[CANCER RESEARCH 64, 3661–3667, May 15, 2004]

Microarray Analysis Reveals Differences in Gene Expression of Circulating CD8+ T Cells in Melanoma Patients and Healthy Donors

Tong Xu,[1] Chen-Tsen Shu,[1] Elizabeth Purdom,[2] Demi Dang,[3] Diane Ilsley,[3] Yaqian Guo,[2] Jeffrey Weber,[4] Susan P. Holmes,[2] and Peter P. Lee[1]

[1]Division of Hematology, Stanford University School of Medicine, Stanford; [2]Department of Statistics, Stanford University, Stanford; [3]Agilent Technologies, Palo Alto; and [4]University of Southern California/Norris Cancer Center, Los Angeles, California

ABSTRACT

Circulating T cells from many cancer patients are known to be dysfunctional and undergo spontaneous apoptosis. We used microarray technology to determine whether gene expression differences exist in T cells from melanoma patients *versus* healthy subjects, which may underlie these abnormalities. To maximize the resolution of our data, we sort purified CD8+ subsets and amplified the extracted RNA for microarray analysis. These analyses show subtle but statistically significant expression differences for 10 genes in T cells from melanoma patients *versus* healthy controls, which were additionally confirmed by quantitative real-time PCR analysis. Whereas none of these genes are members of the classical apoptosis pathways, several may be linked to apoptosis. To additionally investigate the significance of these 10 genes, we combined them into a classifier and found that they provide a much better discrimination between melanoma and healthy T cells as compared with a classifier built uniquely with classical apoptosis-related genes. These results suggest the possible engagement of an alternative apoptosis pathway in circulating T cells from cancer patients.

INTRODUCTION

Cancer impacts the immune response of the host in a number of ways. Dysfunction or anergy of tumor-infiltrating lymphocytes and circulating tumor-specific T cells may be early events in tumor progression (1, 2), whereas global immune dysfunction develops in many patients with advanced tumor burden (3). Defects include alterations in T-cell receptor signaling events, reduced proliferation, and increased apoptosis. In patients with melanoma, breast, oral, and head and neck cancer, spontaneous apoptosis occurs in a high percentage of peripheral CD8+ T cells and natural killer cells (4, 5). It remains unclear what drives the dysfunction or apoptosis of lymphocytes in the cancer setting. To address the underlying molecular defects within T cells in cancer, we used cDNA microarray technology to study the gene expression of CD8+ T-cell subsets (naïve, memory, and effector) in melanoma patients and compare to T cells from healthy controls.

DNA microarray technology provides a powerful tool to study differences in transcript abundance in parallel (6). Typical microarray labeling procedures require 3–6 μg polyadenylated RNA or 20–100 μg total RNA per cDNA microarray. This amount of RNA can only be reliably obtained from cell lines or tissue samples, which contain heterogeneous cell types. Because it is difficult to obtain sufficient material for microarray analysis from pure primary cells, RNA amplification methods have been developed. Among the currently available amplification methods, the most commonly used is T7 promoter-based linear amplification first developed by Eberwine (7). In this report, we used a modified T7-based amplification protocol (developed by Agilent Technologies Inc.). Systematic evaluation of the reproducibility of this amplification protocol demonstrated strong correlations of >0.9, thus ensuring a high quality of amplification. This protocol enabled us to study small sort-purified populations of CD8+ T-cell subsets. We found subtle gene expression differences between T cells from melanoma patients *versus* those from healthy controls. These subtle differences were revealed using a new data transformation that stabilizes the variance across the whole set of expression levels. These results were confirmed via quantitative real-time PCR (RQ-PCR) analysis of the original, unamplified RNA samples. This represents an important step toward the elucidation of the mechanism of immune dysfunction in cancer.

MATERIALS AND METHODS

Patient Samples. Peripheral blood mononuclear cells (PBMC) were isolated from healthy donors (Stanford Blood Center, Stanford, CA) or patients with American Joint Committee on Cancer stage III-IV melanoma before any systemic therapy (Norris Cancer Center, Los Angeles, CA) following informed consent. Cells were washed and cryopreserved in 90% FCS and 10% DMSO.

Flow Sorting and Cell Treatment. PBMCs were thawed and cultured (RPMI 1640 supplemented with 10% fetal bovine serum, penicillin, streptomycin, and glutamine) overnight at 37°C, 5% CO_2. CD8+ T cells were enriched using RosetteSep (StemCell Technologies, Vancouver, British Columbia, Canada). The resulting cells were incubated with the antibodies CD8-phycoerythrin (Caltag, San Francisco, CA), CD45RA-Cy5PE (Phar-Mingen, San Diego, CA), and CD27-FITC (Caltag) at room temperature for 30 min. After staining, cells were washed and sorted using a FACSVantage (Becton Dickinson, San Jose, CA). CD8+CD45RA+CD27+, CD8+CD45RA+CD27−, and CD8+CD45RA-CD27+ cells were collected separately using gates shown in Fig. 1. Cell purity was confirmed after the sorting by analyzing each sorted fraction using a FACSCalibur (Becton Dickinson). Cells (100,000) from each fraction were homogenized into 1 ml of TRIzol (Invitrogen) and frozen at −80°C, with the addition of 10 μg of linear acrylamide (Ambion).

RNA Isolation and Sample Preparation. Total cellular RNA was isolated from the TRIzol cell extracts following the manufacturer's protocol. The RNA was then resuspended in 15 μl of 10 mM Tris-HCl (pH 8.0) and 1 mM EDTA. Approximately half of the purified sample RNA was used in the microarray analysis, and the remaining fraction was frozen at −80°C to be used for additional analysis. One μl was diluted 1:5 in nuclease-free water, and the diluted RNA was used for analysis of RNA integrity on the RNA 6000 Pico Labchip kit (Agilent Technologies, Palo Alto, CA). To preserve the RNA sample, we did not use UV spectrophotometry to quantitate the amount of RNA isolated from each sorted T-cell sample. Rather, the RNA pellets for each sample were resuspended in identical volumes, and the same volume for each sample (approximately half of the total volume) was used in the amplification reactions.

