[CANCER RESEARCH 63, 5352–5356, September 1, 2003]

Array Comparative Genome Hybridization for Tumor Classification and Gene Discovery in Mouse Models of Malignant Melanoma[1]

Rónán C. O'Hagan,[2] Cameron W. Brennan,[2] Andrew Strahs, Xuegong Zhang, Karuppiah Kannan, Melissa Donovan, Craig Cauwels, Norman E. Sharpless, Wing Hung Wong, and Lynda Chin[3]

Department of Medical Oncology, Dana Farber Cancer Institute [R. C. O., C. W. B., K. K., M. D., C. C., N. E. S., L. C.], Department of Biostatistics, Harvard School of Public Health [A. S., X. Z., W. H. W.], Belfer Cancer Genomics Centre, Dana Farber Cancer Institute [M. D., C. C.], and Department of Dermatology, Harvard Medical School [L. C.], Boston, Massachusetts 02115; Department of Automation, Tsinghua University, Beijing 100084, China [X. Z.]; and Department of Statistics, Harvard University, Cambridge, Massachusetts 02138 [W. H. W.]

ABSTRACT

Chromosomal numerical aberrations (CNAs), particularly regional amplifications and deletions, are a hallmark of solid tumor genomes. These genomic alterations carry the potential to convey etiologic and clinical significance by virtue of their clonality within a tumor cell population, their distinctive patterns in relation to tumor staging, and their recurrence across different tumor types. In this study, we showed that array-based comparative genomic hybridization (CGH) analysis of genome-wide CNAs can classify tumors on the basis of differing etiologies and provide mechanistic insights to specific biological processes. In a RAS-induced $p19^{Arf-/-}$ mouse model that experienced accelerated melanoma formation after UV exposure, array-CGH analysis was effective in distinguishing phenotypically identical melanomas that differed solely by previous UV exposure. Moreover, classification by array-CGH identified key CNAs unique to each class, including amplification of cyclin-dependent kinase 6 in UV-treated cohort, a finding consistent with our recent report that UVB targets components of the $p16^{INK4a}$-cyclin-dependent kinase-RB pathway in melanoma genesis (K. Kannan, *et al.*, Proc. Natl. Acad. Sci. USA, *21:* 2003). These results are the first to establish the utility of array-CGH as a means of etiology-based tumor classification in genetically defined cancer-prone models.

INTRODUCTION

Identification of etiological factors and associated genetic events underlying malignant transformation and progression remains a central focus in the field of cancer. Chromosomal imbalances, particularly amplifications and deletions, are common in solid tumors, including melanoma (1–3). In human melanoma, conventional cytogenetic methods have detected specific CNAs,[4] most prominently LOH of 1p, deletion of 6q22–27, amplification of 7, LOH or deletion of 9p21, and LOH of 10q24–26 (2, 3). The notion that recurrent CNAs represent selected genetic events driving specific processes in the evolution of cancers suggests that the patterns of CNAs may convey etiologic and prognostic information; thus, characterization of such CNAs will provide insight on the molecular mechanism underlying such processes.

Recent advances in microarray technology have revolutionized our ability to catalogue chromosomal imbalances and gene expression patterns present in tumors. Microarray-based expression studies have established the feasibility of tumor classification based on gene expression patterns (4, 6, 7–12). In addition, profiling studies in human tumors, such as melanomas, have implicated tumor suppressor genes (13) and oncogenes with plausible links to metastasis (14), suggesting that genome-wide expression profiles might serve as a means of identifying genes and pathways governing specific cancer-relevant processes. Although array-CGH has been highly effective in cataloging recurrent CNAs associated with a number of different tumor types in humans (15–20) and in engineered cancer-prone mouse strains (21–24), array-CGH analyses have not yet been assessed as an approach to segregate tumors on the basis of differing etiologies or genetics or to identify those CNAs with potential links to specific tumor biological processes. To investigate whether array-CGH profiles can be used for these purposes, we compared the scope and nature of the chromosomal CNAs by array-CGH in phenotypically and histopathologically homogeneous melanomas arising spontaneously or after exposure to UVB in mouse models.

A Tyr-RAS transgenic model of melanoma has demonstrated that activated RAS mutation can cooperate with loss of $p16^{INK4a}$, $p19^{ARF}$, or both to drive melanoma formation.[5] To explore the role of neonatal UV exposure in melanoma development, we subjected *Tyr-RAS+* $p16^{Ink4a\ -/-}$ and *Tyr-RAS+* $p19^{ARF-/-}$ animals to a single erythrogenic dose of UVB on postnatal day 1 and compared the latency and incidence of melanoma in these UV-treated animals with that in the untreated cohort (26). We found that UV exposure greatly accelerated melanoma formation in *Tyr-RAS+* $p19^{ARF-/-}$ animals but not in *Tyr-RAS+* $p16^{Ink4a\ -/-}$ mice. Despite different germ-line mutations ($p16^{INK4a}$ or $p19^{ARF}$ deficiency) and etiologies (spontaneous *versus* UV treated), the resultant melanomas were histopathologically indistinguishable and frequently sustained mutations that inactivated the Rb or p53 pathways in the $p19^{ARF}$ or $p16^{INK4a}$ mutant mice, respectively (26). Given their common tumor biological end points, these highly related melanomas from genetically equivalent *Tyr-RAS+* $p19^{ARF-/-}$ mice, differing only in their history of UV exposure, provided a stringent system in which to test the hypothesis that profiles of CNAs contain discriminatory information on tumor etiology and pathogenesis, and such unbiased information can uncover molecular events underlying etiologic difference and lead to new mechanistic insights.

