difference in TPA treatment. Surprisingly, as shown in Table 1, 64% (51/79) of the genes that distinguish low- and high-risk papillomas/SCC follow the high-risk/SCC pattern in chronic TPA-treated normal skin, while only 9/79 genes were similar between low-risk papillomas and chronic TPA treatment. Thus, chronic TPA promotion of normal skin causes changes in gene expression distinct from the majority of chronically promoted benign tumors, but produces an expression signature similar to high-risk papillomas and SCC.

*Generation of a risk predictor for benign lesions*

To test the robustness of the benign tumor classifier we utilized a different set of class prediction tools in the BRBArray Tools software program (NCI) to compare low and high-risk papillomas, and then used this classifier to predict the risk of independent coded unknown tumors generated using the two-stage carcinogenesis protocol. The sensitivity and specificity (Supplementary methods) of the classifiers was between 0.857 and 1 depending on the prediction method. Three low-risk tumors and three high-risk tumors were designated as test unknowns and the remainder used as a training set for class prediction. Genes were included at a univariate parametric significance level of $P<0.0001$. Except for one tumor that was misclassified by the support vector machines algorithm, all six classifiers were in agreement and correctly classified the test tumors based on keratin profiling, promotion protocols and previous hierarchical clustering. This list of 125 genes contained close to 50% of the genes present on the previously generated PAM gene list (Supplementary Table III), and included many of the immune function genes such as Lgals2, Hamp1, IgK, Laf4, Traf2, and TCR$\beta$ as well as CD2. We then used this $P<0.0001$ predictor gene set to classify eight coded not previously classified tumor samples with two previously classified as either low- or high-risk. Hierarchical clustering of these unknown tumors placed three high-risk papillomas (as determined by tumor induction protocol) in one cluster with a previously identified high-risk papilloma (1B-50C) and an unknown that was determined histologically to be a SCC, and four unclassified low-risk papillomas with a previously classified low-risk papilloma (2-53-D) (Figure 6). To correlate the molecular risk prediction with known biomarkers, we stained the tumors with different keratin antibodies. There was a strong relationship between the keratin immunostaining pattern and the predicted risk group based on the array clustering. These results show that this methodology can predict the risk potential of a benign papilloma with high accuracy.

## Discussion

Microarray technology has been used to identify gene expression patterns associated with different stages of cancer, different cancer types and to generate predictors of therapeutic response (Glinsky *et al*., 2004; Lossos *et al*., 2004; Nambiar *et al*., 2004; Juric *et al*., 2005). In the mouse skin carcinogenesis model microarrays have identified genes whose expression is associated with tumor promotion, with benign and malignant stages of tumor progression or with malignant progression and metastasis (Dong *et al*., 1997; Serewko *et al*., 2002; Dooley *et al*., 2003; Hummerich *et al*., 2006; Ridd *et al*., 2006). Similar to many human epithelial cancers that arise from premalignant lesions such as superficial bladder dysplasia, oral leukoplakia or actinic keratoses with unknown risk for malignant conversion (Callen *et al*., 1997), benign tumors generated in the mouse multistage skin carcinogenesis model differ in malignant potential and express unique phenotypic markers (Hennings *et al*., 1985; Glick *et al*., 1993; Tennenbaum *et al*., 1993; Darwiche *et al*., 1996). In this report, we have used microarray analysis to test the hypothesis that high-risk papillomas share a gene expression pattern with and thereby represent the likely precursor lesion to SCC. While there were patterns of gene expression that were common to all tumors, two clustering methods using a database of 514 differentially expressed genes closely linked the gene expression profile of high-risk papillomas and SCC. These results demonstrate that within a large cohort of histologically and biologically benign tumors subpopulations of tumors exist with distinct gene expression patterns that relate to their malignant potential. In addition, this expression clustering