RNA Amplification, Labeling, and Hybridization. To minimize potential amplification bias, we elected to use only one round of amplification in this study. The typical yield of amplified RNA (aRNA) using the Agilent amplification protocol is between 400 ng and 10 μg. The polyadenylated RNA was amplified using the Agilent Low Input Linear Amplification kit following the protocol described in the users manual for RNA Amplification and Fluorescent cDNA Synthesis. Briefly, total RNA (7.8 μl) was mixed with 1.2 μL T7 promotor primer. The primer and template were annealed by incubating the reaction at 70ºC for 10 min and then placing the reaction on ice. The cDNA synthesis was performed by adding 1× First Strand Buffer, DTT, RNaseOut, deoxynucleoside triphosphates, and Moloney murine leukemia virus reverse

Received 10/29/03; revised 2/11/04; accepted 3/5/04.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: Peter P. Lee, Stanford University, Center for Clincal Sciences Research, Room 1155, 269 Campus Drive, Stanford, CA 94305. Phone: (650) 498-7942; E-mail: ppl@stanford.edu.

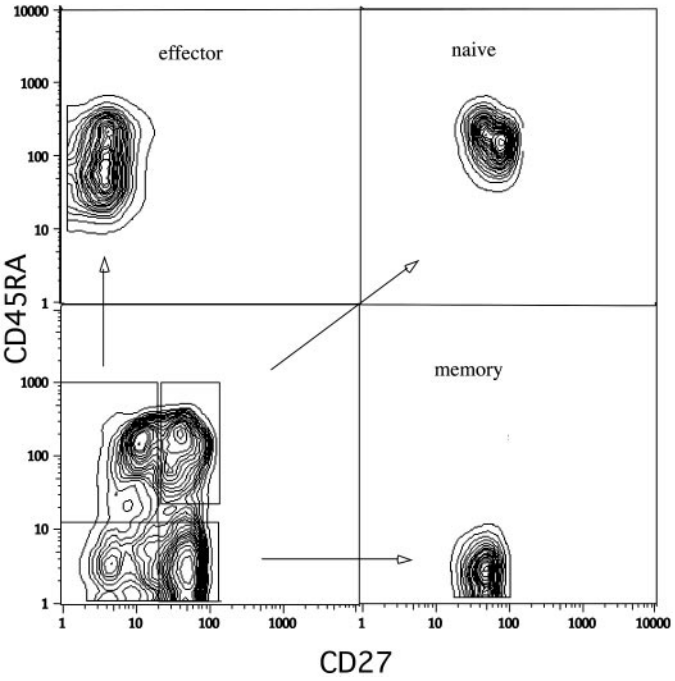

Fig. 1 Fluorescence-activated cell sorter plot of sample before and after sorting naïve, effector, and memory T-cell subsets were differentiated based on expression of CD27 and CD45RA. These subsets were sorted individually based on the indicated sort gates.

transcriptase. The reaction was incubated at 40°C for 2 h. The enzyme was then inactivated by heating the reaction to 65°C for 15 min and transferring the reaction to ice. *In vitro* transcription was initiated by the addition of 60 μl of Transcription Master Mix containing 1× Transcription Buffer, DTT, nucleotide triphosphates, polyethylene glycol, RNaseOut, Inorganic Pyrophosphatase, and T7 RNA polymerase to the cDNA synthesis reaction (total volume 80 μl). The reaction was incubated at 40°C for 4 h. The aRNA was purified using Qiagen RNeasy mini spin columns, as described in the Agilent user manual. The eluted aRNA was dried under vacuum in a rotary dessicator. The pellet was resuspended in 15 μl of RNase-free water. The aRNA concentration was determined using the Nanodrop ND-1000, and aRNA was visually qualified using the Agilent 2100 BioAnalyzer (RNA 6000 Nano Labchip kit). The reference sample was Stratagene human universal reference and was amplified at the same time as the sample RNA using the same method. Individual reference aRNA reactions were pooled before labeling. After amplification, each aRNA sample was checked using the RNA Nano Labchip before being converted to fluorescently labeled cDNA (data not shown). The patient samples were labeled using Cy5-dCTP and the reference sample was labeled using Cy3-dCTP. After RNase A digestion, the Cy3- and Cy5-labeled cDNA targets were pooled and purified using Qiaquick PCR purification kit (Qiagen) following the Agilent recommended protocol. The eluted targets were dried under vacuum in a rotary dessicator and resuspended in 7.5 μl of water. Cot 1 DNA, control targets, and hybridization buffer were added to the target solution. After incubating at 98°C for 2 min, the target was cooled. Fluorescently labeled cDNA targets were hybridized onto Agilent Human 1 cDNA microarrays following the Agilent cDNA Microarray user manual.

Microarray Imaging and Data Analysis. The microarrays were washed and dried following the user manual instructions and scanned on the Agilent dual laser DNA microarray array scanner. Data were corrected with regards to local background as implemented by the Agilent feature extraction software. Microarrays that were hybridized on the same day were labeled as belonging to the same batch. The median red and green background and foreground variables were transferred into the R software package (8). The data were first transformed batch by batch using a variance stabilizing procedure as described in the next section. Remaining batch effects were eliminated through subtraction of the median from each batch, thus aligning all of the batch medians to a common value. In some analyses, the cell type effect (naïve, effector, and memory) was also removed through an additive model to make the averages of all of the genes for each of the cell types zero in the residual data. Features that show little overall variability across arrays were filtered out by excluding the features that had a t-statistic of <1.6 in all of the cell type groups. This left a subset of 2150 features in the data set.