MATERIALS AND METHODS

Tumor Genomic DNA Isolation. Genomic DNA was isolated from melanomas using the PUREGENE DNA isolation kit (Gentra Systems). Five μg of gDNA were digested with *Eco*RI, extracted with phenol:chloroform, ethanol precipitated, and resuspended in sterile distilled water.

Labeling and Hybridization to Arrays. Array-CGH experiments were performed with SpectralChip arrays (Spectral Genomics, Houston, TX) according to the manufacturer's protocol. These arrays comprise 976 mapped mouse BACs from the RPCI11 BAC library. One μg of tumor and normal genomic DNA was labeled with Cy3-cCTP or Cy5-dCTP (Amersham Pharmacia Biotech). Fluor-reversal experiments were performed in all cases. Hybridizations were performed in sealed chambers for 20 h at 60°C. After

Received 3/11/03; revised 5/16/03; accepted 6/25/03.

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.

[1] Supported in part by grants to L. C. (NIH UO1 CA84313, the Claudia Adams Barr Program in Cancer Research, and the Rockefeller Brothers Fund) and by a grant to W. H. W. (NIH P20 CA96470). L. C. is a Charles E. Culpeper Medical Scholar.

[2] R. C. O. and C. W. B. contributed equally to this work.

[3] To whom requests for reprints should be addressed, at Dana-Farber Cancer Institute, Harvard Medical School, 44 Binney Street (M413), Boston, MA 02115. Phone: (617) 632-6091; Fax: (617) 632-6069.

[4] The abbreviations used are: CNA, chromosomal numerical alteration; SVM, support vector machine; CGH, comparative genomic hybridization; LOH, loss of heterozygosity; Rb, retinoblastoma; cdk, cyclin-dependent kinase; BAC, bacterial artificial chromosome; Tyr-RAS, tyrosinase-H-RAS^{V12G}; RT-PCR, reverse transcription-PCR.

[5] L. Chin *et al.*, unpublished observations.

hybridization, slides were rinsed in 2 × SSC, 0.5% SDS at room temperature; washed for 40 min in 2 × SSC, 0.5% SDS at 65C; washed for 10 min in 0.2 × SSC at 65°C; and washed for 1 min in 0.2 × SSC at 2°C. Slides were then air dried by centrifugation before imaging.

Image and Data Analysis. Sixteen-bit TIF images were collected using an Axon 4000B scanner and processed initially using GenePix Pro. Subsequently, custom software was used to exclude spots that demonstrated low signal:noise ratios or poor spot morphology. Data were normalized to the ratio of medians of Cy3 and Cy5, and the normalized values converted to log2 ratios to weight gains and losses equally. Dye swap experiments were merged to calculate the mean and SD for each BAC.

Clustering and Classification. Cluster (http://rana.ibl.gov/) was used to cluster log2 values for array-CGH data processed as described above, using all BACs. Subsequently, BACs were excluded if the processing software generated at least one "missing value" for any tumor. Under this condition, 874 BACs were included in additional hierarchical clustering and classification experiments. The data were standardized so that BACs had common mean and variance.

A method based on the linear SVM (27, 28) was used to classify tumors using their array-CGH profiles for the remaining 874 BACs. Given a training set, in this case, tumors whose UV status is known, SVM builds a linear classifier:

$$f(\mathbf{x}) = \sum_i w_i x_i + b,$$

where x_i is the observed log florescence ratio for one tumor for BAC i sample, and w_i is its corresponding weight in the classifier. If $f(x) > 0$, the sample is assigned to class 1; otherwise, it is assigned to class 2. SVM minimizes the number of misclassifications in the training data and maximizes the separation margin between the two classes. $f(x)$ is based on the subset of training samples that best defines the separation between the two classes. These informative samples are called support vectors.

A software package, written by X. Z. and W. H. W., was used to implement the recursive classification and feature selection procedure introduced in "Results." This package incorporates the SVMTorch (29) software. Initially, SVMTorch is used to build an SVM on the 874 BACs identified above. The BACs with the largest contribution to the classifier are then selected. The contribution for a given BAC is measured by the difference between its two class means weighted by its coefficient in the SVM. A new SVM is generated on the reduced set, and a subset of these BACs is selected using the same procedure. This is done recursively, resulting in a series of nested subsets of BACs and a series of SVM models built using these BACs. For the current purpose of evaluating the feasibility of classification, we arbitrarily set the number of selected BACs in the above successive stages to be 500, 200, 100, 50, 30, and 20.