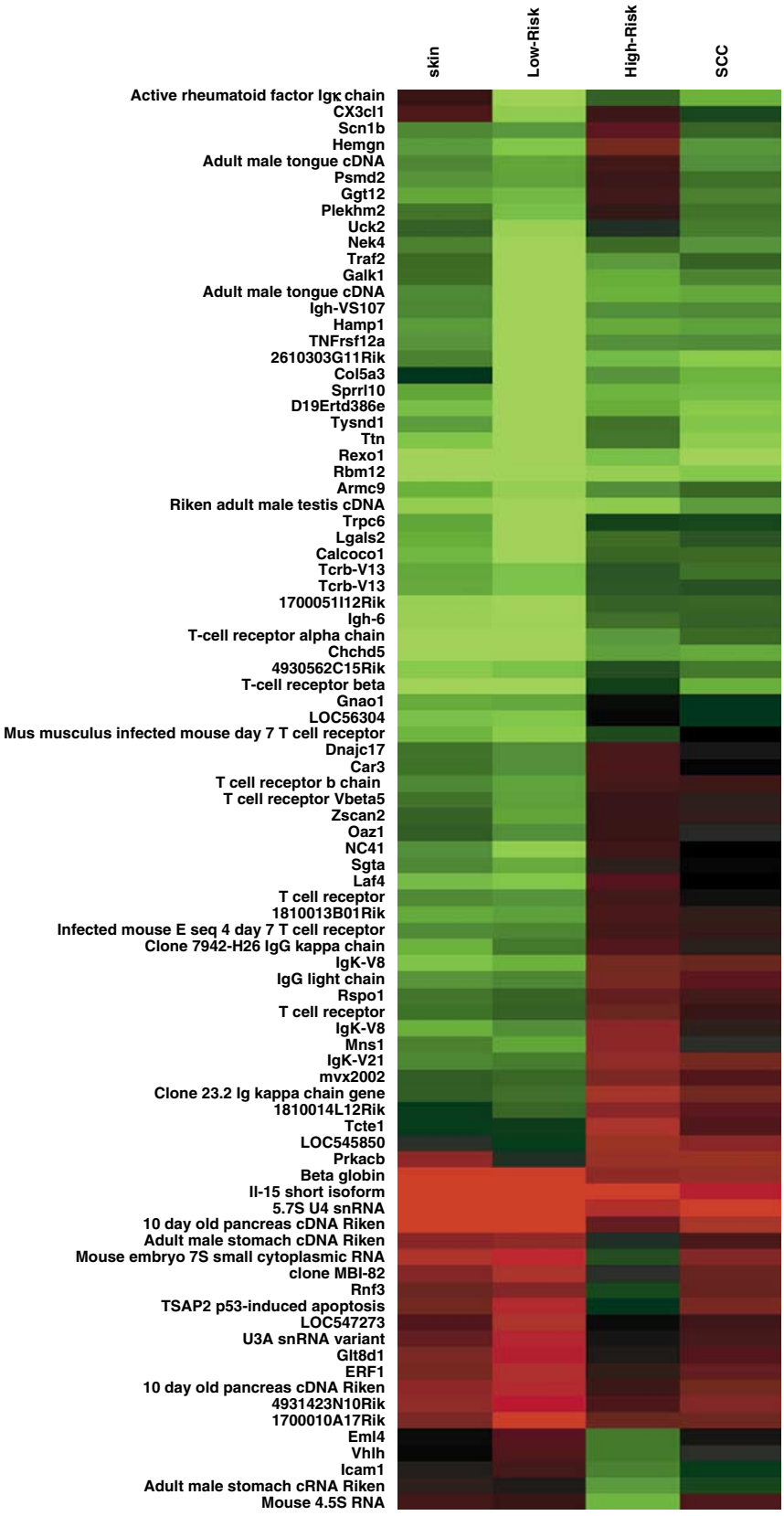



**Figure 4** An expression signature predicts risk group of unknown squamous tumors. Genes that distinguish low- and high-risk papillomas show similar pattern of expression in normal skin and squamous cell carcinoma (SCC), respectively. Arrays from normal skin, benign tumors and SCC were filtered using an 87 gene list generated using prediction analysis for microarray (PAM) comparison between low-risk and high-risk papillomas with a misclassification rate of 0.048. Filtered expression ratios were averaged within each group to generate heat map.