Data Normalization. Many microarray studies have shown that the usual log ratios tend to have variances that change with gene expression intensity (9). This makes inference based on fold difference (10) alone difficult due to heteroscedasticity. Also, the features in which the intensity of one or two channels is nonpositive have to be dismissed. The variance stabilizing normalization (9) procedure we used avoids these drawbacks. It applies a transformation h to the raw intensities of both green and red channels to remove the dependency of the variance on the mean. The differences between the transformed values can be viewed as "generalized log ratios" because for high intensities, x_i and x_j,

$$h = h(x_i) - h(x_j) \cong \log(x_i) - \log(x_j)$$

After normalization, not only was the Lowess curve flat around 0, but also the majority of the data points have a constant variability.

Using Bioconductor's multitest, adjusted *P*s were calculated using an implementation of the step-down procedure for multiple testing of Westfall and Young (11). The adjusted *P*s given to the genes that have marked differential expression between melanoma and healthy donors were rank ordered, and the 50 genes with the smallest were retained and submitted to the choice of discriminatory genes procedure detailed below.

Real-Time RQ-PCR. To validate the hypotheses generated by the microarray study, quantitative RQ-PCR was performed. Five genes were chosen according to multiple testing procedures. The PCR primers specific to these genes were designed using Primer 3 software.[5] All of the primers were designed with melting temperature 58-60°C and resulting products between 100 and 150 bp. For each healthy donor and melanoma patient, a small aliquot of total RNA extracted from each CD8+ subset was saved for quantitative PCR (separate from the total RNA used for linear amplification). cDNA was transcribed from total RNA using oligodeoxythymidylate primer and Superscript II (Invitrogen) in 20 μl of reaction volume. One μl of cDNA was then carried out in triplicate using iQ SYBR Green Supermix (Bio-Rad). The PCR conditions were as follows: 7 min at 95°C for initial denaturing, followed by 40 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 30 s in the Bio-Rad iQ real-time sequence detection system. Levels of each gene were normalized to expression levels of a housekeeping gene, glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*).

The primers for each gene are: FLJ22059 clone, 5′-GGAGACCATAGCAGCGAGTC-3′5′-TCTTCCCACATGGACATGAA-3′; DDB2, 5′-CTCCTCAATGGAGGGAACAA-3′5′-GTGACCACCATTCGGCTACT-3′; *N*-myristoyltransferase 2, 5′-GAACATTGATGAGGCTGCAA-3′5′-ACCCTGTGGCAAAGAATACG-3′; Galectin-1, 5′-GGAACATCCTCCTGGACTCA-3′5′-CAGGTTGTTGCTGTCTTTGC-3′; Hypothetical Protein 669, 5′-ATGGAAAAGCCATCAAGGTG-3′5′-GGTTCCTCCACTTCCTCCTC-3′; Paraspeckle 1, 5′-CTACGGATTCGCTTCGCTAC-3′5′-CGATCATCCACAACCACAAC-3′; and GAPDH, 5′-CAGCCTCAAGATCATCAGCA-3′5′-GTCTTCTGGGTGGCAGTGAT-3′.

RESULTS

Frequencies of Naïve, Memory, and Effector CD8+ T Cells in Healthy Controls and Melanoma Patients. The percentages of naïve, memory, and effector CD8+ subsets in each healthy donor or melanoma patient sample were determined based on CD27 and CD45RA expression. There was no statistically significant differences ($P > 0.05$) in the relative distribution of the subsets either among subjects within each group or between melanoma and healthy subjects, confirming that melanoma patients do not have gross perturbations of their CD8+ T-cell subsets. CD8+ T-cell subsets were then isolated from PBMC samples by FACSorting based on their expression of CD27 and CD45RA (as shown in Fig. 1), then their RNA extracted, amplified, and hybridized onto microarrays.

[5] Internet address: http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi.