Cross-validation is implemented by removing one tumor sample from the data set and performing the above procedure on the remaining 37 samples. The series of models obtained at different selection levels are then applied to the excluded sample to predict its UV status. This process is duplicated for each tumor in the sample. The accuracy of the predictions across experiments is summarized for each size of classifier.

qPCR and RT-PCR. Primers for PCR and RT-PCR corresponded to the primers for the indicated MIT markers or were designed from the 3′untranslated region of the indicated genes using Primer 3. Reverse transcription was performed on 3 μg of total RNA from each tumor using the Superscript reverse transcriptase kit (Stratagene) and an oligo-dT primer according to the manufacturer's specifications in 20 μl of total volume. PCR was performed using 100 ng of genomic DNA or 1 μl of reverse transcriptase reaction mix with marker- or gene-specific primers for 17 cycles of 95°C for 30 s, 56°C for 1 min, and 72°C for 45 s. PCR products were run on a 1% agarose Tris-borate EDTA gel and transferred to Hybond N+. *cdk6* and *glyceraldehyde-3-phosphate dehydrogenase* were not visible on eithidium bromide-stained gels and only detected by hybridization to specific probes. Probes were generated from pooled normal genomic DNA from the mice of origin for each tumor by PCR using the same primers as for the quantitative PCR analysis. Random-primed ^{32}P-labeled probes were hybridized for 2 h at 65°C in RapidHyb (Amersham Pharmacia). Quantification was performed using a phosphorimager.

Mapping of Genes and Markers. Physical map positions of genes and markers were obtained from the Ensembl mouse genome database v4.1.1, January 31, 2002.

RESULTS AND DISCUSSION

UV-induced and spontaneous melanomas from *Tyr-RAS* transgenic animals on $p19^{ARF-/-}$ and $p16^{Ink4a-/-}$ backgrounds were subjected to array-CGH analysis (Table 1). For each tumor, regional chromosomal CNAs in tumor genomic DNA were compared with somatic DNA from the same mouse. The threshold for defining a CNA was based on the distribution of log ratios in self-self hybridizations (data not shown). Regions were scored as significantly increased/decreased if the CGH ratio was 4 SDs above/below the mean (+/− 0.35, log base 2).

Analysis of the overall genomic integrity, as measured by frequency of CNAs detectable by array-CGH, revealed a significantly higher rate of CNAs in spontaneous (non-UV) than UV-induced tumors from $p19^{ARF-/-}$ animals. On average, 4.2% of the informative genomic regions showed detectable copy number changes in spontaneous $p19^{ARF-/-}$ tumors, compared with 0.6% among UV-induced tumors ($P < 0.001$; Table 1). This global difference in genomic profiles suggests that UV exposure significantly affected the nature of accumulated mutations in $p19^{ARF-/-}$ melanomas and that different patterns of CNAs are likely to be identified in UV *versus* non-UV tumors. In contrast, UV treatment did not alter the frequency of CNAs in melanoma from *Tyr-RAS*+ $p16^{Ink4a\ -/-}$ animals (Table 1), consistent with the lack of biological impact by UV exposure in $p16^{INK4a-/-}$ animals (as measured by incidence, latency, and clinical behavior of melanoma, as well as status of the pRB and p53 pathways in the tumors; Ref. 26).

We first used a method based on the linear SVM to classify the non-UV and UV-induced $p19^{ARF-/-}$ tumors based on their array-CGH profiles (see "Materials and Methods"; Refs. 27 and 28). Thirty-eight $p19^{ARF-/-}$ tumors with high quality array-CGH data were used for this part of the analysis. Only 874 BACs showed valid readings in all 38 of these tumors. This filtering procedure eliminated the need to make assumptions about the causes of the low information content in the discarded tumors or BACs. Hierarchical clustering of these 38 tumors using 874 BACs was virtually identical to that based on the full set of 931 BACs (data not shown), suggesting that the ability to classify tumors is not substantially reduced by this filter.

Because the number of descriptive spots (BACs) exceeded the number of tumor samples tested, we reduced the number of BACs considered in the classification scheme by feature selection. In designing a feature selection procedure to form classifiers based on smaller numbers of BACs, an optimal SVM model was built using all available BACs, and the BACs were ranked according to their contribution to this model. A subset of the BACs that made the largest contribution to this model was selected, and a new SVM model was built with these selected BACs. The selected BACs were then reranked according to their contribution to the new model, and the selection was repeated. For each model, the performance of SVM can

Table 1 *Array-CGH analysis of Tyr-RAS+ $p16^{INK4a-/-}$ and $p19^{ARF-/-}$ melanomas*

A.	Genotype	UV Rx	No. tumors	Mean instability[a]	SE
	$p19^{ARF-/-}$	−	20	4.2%	0.83
		+	20	0.6%	0.22
	$p16^{Ink4a-/-}$	−	11	1.5%	0.42
		+	7	1.6%	1.02
	Control[b]			0.1%	