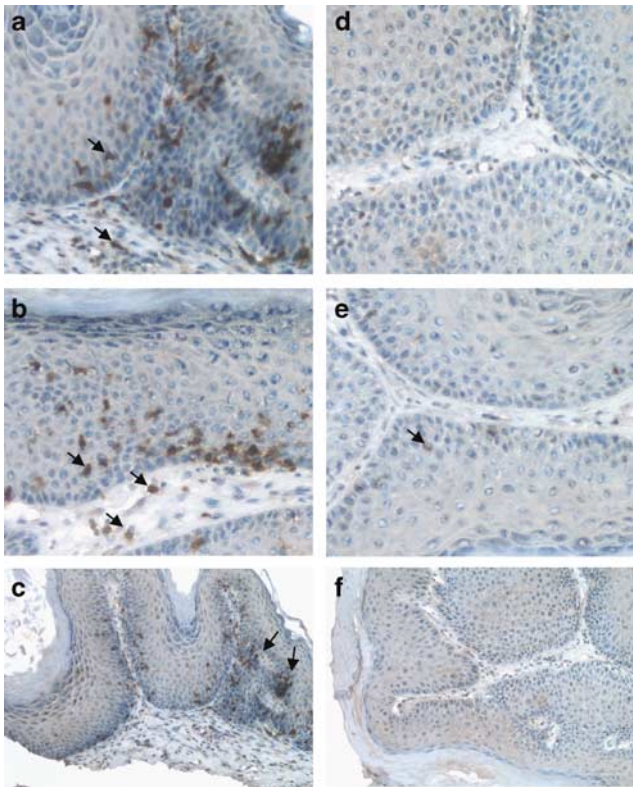


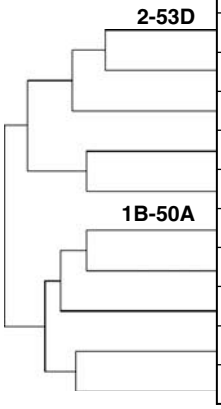

**Figure 5** High-risk papillomas exhibit reduced T-cell infiltration. Tissue sections from ethanol-fixed low (**a**–**c**) and high-risk papillomas (**d**–**f**) were stained for T cells using an anti-CD3ε antibody. Photomicrographs (**a**, **b**, **d**, **e**) are high-magnification ( × 400) images of two low-risk papillomas showing focal regions of numerous intraepithelial and stromal CD3ε+ cells, and the absence of T cells in two representative high-risk papillomas. (**c** and **f**) ( × 100 magnification) show global absence of $CD3^+$ cells in high-risk papillomas and multiple focal regions of intraepithelial $CD3^+$ cells in high-risk papillomas. Six out of six tumors of each class exhibited the staining pattern of the tumors shown. Arrows denote CD3ε-positive cells.

correctly distinguished a tumor generated by the low-risk promotion protocol that had a phenotypic profile characteristic of the low-risk papillomas. Of the 514 differentially regulated genes, we identified a set of 87 that represented significant classifiers for either the low- or high-risk papilloma state. When these arrays were divided into training and unknown sets this classifier list correctly predicted the risk group of the six 'unknown' tumors. Similarly, using a second classification method we correctly grouped 9/10 unknown tumors into low or high-risk groups. While a larger unknown tumor set is needed to test the predictive nature in this model, and we have not determined the minimum classifier needed, these results suggest that this methodology can be used to distinguish benign lesions at differing risks for progression.

A large fraction of the genes whose expression changes during tumor progression is associated with inflammation and adaptive immunity. Based on human studies and some experimental models that show a link between inflammation and tumor development (Coussens *et al*., 1999; Balkwill and Mantovani, 2001; Coussens and Werb, 2002; de Visser *et al*., 2005), it is widely believed that chronic inflammation acts to promote cancer development. Indeed in both this report and a previous microarray study (Hummerich *et al*., 2006) many proinflammatory genes are upregulated during squamous tumor development with the highest expression in benign low-risk papillomas. However, in contrast to a model of HPV-E6/E7 driven cutaneous tumor formation where inflammation and antibody deposition underlie the pathogenesis of the disease (Coussens *et al*., 1999; de Visser *et al*., 2005), our data show that high-risk papillomas and SCC are distinguished by reduced expression of genes associated with T cells, B cells and adaptive immunity. These results support the concept that immunosurveillance mechanisms exist that function to recognize and eliminate developing tumor cells, and that defects in immunosurveillance are associated with tumor progression (Dighe *et al*., 1994; Kaplan *et al*., 1998; Smyth *et al*., 2000; Girardi *et al*., 2001, 2003). Our results suggest that premalignant tumors with high expression of inflammation-associated genes and T- and B-cell-specific products remain terminally benign, and that the rapid appearance of high-risk lesions after onset of promotion may be due to reduced anti-tumor immunity acting to slow outgrowth of initiated keratinocytes.