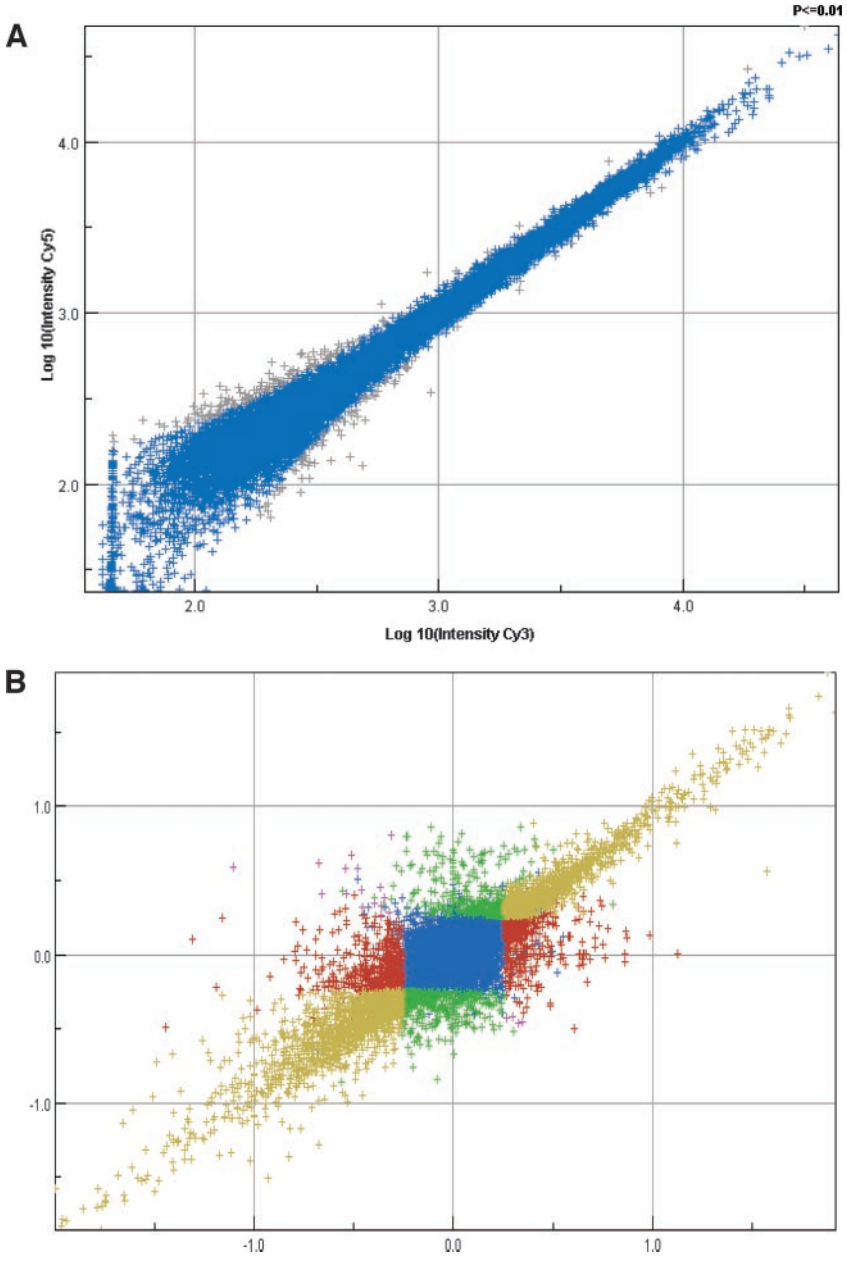

Fig. 2 Quality control plots. *A*, microarray data from a self-self experiment hybridized with Cy-3- and Cy-5-labeled cDNA generated from amplified RNA from 100,000 sorted CD8+ T cells. *Data points* represent the red fluorescent signal intensity *versus* the green fluorescent signal intensity. *B*, combined microarray data from two microarrays representing the two halves of a dye swap experiment. In one experiment, the microarray was hybridized with Cy-3-labeled T-cell cDNA and Cy-5-labeled reference cDNA generated from 500 ng of the corresponding amplified RNA. In the second experiment, the microarray was hybridized with Cy-5-labeled T-cell cDNA and Cy-3-labeled reference cDNA. Together, the combined plot represents the two polarities in a dye swap experiment. The *yellow points* are genes that displayed good correlation with similar differential expression for both polarities of the dye swap experiment. Shown in *blue* are genes that are unchanged and in *red* are genes that were found to be differentially expressed in one polarity but unchanged in the other polarity. The *pink points* represent anticorrelated genes in the two polarities.

Quality of RNA Extraction, Linear Amplification, and Labeling. To determine the amount, quality, and reproducibility of RNA extracted from small numbers of cells, we sorted various numbers of CD8+ T cells from healthy donor PBMC ranging from 50,000 to 500,000 cells.

We optimized the cell sorting conditions to maintain RNA integrity and the RNA isolation method to minimize RNA degradation and genomic DNA contamination. Genomic DNA contamination can become problematic when isolating RNA from small cell numbers because the fraction of genomic DNA relative to the percentage of RNA increases with decreasing cell number. One of the main problems with genomic DNA contamination is elevation of the RNA concentration when using UV spectrophotometry, because RNA and genomic DNA absorb at the same wavelength. Because the RNA yield was too low to measure by UV spectrophotometry, an estimation of RNA concentration was determined using the Bioanalyzer Pico assay software, which enabled us to compare RNA yields from each patient sample to ensure that the yields were comparable (data not shown). It was estimated that the average yield of total RNA from 100,000 cells was 20–60 ng based on the concentration reported by the RNA Pico Assay software and generating a standard curve from a sample of known concentration. The limited amount of sample RNA did not allow for replicate hybridizations of each sample. Thus, we performed control experiments to test the reproducibility of our sample labeling method. RNA was isolated from 200,000 T cells (unsorted) and divided into several aliquots. Each sample was linearly amplified and then converted to fluorescently labeled cDNA. The system noise was determined using self-self hybridizations in which the same sample was labeled in both cyanine-3 and cyanine-5. In Fig. 2*A*, data points represent the red signal *versus* the green signal for each feature on the microarray. The data converged to a tight line corresponding to the expected log ratio of 0, indicating that the signal from the two individual labeling reactions is nearly identical. Less than 1.5% of the genes on the microarray were identified as genes differentially expressed ($P < 0.01$), indicating the low level of system noise. To additionally check that the difference in the red and green dyes was

not introducing additional variability, we performed dye swap experiments and analyzed the combined plot using the Rosetta Resolver system shown in Fig. 2*B*. The scatter plot shows that 97% of the genes were correlated, with 2773 genes showing strong correlations in both polarities. The anticorrelation was <1% (pink crosses).

Together, these data show that good quality RNA could be extracted from small numbers of sorted human T cells, and the linear amplification protocol is reliable and reproducible, with low system noise.

Discriminant Analysis between Healthy and Melanoma T Cells. Gene expression data were extracted from each hybridized array using the Agilent Feature Extraction software with Lowess background correction. After the variance stabilizing normalization transformation, an initial analysis of these data showed no statistically significant differences in gene expression between T cells from healthy and melanoma patients. Additional analysis suggested that this may have occurred due to large gene expression differences among naïve, effector, and memory T cells, which could mask potentially more subtle differences between melanoma and healthy T cells. To maximize our ability to detect subtle differences in gene expression, we thus removed the cell type effects and pooled the gene expression data from all three of the T-cell subsets from each subject using a simple linear model on the transformed data. Next, the Westfall and Young multiple testing procedure was applied to detect any differences between T cells from melanoma patients and healthy donors. A subset of 50 genes with the smallest *P*s was selected by taking the 50 features with smallest adjusted *P*s. These were used as the input into a discriminant analyses (a simple, efficient supervised learning procedure; see Ref. 12), with cross-validation that chose the best subset for discriminating the arrays into healthy and melanoma groups with low classification error.

Eleven features (10 genes) were found to consistently discriminate between T cells from melanoma patients *versus* healthy controls with adjusted *P*s $\leq$ 0.05 (Table 1).