B.	Comparisons: Genotype	Treatment	*P*
	$p19^{ARF-/-}$	UV *vs.* non-UV	<0.001
	$p16^{Ink4a-/-}$	UV *vs.* non-UV	0.931
	$p19^{ARF-/-}$ *vs.* $p16^{Ink4a-/-}$	Non-UV	0.013

[a] Instability = % of valid BACs with change > 4 SD from the mean of self-to-self hybridizations.
[b] Control = self-to-self hybridization.

be assessed. For our current study, we have arbitrarily chosen the numbers of selected features (BACs) for this recursive procedure to be 500, 200, 100, 50, 30, and 20. As shown in Fig. 1, the 874-features classifier achieved a higher accuracy in the UV-treated cohort alone, whereas the 20-feature classifier achieved a higher accuracy in the non-UV cohort. However, the overall classification accuracy (combined for both cohorts) is similar for both. The observation that a small-feature classifier was able to achieve similar accuracy as the large data set suggests the likelihood of identifying key stereotypical changes that may serve as potential prognostic/predictive markers.

Cross-validation was used to assess the quality of the SVM classification models given the small set of tumor profiles. For each tumor, the feature selection procedure described above was performed to establish the SVMs using the remaining set of 37 tumors. At each step, the classifier was applied to the excluded tumor, and the accuracy of the classification was recorded. Fig. 1 shows the results of this validation procedure for the 38 tumors, where the data are standardized so that each BAC has common mean and variance. This analysis revealed that we were able to achieve a maximum overall accuracy of classification of 76.5%. To assess the significance of this finding, we repeated the cross-validation procedure on 100 random permutations of the UV labels for all of the 38 tumors. We found that none of the 100 runs of R-SVM permutation test achieved a better cross-validation error rate than that observed in the actual classification, indicating that the 76.5% accuracy is significant. In other words, UV exposure history can be predicted retrospectively based on the array-CGH profiles of the resulting tumors. This finding is particularly encouraging in view of the fact that detailed histopathological and classical analyses were not able to distinguish tumors with or without UV exposure history, suggesting that array-CGH profiles reveal aspects of cancer etiology that are not detectable by standard approaches. Finally, it is unlikely that array-CGH merely identified a UV signature profile, because SVM analysis was unable to distinguish the UV status of melanomas arising from *Tyr-RAS+* $p16^{INK4a-/-}$ animals (data not shown). This lack of distinction on the genomic level is consistent with the absence of any observed UV effect in melanoma genesis on the $p16^{Ink4a-/-}$ background (26). In summary, the computational analysis of array-CGH profiles was capable of correctly classifying tumors that exhibited differences in biological behavior ($p19^{ARF-/-}$ UV *versus* non-UV) but could not differentiate among tumors with similar biological behavior ($p16^{Ink4a-/-\ UV}$ *versus* non-UV).

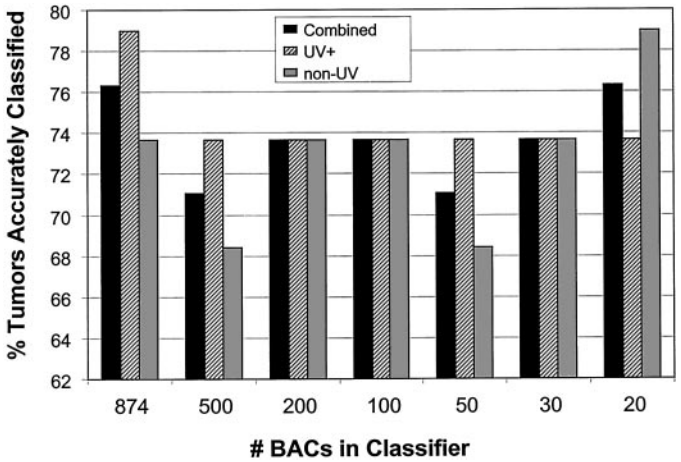

Fig. 1. UV Status of melanomas is predicted by array-CGH profiles. The ability of linear SVM to predict accurately whether a tumor is UV treated (UV+) or untreated (non-UV) is indicated for classification analyses performed with different numbers of BACs as described in the text and "Materials and Methods." *Black bars*, the combined/overall accuracy of identifying tumors as either UV treated or untreated; *hatched bars*, the accuracy with which UV-treated tumors were classified as UV treated; *grey bars*, accuracy with which untreated tumors were classified as untreated. Note, the 874-features classifier achieved a higher accuracy in the UV-treated cohort alone, whereas the 20-feature classifier achieved a higher accuracy in the non-UV cohort. However, the overall classification accuracy (combined for both cohorts) is similar for both.

Having shown that the array-CGH profiles contain statistically significant information to predict the UV etiology, we asked whether we could identify a subset of CNAs that are most informative at telling the two classes of tumors apart and whether such signature-CNAs can provide clues to the mechanism of UV's melanoma-promoting activity. SVM is a powerful tool for the identification of classifying information among linear combinations of BACs, but it is not designed to test the classification power of individual BACs. Because most loci identified by array-CGH at the current resolution are represented by changes in a single BAC, we investigated an approach that could be applied to individual BACs to identify ones with the greatest classification power.