It was expected, but not observed, that low-risk papillomas would have a largely overlapping gene expression profile with chronically TPA-treated normal skin due to the repeated application of TPA to these tumors. While there was overlap between chronic TPA and all tumors, there were very few genes whose expression pattern was uniquely common to low-risk papillomas and TPA-treated skin. Instead, chronic TPA promotion produced an expression pattern similar to that of high-risk papillomas and SCC within the context of genes that distinguished these tumors from low-risk papillomas. The expression pattern of genes that distinguish tumors that ultimately progress to malignancy can be detected in as little as 5 weeks of promotion of normal uninitiated skin. Thus, even 8 weeks after last TPA treatment high-risk papillomas retain an expression pattern similar to chronically promoted skin, suggesting that these represent autonomous or self-promoting lesions. That the high-risk expression signature is evident after 5 weeks of promotion in normal skin suggests that the keratinocyte progenitor of the high-risk papilloma and the keratinocytes that persist in the epidermis after long-term promotion are closely linked. It remains to be determined, however, if the precursor to high-risk papillomas originates in follicular stem cells, while low-risk papillomas originate from interfollicular keratinocytes. Recent studies showing that SCC arise from hair follicle stem cells, while the majority of benign papillomas arise from interfollicular keratinocytes support this concept (Morris *et al*., 1997, 2000; Brown *et al*., 1998; Perez-Losada and Balmain, 2003).

Validation of these gene lists as predictive markers for human cancers remains to be determined. Using the online Oncomine human cancer microarray database




| unknown | Predicted Risk: Array | Promotion Protocol | Histology | Keratin Profiling | Predicted Risk: Phenotpying |
|---|---|---|---|---|---|
| 1 | Low | 2 | benign/pap | K1+++ K13- | Low |
| 5 | Low | 2 | benign/pap | K1+++ K13- | Low |
| 9 | Low | 2 | benign/pap | K1+++ K13+ | Low |
| 10 | Low | 2 | benign/pap | nd | Low |
| 6 | Low | 2 | benign/pap | K1+++ K13++ | Low |
| 2 | High | 1 | benign/pap | K1+ K13+++ | High |
| 4 | High | 1 | benign/pap | K1++ K13++ | High |
| 3 | High | 1 | SCC | K1-K13++K8+ | High |
| 7 | High | 1 | benign/pap | nd | High |
| 8 | High | 1 | benign/pap | K1+ K13+++ | High |

**Figure 6** A classifier that predicts risk for unknown benign papillomas. Arrays performed on known and unknown skin tumors were filtered using a 125 gene list obtained using a covariate classifier algorithm in BRBArray tools (Supplementary methods) and then clustered hierarchically using centered correlation and average linkage. The associated table for each tumor on the dendrogram shows the tumor promotion protocol, histology, keratin staining profile and risk group as determined by expression of genes within the classifier, and phenotypic analysis (histology and keratin immunostaining). Unknowns 1 and 2 were previously analysed as low- and high-risk papillomas 2-53-D and 1B-50A, respectively, which were used as internal controls for clustering.