Check and Recalibration. Classical fold changes are no longer meaningful with the renormalization and variance stabilizing transformation procedures necessary to render this large data set homogeneous for analyses. To derive estimates of the approximate fold differences in the genes found to be differentially expressed, we made use of reference genes for which we knew the fold difference from a covariable; in this case the covariable was gender. Male subjects were in a 2:1 ratio between the melanoma and healthy patients in this study, creating an expected fold difference that was observed on several specific Y-chromosome genes. When we performed the multiple testing procedure with the Y genes reinserted into the dataset, the SMC (mouse) homologue Y gene appears similar to galectin-1, with an adjusted *P* of 0.034. This suggests that if the variances are comparable, the level of the differential expression between the two groups of patients is at ~2-fold for the galectin-1 gene.

Hierarchical Clusterings. A hierarchical clustering of the data was performed using the mva package in R. The results are displayed here for the 11 retained features.

Fig. 3*A* shows that this smaller subset of genes classifies healthy donors and melanoma patients into two groups. To find the most discriminatory genes, we used a cross-validation method (13). This method consists in removing one observation (melanoma patient or healthy patient), calculating a new discriminating rule on the 29 remaining patients, and then using the new rule to attribute the deleted observation to a class (melanoma or healthy) and recording whether the observation was actually well classified by the rule. This is repeated 30 times, once for each observation, and the prediction error of the discriminating rule is assigned the percentage of improperly classified observations. We took all subsets of genes of size <20 from the 50 genes prefiltered to have the smallest multiple testing *P*s and chose the subset with the best cross-validated prediction score. This turned out 11 features (10 genes). This particular subset gave a cross-validation score of 100% of the observations well classified, encouraging us to think that these 10 genes were indeed good predictors of T cells from melanoma *versus* healthy subjects. The expression for each gene in this subset was significantly different between healthy donors and melanoma patients ($P < 0.05$). These data confirm that CD8+ T cells in melanoma patients have subtle differences in gene expression from CD8+ T cells from healthy donors.

The same analysis was carried out on a set of 86 classical apoptosis genes, of which the 9 most discriminant were chosen based on the same cross-validation estimation of classification quality. These 9 genes were then analyzed using the same hierarchical clustering as the 10 overall best genes. Clustering done on classical apoptosis genes shows that the distinction between melanoma and healthy patients cannot be made simply on the evidence of genes involved in the classical apoptosis pathway (Fig. 3*B*).

Gene Confirmation with Real-Time Reverse Transcription-PCR. To confirm the significance of the genes identified to be differentially expressed using microarrays, we performed RQ-PCR analysis using the original, unamplified RNA materials. CD8+ T-cell subsets were not combined in these experiments. Due to limited quantities of unamplified RNA, 6 of 10 genes were selected for RQ-PCR analysis: *DDB2, PSPC1, FLJ10955, Galectin-1, N- myristoyltransferase,* and *Hypothetical Protein 669*. A housekeeping gene, *GAPDH,* was also amplified for normalization of data. The threshold cycle for each sample is calculated by iCycler Real Time Detection System. After normalization with GAPDH, the threshold cycle of the 6 genes was analyzed using a Wilcoxon test. As shown in Table 2, significant differences ($P < 0.05$) were found in at least one CD8+ T-cell subset between melanoma patients and healthy donors confirming results from the microarray data. The only exception was expression of gene FLJ22059, which did not show significant differences between healthy and melanoma in any of the three subsets.

Table 1 *Discriminating features, ranked in order of significance*

Unigene number	UniGene abbreviation	UniGene name/description	Adjusted *P*	Relative expression[a]
Hs.16364	PSPC1	Paraspeckle component 1	0.0002	⇓
Hs.446564	DDB2	Damage-specific DNA binding protein 2 (48 kD)	0.0160	⇑
Hs.403997	VIL2	Villin 2 (ezrin)	0.0278	⇑
Hs.407909	LGALS1	Lectin, galactoside-binding, soluble, 1 (galectin 1)	0.0278	⇑
Hs.78877	ITPKB	Inositol 1,4,5-trisphosphate 3-kinase B	0.0340	⇑
Hs.407909	LGALS1	Lectin, galactoside-binding, soluble, 1 (galectin 1)	0.0340	⇑
—	—	Hypothetical protein 669[b]	0.0492	⇓
Hs.60339	NMT2	*N*-myristoyltransferase 2	0.0492	⇓
Hs.152335	—	LOC154084; Homo sapiens clone 24889 mRNA sequence	0.0492	⇑
Hs.433416	NME2	Nonmetastatic cells 2, protein (NM23B) expressed in	0.0492	⇑
Hs.13323	FLJ22059	Hypothetical protein FLJ22059	0.0492	⇑

[a] Melanoma versus healthy controls.
[b] Sequence does not have a Unigene number.