For this exercise, we chose to focus on regions of gain rather than loss attributable to the greater sensitivity of array-CGH to copy number increase, which can be many-fold, than typical single copy losses in regions of deletion. BACs were scored as "amplified or gained" if the CGH ratio was 4 SDs above the mean log ratio for self-self hybridization controls (>0.35, log base 2). For each BAC, we calculated the difference in the number of tumors that show copy number increase (*i.e.,* "amplification" or "gain") in the UV *versus* non-UV cohorts. Next, we rank ordered all of the informative BACs by the magnitude of such difference and listed the top 20 in Table 2. In other words, the BACs listed are those with the greatest classification power (*i.e.,* most informative in separating the UV and non-UV tumors) and may thus point to gene(s) governing biological differences between the two tumor classes. It is also noteworthy that these top classifying features are found not only in the cohort with significant higher frequency of CNAs (*e.g.,* the spontaneous cohort) but also in the cohort with low frequency of CNAs. This indicates that the primary driving force of this classification is recurrence of specific CNAs, and significant predictors are not dependent on the frequency of CNAs.

To test whether the top 20 prioritized CNAs carry any biological significance, we focused on the region of gain with the greatest representation in our sample set. When sorted by their chromosomal locations, the top 20 list of most informative BACs revealed an overrepresentation of those mapping to chromosome 5 (Table 2). Specifically, 5 of the top 20 BACs localized to chromosome 5, and in total, these 5 BACs were amplified 28 times in 19 UV-induced tumors (indicating that the regions of amplification involved more than one BAC in some tumors; Fig. 2) *versus* only once among the non-UV tumors (Table 2). Moreover, three of these five BACs (D5MIT249, D5MIT102, and D5MIT346) were localized to a contiguous region at the proximal end of chromosome 5 (Table 2 and Fig. 2). Gain of the proximal end of chromosome 5 was apparent even when the log ratio threshold for defining amplification was raised or lowered by 50% (data not shown). It is also worth noting that the same three BACs from proximal chromosome 5 were identified in the list of top 20 generated when BACs were rank ordered based on metrics that account for both amplifications and deletions, including T-score and the related correlation metric defined in Golub *et al.* (data not shown; Ref. 8). From these data, we surmised that gain of a gene or genes residing on proximal chromosome 5 represented an important genetic event in the development of UV-induced melanomas in this model system.

To map more precisely the target gene within this locus, 13 UV-treated tumors were analyzed further by combined array-CGH and quantitative PCR (see "Materials and Methods"). These assays defined the minimal region of gain (*i.e.,* common overlapping region among the 13 tumors) to be 1.5 Mb at the telomeric end of chromosome 5 (Fig. 2). Bioinformatic analysis of the mouse genome se-

Table 2 *List of the top 20 BACs differentially amplified in UV and non-UV melanomas*

BAC ID	Chromosome	Melanoma with gain by array-CGH	
		UV treated	Spontaneous
D1MIT243	1	3	0
D1MIT369	1	5	0
D2MIT167	2	0	3
D5MIT249	5	13	1
D5MIT103	5	12	0
D5MIT346	5	4	0
D5MIT160	5	8	0
D5MIT122	5	4	0
D8MIT258	8	4	0
D8MIT232	8	3	0
D9MIT24	9	0	3
D10MIT218	10	3	0
D10MIT163	10	3	0
D12MIT237	12	0	3
D12MIT54	12	3	0
D12MIT275	12	0	3
D16MIT178	16	3	0
D17MIT113	17	3	0
DXMIT211	X	0	3
DXMIT100	X	3	0

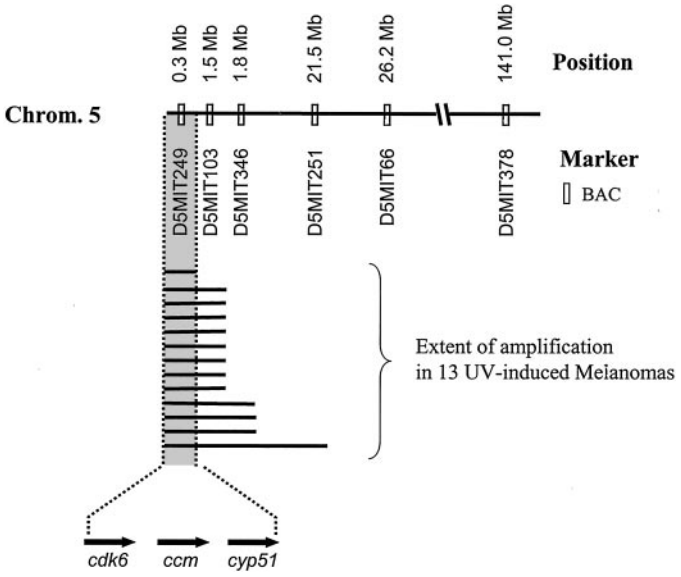

Fig. 2. Mapping of the minimal amplicon on chromosome 5. *Black bars*, the extent of the region of amplification at proximal chromosome 5 for 13 UV-treated tumors that contained amplifications in this region. Names of the markers (*open boxes*) and their positions along chromosome 5 (defined by Ensembl mouse genome database v4.1.1, Jan. 31, 2002) are as indicated.

quence (Ensembl mouse genome database v4.1.1, Jan. 31, 2002) revealed three annotated genes residing within the minimal region, namely, *cdk6*, *cyp51*, and *ccm1*.