(http://www.oncomine.org/main/index.jsp; Rhodes *et al*., 2004) which allows crossplatform comparison of cancer array data from multiple published studies, we find that 22 of the PAM list genes were significantly changed in several human cancers relative to normal. In all lung cancers arrays, 9 of the 13 total identified changed genes followed a similar pattern as the low-risk papillomas in this study. CXC3CL1 (fractalkine) was decreased in multiple lung cancer types as well as liver, breast and prostate cancer relative to normal suggesting that decreased expression of this chemokine may be of predictive value in some human cancers. Other genes in the high-risk signature that are expressed in a similar pattern in human cancers include Laf4, decreased in lung SCC and adenocarcinoma, RNF3 increased in prostate carcinoma, intercellular adhesion molecule 1 (ICAM1) increased in HNSCC, and VHL increased in lung adenocarcinoma and metastatic prostate carcinoma. While these specific genes may be only directly relevant to the two-stage skin carcinogenesis model, these results provide strong evidence for heterogeneity in global gene expression patterns of premalignant lesions that is linked to phenotypic heterogeneity and risk for malignant conversion. This suggests that expression analysis of human premalignant lesions could provide diagnostic biomarkers for cancer risk as well as novel therapeutic targets for high-risk lesions.

## Materials and methods

*Tumor generation and characterization*
Tumors were induced in 7–8-week-old outbred female SENCAR mice using a one time treatment with DMBA (5 $\mu$g) and TPA (2 $\mu$g) once per week for variable times as described (Hennings *et al*., 1985). Low-risk papillomas were generated by continuous TPA promotion for 14–15 weeks and were harvested 3 days after the last TPA treatment. High-risk papillomas were induced by 7 weeks of TPA promotion and harvested at 14–15 weeks. SCCs generated in either tumor induction protocol were harvested as they appeared. Normal skin was isolated from 7–8-week-old SENCAR mice in the telogen stage of hair cycle. Hyperplastic non-initiated skin was generated by 5 weeks of TPA treatment (2 $\mu$g) once per week and was harvested 3 days after the last TPA treatment. Histology of tumors and staging were characterized by hematoxylin and eosin staining and through anti-keratin immunostaining.

*Microarray design and analysis and validation*
See Supplementary Information for detailed description of microarray methodology, analysis and validation by semi-quantitative and quantitative reverse transcription–PCR. All original microarray data has been deposited to the GEO website http://www.ncbi.nlm.nih.gov/geo/, accession number GSE5576.

*Immunohistochemistry*
Keratins were detected in 70% ethanol-fixed tissue sections using monospecific rabbit antibodies to keratins K1, K10, K13 (Covance Research Products, Berkeley, CA, USA) or a monoclonal anti-K8 (Developmental Studies Hybridoma Bank, University of Iowa, IA, USA), and T cells were detected in 70% ethanol fixed tissue sections using an anti-CD3$\varepsilon$ antibody (Santa Cruz Biotechnology, Santa Cruz, CA, USA). Bound antibodies were detected using appropriate biotinylated secondary antibodies and the Vectastain Elite Kit (Vector Laboratories, Burlingame, CA, USA) as described previously (Glick *et al*., 1993).

## Abbreviations

DMBA, dimethyl benz[*a*]-anthracene; FDR, false discovery rate; K, keratin; MDS, multidimensional scaling; PAM, prediction analysis for microarray; SAM, significance analysis for microarrays; SCC, squamous cell carcinomas; TPA,

12-*O*-tetradecanoylphorbol-13-acetate; TCR, T cell receptor; TPR, tetratrico peptide repeat; ETS, E26 transformation specific.

**Acknowledgements**

ND was supported by the American University of Beirut in Beirut/Lebanon and by the National Cancer Institute Scientist Exchange Program Fellowship. RP-L and AG were supported by the Penn State Huck Institutes of Life Sciences the Penn State Institute of the Environment, and NIH Grant CA117957 to AG This research was also supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

**References**

Balkwill F, Mantovani A. (2001). Inflammation and cancer: back to Virchow? *Lancet* **357**: 539–545.

Brown K, Strathdee D, Bryson S, Lambie W, Balmain A. (1998). The malignant capacity of skin tumours induced by expression of a mutant H-ras transgene depends on the cell type targeted. *Curr Biol* **8**: 516–524.

Callen JP, Bickers DR, Moy RL. (1997). Actinic keratoses. *J Am Acad Dermatol* **36**: 650–653.