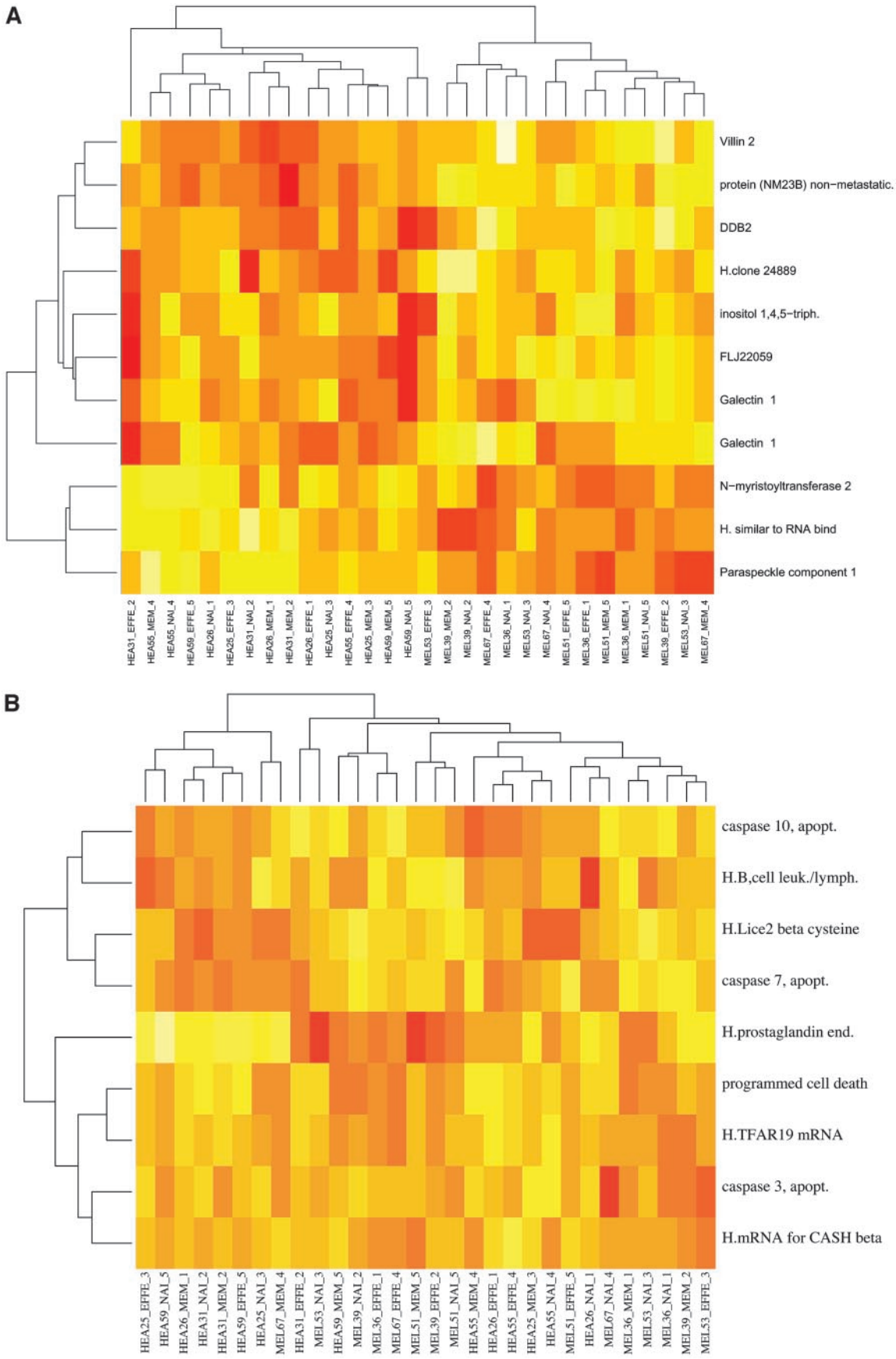

Fig. 3 Hierarchical clustering of two filtered gene set. The *white* are the most expressed, the *red* the least, with *yellow* as the intermediary. Hierarchical clustering (unsupervised) was done using the subsets of genes chosen as best classifiers by a cross-validation analysis. *A* shows the results when the subset of genes chosen was taken from all genes filtered to have a significant between-variance:within-variance ratio. Note that the clustering separates the two groups neatly between melanoma and healthy patients. *B* shows the same procedure but where the original gene set was restricted to genes involved in apoptosis. This clustering, although done using the best possible classifier, is unable to separate the two groups distinctly.

Table 2 *Results from two sample t tests, by cell types*

Unigene number	UniGene abbreviation	UniGene name/description	Naïve	Effector	Memory
Hs.16364	PSPC1	Paraspeckle component 1	0.53	**0.0498**	0.54
Hs.446564	DDB2	Damage-specific DNA binding protein 2 (48 kD)	0.28	0.27	**0.027**
Hs.407909	LGALS1	Lectin, galactoside-binding, soluble, 1 (galectin 1)	**0.048**	**0.041**	**0.021**
—	—	Hypothetical protein 669	0.08	$\mathbf{6.10^{-6}}$	0.69
Hs.60339	NMT2	*N*-myristoyltransferase 2	0.27	**0.0003**	0.45
Hs.13323	FLJ22059	Hypothetical protein FLJ22059	0.75	0.62	0.21

DISCUSSION

Specific and global T-cell abnormalities have been found in patients with a variety of cancers (4, 5). The underlying mechanism(s) of this remain unclear. In this study, we used DNA microarray technology to determine whether gene expression differences exist between circulating T cells from melanoma patients *versus* from healthy donors as a basis of T-cell abnormalities in cancer. A recent report showed that there are indeed gene expression differences in PBMC samples isolated from patients with advanced renal cell carcinoma as compared with those from healthy subjects (14). This is an important observation showing that cancer can lead to changes in gene expression in circulating blood cells. However, because PBMCs consist of heterogeneous mixtures of CD4+T cells, CD8+T cells, B cells, natural killer cells, monocytes, and some granulocytes, it is unclear whether gene expression differences exist within all cell populations or in specific populations between cancer patients and healthy subjects. To derive maximum resolution in our data, we focused on CD8+ T cells and sorted them into naïve, effector, and memory subsets. An important technical hurdle we faced was the amounts of RNA available from such sorted purified cell populations. Analysis of transcript abundance on microarrays requires at least micrograms of total RNA. Without amplification, such amounts can only be obtained from millions of cells. RNA isolated from a few cells as in aspiration biopsies or rare populations isolated by cell sorting and laser capture is far below this amount. Two approaches have been developed to solve this problem, signal amplification and RNA amplification. Signal amplification methods include indirect labeling technique (15). Thus far, signal amplification methods can cut down the sample RNA for labeling to 2.5 μg. Global RNA amplification methods require less RNA template than signal amplification. There are two global RNA amplification methods, T7-based linear amplification and PCR exponential amplification (16). Exponential amplification is believed to decrease transcript abundance relationships and introduce bias (17). Several systematic studies of linear amplification have already been published (18–21).

Most systematic studies use serial dilution of standard RNA or serial dilution of total RNA extracted from large numbers of cells, so the RNA quality is easy to control. To keep the integrity of RNA, we optimized the sorting condition to prevent RNA degradation in our small, individually sorted T-cell samples. From 100K sorted CD8+ T cells, 20–60 ng total RNA were isolated. Of the extracted RNA, 10% was measured using Bioanalyzer, ensuring the quality of the RNA used for amplification. After one round of amplification, the yield of antisense RNA from the 100K CD8+ T cells was $\sim$1 μg available for the labeling and hybridization steps.