The finding of *cdk6* within this locus is remarkable in view of recent reports demonstrating that components of the Rb pathway are critical targets of UV's melanoma-promoting action (26). Indeed, in this $p19^{ARF}$-deficient model, a frequent acquired lesion in non-UV melanomas is loss of $p16^{Ink4a}$ expression or function. In contrast, $p16^{Ink4a}$ expression and function is retained in a greater proportion of the UV-treated tumors, which instead appear to preferentially amplify the *cdk6* locus on chromosome 5 (Fig. 3*A*). Moreover, *cdk6* amplification and $p16^{Ink4a}$ loss were mutually exclusive (Fig. 3*A*; Ref. 26). The c*dk6* gene amplification was confirmed in each tumor that showed proximal chromosome 5 amplification by array-CGH (Fig. 3*B*). *cdk6* amplification was also identified in two tumors in which none of the proximal chromosome 5 array-CGH markers were amplified (tumors 16 and 19; Fig. 3*B*). This more focal amplification strongly suggests that *cdk6* is the target of the amplification event. Moreover, *cdk6* gene amplification is associated with *cdk6* overexpression as determined by quantitative RT-PCR, which showed increased *cdk6* mRNA levels in UV-treated tumors that exhibited *cdk6* amplification by comparison with normal melanocytes and by comparison with non-UV-treated tumors (Fig. 3*C*). In addition, *cdk6* is overexpressed in UV-treated tumors in the absence of amplification of proximal chromosome 5 or the *cdk6* gene, suggesting that UV treatment can enhance cdk6 expression through multiple mechanisms (tumor 9; Fig. 3, *B* and *C*). In contrast, *cdk6* amplification was not detected in tumors arising in $p16^{Ink4a-/-}$ mice spontaneously or as a result of UV exposure (data not shown). In conclusion, the combined use of computational and array-CGH methods in this study has established that genome-wide CNA patterns can be used for tumor classification. Furthermore, prioritization of key features distinguishing between two subclasses of melanoma in this study has revealed *cdk6* as an important target associated with UV exposure. This is consistent with recent reports identifying components of the Rb pathway, including Cdk6, as critical targets of UV's melanoma-promoting activity (26).

More broadly, this study demonstrates that distinct etiologies driving oncogenesis are embedded within the pattern of cytogenetic alterations present in fully established tumors, that such alterations convey prognostic information and that characteristic alterations differentiating tumor classes can be identified through combined array-

A

Tyr-RAS+ $p19^{ARF}$-/- melanoma	With Loss of $p16^{INK4a}$		With retention of $p16^{INK4a}$	
	Total	# with Chr 5 Gain	Total	# with Chr 5 Gain
Non-UV (n=18)	7	0	11	0
UV-induced (n=19)	1	0	18	12

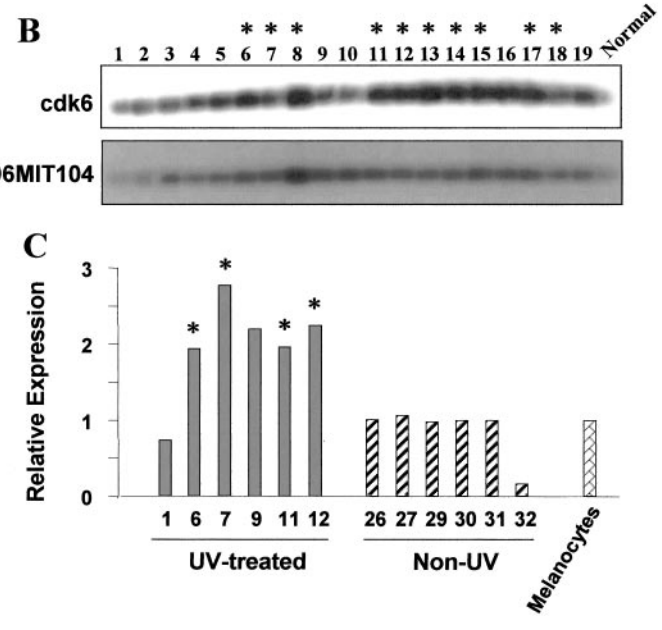

Fig. 3. Amplification of *cdk6* represents the Rb lesion in UV-treated melanomas. *A*, table indicates the inverse correlation between loss of p16 expression or function and amplification of cdk6. Loss of p16 was determined as described previously (26). Amplification of *cdk6* was determined by array-CGH and quantitative PCR. *B*, quantitative PCR illustrating the genomic status of *cdk6* in each tumor. D6MIT104 serves as a loading control because the copy number of D6MIT104 was not changed in any tumor. Tumor identification numbers are indicated above each lane. *Asterisks*, tumor with gain of proximal chromosome 5 by array-CGH. *C*, quantitative RT-PCR illustrates the level of *cdk6* mRNA in the indicated tumors. *Glyceraldehyde-3-phosphate dehydrogenase* was used for loading control. Tumor identification numbers are indicated *below* each lane and correspond to those in *B*. *Asterisks*, tumor with gain of proximal chromosome 5 identified by array-CGH.