Coussens LM, Raymond WW, Bergers G, Laig-Webster M, Behrendtsen O, Werb Z *et al*. (1999). Inflammatory mast cells up-regulate angiogenesis during squamous epithelial carcinogenesis. *Genes Dev* **13**: 1382–1397.

Coussens LM, Werb Z. (2002). Inflammation and cancer. *Nature* **420**: 860–867.

Darwiche N, Scita G, Jones C, Rutberg S, Greenwald E, Tennenbaum T *et al*. (1996). Loss of retinoic acid receptors in mouse skin and skin tumors is associated with activation of the *ras*$^{Ha}$ oncogene and high risk for premalignant progression. *Cancer Res* **56**: 4942–4959.

de Visser KE, Korets LV, Coussens LM. (2005). De novo carcinogenesis promoted by chronic inflammation is B lymphocyte dependent. *Cancer Cell* **7**: 411–423.

Dighe AS, Richards E, Old LJ, Schreiber RD. (1994). Enhanced *in vivo* growth and resistance to rejection of tumor cells expressing dominant negative IFN gamma receptors. *Immunity* **1**: 447–456.

Dong G, Loukinova E, Smith CW, Chen Z, Van Waes C. (1997). Genes differentially expressed with malignant transformation and metastatic tumor progression of murine squamous cell carcinoma. *J Cell Biochem Suppl* **28–29**: 90–100.

Dooley TP, Reddy SP, Wilborn TW, Davis RL. (2003). Biomarkers of human cutaneous squamous cell carcinoma from tissues and cell lines identified by DNA microarrays and qRT-PCR. *Biochem Biophys Res Commun* **306**: 1026–1036.

Dumitriu IE, Baruah P, Manfredi AA, Bianchi ME, Rovere-Querini P. (2005). HMGB1: guiding immunity from within. *Trends Immunol* **26**: 381–387.

Girardi M, Glusac E, Filler RB, Roberts SJ, Propperova I, Lewis J *et al*. (2003). The distinct contributions of murine T cell receptor (TCR)gammadelta + and TCRalphabeta + T cells to different stages of chemically induced skin cancer. *J Exp Med* **198**: 747–755.

Girardi M, Oppenheim DE, Steele CR, Lewis JM, Glusac E, Filler R *et al*. (2001). Regulation of cutaneous malignancy by gammadelta T cells. *Science* **294**: 605–609.

Glick AB, Kulkarni AB, Tennenbaum T, Hennings H, Flanders KC, O'Reilly M *et al*. (1993). Loss of expression of transforming growth factor $\beta$ in skin and skin tumors is associated with hyperproliferation and a high risk for malignant conversion. *Proc Natl Acad Sci USA* **90**: 6076–6080.

Glinsky GV, Glinskii AB, Stephenson AJ, Hoffman RM, Gerald WL. (2004). Gene expression profiling predicts clinical outcome of prostate cancer. *J Clin Invest* **113**: 913–923.

Hennings H, Shores R, Mitchell P, Spangler EF, Yuspa SH. (1985). Induction of papillomas with a high probability of conversion to malignancy. *Carcinogenesis* **6**: 1607–1610.

Hummerich L, Muller R, Hess J, Kokocinski F, Hahn M, Furstenberger G *et al*. (2006). Identification of novel tumour-associated genes differentially expressed in the process of squamous cell cancer development. *Oncogene* **25**: 111–121.

Juric D, Sale S, Hromas RA, Yu R, Wang Y, Duran GE *et al*. (2005). Gene expression profiling differentiates germ cell tumors from other cancers and defines subtype-specific signatures. *Proc Natl Acad Sci USA* **102**: 17763–17768.

Kaplan DH, Shankaran V, Dighe AS, Stockert E, Aguet M, Old LJ *et al*. (1998). Demonstration of an interferon gamma-dependent tumor surveillance system in immunocompetent mice. *Proc Natl Acad Sci USA* **95**: 7556–7561.

Larcher F, Bauluz C, Diaz-Guerra M, Quintanilla M, Conti CJ, Ballestin C *et al*. (1992). Aberrant expression of the simple epithelial type II keratin 8 by mouse skin carcinomas but not papillomas. *Mol Carcinog* **6**: 112–121.