Another important aspect of amplification is the degree of reproducibility. We showed that the correlation coefficient between individual hybridization for aRNA amplified from the same total RNA is high using the Agilent amplification protocol. As for the fidelity of T7-based linear amplification, there are several publications showing that the correlation coefficient between amplified and unamplified samples range from 0.83 to 0.86 (22). We found similar results in our experiments (data not shown). Bias as compared with unamplified materials is unavoidably induced when samples are amplified. Importantly, bias introduced into gene expression by amplification is reproducible and systematic (17, 23), hence allowing for reliable comparisons of amplified samples against each other. Gene expression profiling using aRNA provides an approximation of the true expression profile of the original sample. Although duplication is always desirable, for clinical samples it is often difficult to obtain sufficient aRNA to hybridize onto two arrays. Data from our dye swap and self-to-self experiments show high correlation coefficients.

Most microarray experiments performed thus far have been on heterogeneous cell populations, such as tumor biopsies (Ref. 6; which may contain tumor cells, stromal cells, immune cells, and other contaminants), and PBMC (Refs. 24, 14; which contain CD4+ T cells, CD8+ T cells, B cells, natural killer cells, monocytes, and some granulocytes). With such heterogeneity in cells, it is impossible to determine from which cells certain gene expression differences arise. In this study, to maximize the resolution of our data, we focused on CD8+ T cells only, and sorted them into naïve, effector, and memory subsets and hybridized these individually on microarrays. However, because our sample size was small (5 healthy donors and 5 melanoma patients), we removed the cell subset effect in our initial microarray data analysis, that is, we normalized the median of expression of all three of the subsets. This eliminated the subset effect and allowed us to pool the three subsets to look for other more subtle differences than those depending on cell type. Importantly, we confirmed our microarray data using quantitative (real-time) PCR on unamplified RNA extracted from individual T-cell subsets.

On the basis of the analytical model in this study, 11 features (10 genes) provide a robust classifier between CD8+ from melanoma patients and CD8+ from healthy donors. In other words, the combination of these 10 genes is more reflective of molecular behavior of the CD8+ in melanoma patients. Because T cells in cancer patients are known to undergo higher rates of apoptosis, we specifically looked for differential expression of genes classical linked to apoptosis. We generated a list of 86 genes commonly associated with apoptosis, including *caspases 1–9, bcl-2, Fas (CD95),* and many others. Importantly, the same discriminant analysis applied to this sublist of apoptosis genes did not yield a set that could differentiate between the melanoma and healthy patients.

Possible mechanism(s) underlying T-cell abnormalities in cancer patients include factors secreted by tumor cells (*e.g.,* tumor necrosis factor α), cell-cell contact (*e.g.,* Fas-Fas ligand), and repeated stimulation leading to activation- induced cell death. The classical apoptosis pathway involves a family of cysteine proteases, known as caspases, which cleave many vital cellular proteins and proteolytically activate enzymes involved in cell death (25). Several genes we found to be differentially expressed in T cells of melanoma *versus* healthy subjects may be linked directly or indirectly to apoptosis. Galectin-1 is a galactoside binding protein that has been linked to apoptosis in T cells (26–28). Damage-specific DNA binding protein 2 mediates damage repair to DNA; recent evidence suggests that DNA binding protein 2 interacts with p53 in regulating apoptosis (29). The addition of the carbohydrate moiety *N*-myristoyl to certain apoptosis regulators, such

as BID (30), could modulate their activity. Hence, the enzyme *N*-myristoyltransferase 2 may play an indirect role in apoptosis. Ezrin (villin) is a key protein in membrane-cytoskeleton interaction, and is involved in membrane polarization and shape modulation. Ezrin has been shown to be involved in CD95-mediated apoptosis by modulating CD95 linkage to the actin cytoskeleton (31). Whereas the function of PSPC1 is unknown, it belongs to the same family of proteins as p54/nrb, which has been linked to CD95-induced apoptosis in T cells (32).

Importantly, we could not find evidence for the differential expression of classical apoptosis mediators such as caspases, Fas/Fas ligand, bcl-2, bcl-xL, BH3, Bim, Bax, Bak, FLIP, FADD, and Apaf-1, by T cells from melanoma *versus* healthy subjects (as shown in Fig. 3). This raises the possibility that a secondary (alternate) apoptosis pathway may be triggered in circulating T cells of cancer patients. Of note, as many caspases are already present in cells as catalytically dormant proenzymes (zymogens). which are activated in response to stimuli that trigger apoptosis, perhaps it is not unexpected that transcription of some of these genes is not altered in cells undergoing apoptosis. Nonetheless, tumor necrosis factor α (which is known to be secreted by many cancers) has been reported to trigger a secondary (non-caspase mediated) apoptosis pathway. Tumor necrosis factor α seems to trigger cathepsin B and L (not caspase), which activates apoptosis through the mitochondria.

Microarrays may emerge to be a powerful new tool to elucidate the mechanism of cancer-induced immune dysfunction. Our approach aimed to minimize spurious sources of variation such as heterogeneity in cell populations and dye effects. By using a variance stabilization transformation, we were able to detect subtle differences in the expression of a small set of genes that differentiate CD8+ T cells from healthy and melanoma patients at an ~2-fold level. Importantly, we were able to confirm these microarray results by RQ-PCR experiments. Taken together, our data support previous observations of T-cell apoptosis in cancer patients and raise the possibility of an alternative apoptosis pathway being triggered in T cells by cancer. Our data also suggest that different T-cell types may be impacted differently by cancer, and thereby illustrates the benefit of gene expression analysis of purified cell populations.