CGH and computational means. Although the classification power of array-CGH at present is inferior to that of transcription profiling, the increasing density of BAC arrays used for chromosomal profiling should greatly enhance the predictive power of such data sets. Along this line, it should be noted that, within our sample set of 19 UV-induced melanomas, two of them harbored focal amplifications on Chr 5 involving Cdk6 that were not detectable by array-CGH. Higher resolution arrays with denser coverage of BACs in the genome will likely be able to identify these two lesions, hence increasing the overall accuracy of the classification.

Additionally, there are several distinct advantages to array-CGH over expression profiling. First, the increased stability of DNA over RNA may permit the use of archival or poor quality tumor specimens from which RNA cannot be obtained, as well as samples with high degrees of stromal contamination where laser capture micro-dissection is necessary. In addition, RNA expression profiling is limited in its ability to analyze complex heterogeneous tissues because of the lack of a "normal" control, particularly in cases of tumors where cells of origin are yet to be identified. In contrast, diploid somatic tissue is generally available from cancer patients or normal pool. Finally, array CGH measures information about a tumor not necessarily captured by expression profiling, *e.g.,* cytogenetic complexity, regardless of the specific genetic lesions identified, has been shown to correlate with adverse outcome in both solid and hematological malignancies (5, 30, 31). In summary, through the analysis of array-CGH patterns, we were able to distinguish genetically and histopathologically homogeneous tumors, predict their UV-induced *versus* spontaneous etiologies, and identify novel pathogenic lesions, such as *cdk6* amplification in these tumors.

ACKNOWLEDGMENTS

We thank R. DePinho and M. Meyerson for critical reading of this manuscript. C. W. B. and N. E. S. were supported by grants from the NIH and Howard Hughes Medical Institute.