Lossos IS, Czerwinski DK, Alizadeh AA, Wechser MA, Tibshirani R, Botstein D *et al*. (2004). Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. *N Engl J Med* **350**: 1828–1837.

Ma C, Staudt LM. (1996). LAF-4 encodes a lymphoid nuclear protein with transactivation potential that is homologous to AF-4, the gene fused to MLL in t(4;11) leukemias. *Blood* **87**: 734–745.

Morris RJ, Coulter K, Tryson K, Steinberg SR. (1997). Evidence that cutaneous carcinogen-initiated epithelial cells from mice are quiescent rather than actively cycling. *Cancer Res* **57**: 3436–3443.

Morris RJ, Tryson KA, Wu KQ. (2000). Evidence that the epidermal targets of carcinogen action are found in the interfollicular epidermis of infundibulum as well as in the hair follicles. *Cancer Res* **60**: 226–229.

Nambiar PR, Nakanishi M, Gupta R, Cheung E, Firouzi A, Ma XJ *et al*. (2004). Genetic signatures of high- and low-risk aberrant crypt foci in a mouse model of sporadic colon cancer. *Cancer Res* **64**: 6394–6401.

Nishimura H, Washizu J, Nakamura N, Enomoto A, Yoshikai Y. (1998). Translational efficiency is up-regulated by alternative exon in murine IL-15 mRNA. *J Immunol* **160**: 936–942.

Nishimura H, Yajima T, Naiki Y, Tsunobuchi H, Umemura M, Itano K *et al*. (2000). Differential roles of interleukin 15 mRNA isoforms generated by alternative splicing in immune responses *in vivo*. *J Exp Med* **191**: 157–170.

Ozaki K, Inoue K, Sato H, Iida A, Ohnishi Y, Sekine A *et al*. (2004). Functional variation in LGALS2 confers risk of myocardial infarction and regulates lymphotoxin-alpha secretion *in vitro*. *Nature* **429**: 72–75.

Perez-Losada J, Balmain A. (2003). Stem-cell hierarchy in skin cancer. *Nat Rev Cancer* **3**: 434–443.

Peyssonnaux C, Zinkernagel AS, Datta V, Lauth X, Johnson RS, Nizet V. (2006). TLR-4 dependent hepcidin expression

by myeloid cells in response to bacterial pathogens. *Blood* **107**: 3727–3732.

Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D *et al*. (2004). ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* **6**: 1–6.

Ridd K, Zhang SD, Edwards RE, Davies R, Greaves P, Wolfreys A *et al*. (2006). Association of gene expression with sequential proliferation, differentiation and tumor formation in murine skin. *Carcinogenesis* **27**: 1556–1566.

Schlingemann J, Hess J, Wrobel G, Breitenbach U, Gebhardt C, Steinlein P *et al*. (2003). Profile of gene expression induced by the tumour promotor TPA in murine epithelial cells. *Int J Cancer* **104**: 699–708.

Serewko MM, Popa C, Dahler AL, Smith L, Strutton GM, Coman W *et al*. (2002). Alterations in gene expression and activity during squamous cell carcinoma development. *Cancer Res* **62**: 3759–3765.

Smyth MJ, Thia KY, Street SE, MacGregor D, Godfrey DI, Trapani JA. (2000). Perforin-mediated cytotoxicity is critical for surveillance of spontaneous lymphoma. *J Exp Med* **192**: 755–760.

Tennenbaum T, Weiner AK, Belanger AJ, Glick AB, Hennings H, Yuspa SH. (1993). The suprabasal expression of $\alpha 6\beta 4$ integrin is associated with a high risk for malignant progression in mouse skin carcinogenesis. *Cancer Res* **53**: 4803–4810.

Tibshirani R, Hastie T, Narasimhan B, Chu G. (2002). Diagnosis of multiple cancer types by shrunken centroids of gene expression 2. *Proc Natl Acad Sci USA* **99**: 6567–6572.

Yuspa SH. (1998). The pathogenesis of squamous cell cancer: lessons learned from studies of skin carcinogenesis. *J Dermatol Sci* **17**: 1–7.

Supplementary Information accompanies the paper on the Oncogene website (http://www.nature.com/onc).