REFERENCES

1. Lee PP, Yee C, Savage PA, et al. Characterization of circulating T cells specific for tumor-associated antigens in melanoma patients. Nat Med 1999;5:677–85.
2. Staveley-O'Carroll K, Sotomayor E, Montgomery J, et al. Induction of antigen-specific T cell anergy: an early event in the course of tumor progression. Proc Natl Acad Sci USA 1998;95:1178–83.
3. Ochoa AC, Longo DL. Alteration of signal transduction in T cells from cancer patients. Important Adv Oncol 1995;55:43–54.
4. Kuss I, Donnenberg AD, Gooding W, Whiteside TL. Effector CD8+CD45RO-CD27-T cells have signalling defects in patients with squamous cell carcinoma of the head and neck. Br J Cancer 2003;2:223–30.
5. Bauernhofer T, Kuss I, Henderson B, Baum AS, Whiteside TL. Preferential apoptosis of CD56dim natural killer cell subset in patients with cancer. Eur J Immunol 2003;33:119–24.
6. Brown PO, Botstein D. Exploring the new world of the genome with DNA microarrays. Nat Genet 1999;21(Suppl 1):33–7.
7. Eberwine J. Amplification of mRNA populations using aRNA generated from immobilized oligo(dT)-T7 primed cDNA. Biotechniques 1996;20:584–91.
8. Ihaka R, Gentleman R. R: A language for data analysis and graphics. J Computat Graph Stat 1996;5:299–314.
9. Huber W, Von Heydebreck A, Sultmann H, Poustka A, Vingron M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics (Oxford) 2002;18(Suppl 1):S96–104.
10. Chen Y, Dougherty E, Bittner M. Ratio based decisions and the quantitative analysis of cDNA micro-array images. J Biomed Optics 1997;2:364–74.
11. Westfall P, Young S. Resampling-based multiple testing: examples and methods for p value adjustment. New York: John Wiley & Sons; 1993.
12. Dudoit S, Fridlyand J, Speed TP. Comparison of discrimination methods for the classification of tumors using gene expression data. J Am Statist Assoc 2002;97:77–87.
13. Stone M. Cross-validatory choice and assessment of statistical predictions. J Royal Statistical Society B 1974;36:111–47.
14. Twine NC, Stover JA, Marshall B, et al. Disease-associated expression profiles in peripheral blood mononuclear cells from patients with advanced renal cell carcinoma. Cancer Res 2003;63:6069–75.
15. Manduchi E, Scearce LM, Brestelli JE, et al. Comparison of different labeling methods for two-channel high-density microarray experiments. Physiol Genomics 2002;10:169–79.
16. Iscove NN, Barbara M, Gu M, et al. Representation is faithfully preserved in global cDNA amplified exponentially from sub-picogram quantities of mRNA. Nat Biotechnol 2002;20:940–3.
17. Baugh LR, Hill AA, Brown EL, Hunter CP. Quantitative analysis of mRNA amplification by in vitro transcription. Nucleic Acids Res 2001;29:E29.
18. Gomes L, Silva R, Stolf B, et al. Comparative analysis of amplified and nonamplified RNA for hybridization in cDNA microarray. Anal Biochem 2003;321:244–51.
19. Nygaard V, Loland A, Holden M, et al. Effects of mRNA amplification on gene expression ratios in cdna experiments estimated by analysis of variance. BMC Genomics 2003;23:11.
20. Puskas L, Zvara A, Hackler LJ, Van Hummelen P. RNA amplification results in reproducible microarray data with slight ratio bias. Biotechniques 2002;32:1330–4, 1336, 1338, 1340.
21. Feldman A, Costouros N, Wang E, et al. Advantages of mRNA amplification for microarray analysis. Biotechniques 2002;33:906–12.
22. Zhao H, Hastie T, Whitfield M, Borresen-Dale A, Jeffrey S. Optimization and evaluation of T7 based RNA linear amplification protocols for cDNA microarray analysis. BMC Genomics 2003;3:31.
23. Hu L, Wang J, Baggerly K, et al. Obtaining reliable information from minute amounts of RNA using cDNA microarrays. BMC Genomics 2002;3:16.
24. Whitney AR, Diehn M, Popper SJ, et al. Individuality and variation in gene expression patterns in human blood. Proc Natl Acad Sci USA 2003;100:1896–901.
25. Marsden VS, O'Connor L, O'Reilly LA, et al. Apoptosis initiated by Bcl-2-regulated caspase activation independently of the cytochrome c/Apaf-1/caspase-9 apoptosome. Nature (Lond) 2002;419:634–7.
26. Pace KE, Lee C, Stewart PL, Baum LG. Restricted receptor segregation into membrane microdomains occurs on human T cells during apoptosis induced by galectin-1. J Immunol 1999;163:3801–11.
27. Zuniga E, Rabinovich GA, Iglesias MM, Gruppi A. Regulated expression of galectin-1 during B-cell activation and implications for T-cell apoptosis. J Leukoc Biol 2001;70:73–9.
28. Meissner N, Radke J, Hedges JF, et al. Serial analysis of gene expression in circulating $\gamma\ \delta$ T cell subsets defines distinct immunoregulatory phenotypes and unexpected gene expression profiles. J Immunol 2003;170:356–64.
29. Itoh T, O'Shea C, Linn S. Impaired regulation of tumor suppressor p53 caused by mutations in the xeroderma pigmentosum ddb2 gene: mutual regulatory interactions between p48(ddb2) and p53. Mol Cell Biol 2003;23:7540–53.
30. Zha J, Weiler S, Oh KJ, Wei MC, Korsmeyer SJ. Posttranslational n-myristoylation of bid as a molecular switch for targeting mitochondria and apoptosis. Science (Wash DC) 2000;290:1761–5.
31. Parlato S, Giammarioli AM, Logozzi M, et al. CD95 (APO-1/Fas) linkage to the actin cytoskeleton through ezrin in human T lymphocytes: a novel regulatory mechanism of the CD95 apoptotic pathway. EMBO J 2000;19:5123–34.
32. Thiede B, Dimmler C, Siejak F, Rudel T. Predominant identification of RNA-binding proteins in Fas-induced apoptosis by proteome analysis. J Biol Chem 2001;276:26044–50.