REFERENCES

1. Anderson, G. R., Stoler, D. L., and Brenner, B. M. Cancer: the evolved consequence of a destabilized genome. Bioessays, *23:* 1037–1046, 2001.
2. Bastian, B. C., LeBoit, P. E., Hamm, H., Brocker, E. B., and Pinkel, D. Chromosomal gains and losses in primary cutaneous melanomas detected by comparative genomic hybridization. Cancer Res., *58:* 2170–2175, 1998.
3. Nelson, M. A., Radmacher, M. D., Simon, R., Aickin, M., Yang, J., Panda, L., Emerson, J., Roe, D., Adair, L., Thompson, F., Bangert, J., Leong, S. P., Taetle, R., Salmon, S., and Trent, J. Chromosome abnormalities in malignant melanoma: clinical significance of nonrandom chromosome abnormalities in 206 cases. Cancer Genet. Cytogenet., *122:* 101–109, 2000.
4. Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., Powell, J. I., Yang, L., Marti, G. E., Moore, T., Hudson, J., Jr., Lu, L., Lewis, D. B., Tibshirani, R., Sherlock, G., Chan, W. C., Greiner, T. C., Weisenburger, D. D., Armitage, J. O., Warnke, R., Staudt, L. M., *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature, *403:* 503–511, 2000.
5. Adeyinka, A., Mertens, F., Idvall, I., Bondeson, L., Ingvar, C., Mitelman, F., and Pandis, N. Different patterns of chromosomal imbalances in metastasising and non-metastasising primary breast carcinomas. Int. J. Cancer, *84:* 370–375, 1999.
6. Bhattacharjee, A., Richards, W. G., Staunton, J., Li, C., Monti, S., Vasa, P., Ladd, C., Beheshti, J., Bueno, R., Gillette, M., Loda, M., Weber, G., Mark, E. J., Lander, E. S., Wong, W., Johnson, B. E., Golub, T. R., Sugarbaker, D. J., and Meyerson, M. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc. Natl. Acad. Sci. USA, *98:* 13790–13795, 2001.
7. DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A., and Trent, J. M. Use of a cDNA microarray to analyse gene expression patterns in human cancer. Nat. Genet., *14:* 457–460, 1996.
8. Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., and Lander, E. S. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science (Wash. DC), *286:* 531–537, 1999.
9. Hedenfalk, I., Duggan, D., Chen, Y., Radmacher, M., Bittner, M., Simon, R., Meltzer, P., Gusterson, B., Esteller, M., Kallioniemi, O. P., Wilfond, B., Borg, A., and Trent, J. Gene-expression profiles in hereditary breast cancer. N. Engl. J. Med., *344:* 539–548, 2001.
10. Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Thorsen, T., Quist, H., Matese, J. C., Brown, P. O., Botstein, D., Eystein Lonning, P., and Borresen-Dale, A. L. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. Sci. USA, *98:* 10869–10874, 2001.
11. Takahashi, M., Rhodes, D. R., Furge, K. A., Kanayama, H., Kagawa, S., Haab, B. B., and Teh, B. T. Gene expression profiling of clear cell renal cell carcinoma: gene identification and prognostic classification. Proc. Natl. Acad. Sci. USA, *98:* 9754–9759, 2001.
12. van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R., and Friend, S. H. Gene expression profiling predicts clinical outcome of breast cancer. Nature (Lond.), *415:* 530–536, 2002.
13. Su, Y. A., Bittner, M. L., Chen, Y., Tao, L., Jiang, Y., Zhang, Y., Stephan, D. A., and Trent, J. M. Identification of tumor-suppressor genes using human melanoma cell lines UACC903, UACC903(+6), and SRS3 by comparison of expression profiles. Mol. Carcinog., *28:* 119–127, 2000.
14. Clark, E. A., Golub, T. R., Lander, E. S., and Hynes, R. O. Genomic analysis of metastasis reveals an essential role for RhoC. Nature (Lond.), *406:* 532–535, 2000.
15. Albertson, D. G., Ylstra, B., Segraves, R., Collins, C., Dairkee, S. H., Kowbel, D., Kuo, W. L., Gray, J. W., and Pinkel, D. Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene. Nat. Genet., *25:* 144–146, 2000.
16. Hui, A. B., Lo, K. W., Teo, P. M., To, K. F., and Huang, D. P. Genome wide detection of oncogene amplifications in nasopharyngeal carcinoma by array based comparative genomic hybridization. Int. J. Oncol., *20:* 467–473, 2002.
17. Hui, A. B., Lo, K. W., Yin, X. L., Poon, W. S., and Ng, H. K. Detection of multiple gene amplifications in glioblastoma multiforme using array-based comparative genomic hybridization. Lab. Investig., *81:* 717–723, 2001.
18. Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W. L., Chen, C., Zhai, Y., Dairkee, S. H., Ljung, B. M., Gray, J. W., and Albertson, D. G. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat. Genet., *20:* 207–211, 1998.
19. Takeo, S., Arai, H., Kusano, N., Harada, T., Furuya, T., Kawauchi, S., Oga, A., Hirano, T., Yoshida, T., Okita, K., and Sasaki, K. Examination of oncogene amplification by genomic DNA microarray in hepatocellular carcinomas: comparison with comparative genomic hybridization analysis. Cancer Genet. Cytogenet., *130:* 127–132, 2001.
20. Wilhelm, M., Veltman, J. A., Olshen, A. B., Jain, A. N., Moore, D. H., Presti, J. C., Jr., Kovacs, G., and Waldman, F. M. Array-based comparative genomic hybridization for the differential diagnosis of renal cell cancer. Cancer Res., *62:* 957–960, 2002.
21. Hodgson, G., Hager, J. H., Volik, S., Hariono, S., Wernick, M., Moore, D., Nowak, N., Albertson, D. G., Pinkel, D., Collins, C., Hanahan, D., and Gray, J. W. Genome scanning with array CGH delineates regional alterations in mouse islet carcinomas. Nat. Genet., *29:* 459–464, 2001.
22. Sharpless, N. E., Ferguson, D. O., O'Hagan, R. C., Castrillon, D. H., Lee, C., Farazi, P. A., Alson, S., Fleming, J., Morton, C. C., Frank, K., Chin, L., Alt, F. W., and DePinho, R. A. Impaired nonhomologous end-joining provokes soft tissue sarcomas harboring chromosomal translocations, amplifications, and deletions. Mol. Cell, *8:* 1187–1196, 2001.
23. Snijders, A. M., Nowak, N., Segraves, R., Blackwood, S., Brown, N., Conroy, J., Hamilton, G., Hindle, A. K., Huey, B., Kimura, K., Law, S., Myambo, K., Palmer, J., Ylstra, B., Yue, J. P., Gray, J. W., Jain, A. N., Pinkel, D., and Albertson, D. G. Assembly of microarrays for genome-wide measurement of DNA copy number. Nat. Genet., *29:* 263–264, 2001.
24. You, M. J., Castrillon, D. H., Bastian, B. C., O'Hagan, R. C., Bosenberg, M. W., Parsons, R., Chin, L., and DePinho, R. A. Genetic analysis of Pten and Ink4a/Arf interactions in the suppression of tumorigenesis in mice. Proc. Natl. Acad. Sci. USA, *99:* 1455–1460, 2002.
25. Chin, L., Pomerantz, J., Polsky, D., Jacobson, M., Cohen, C., Cordon-Cardo, C., Horner, J. W., II, and DePinho, R. A. Cooperative effects of INK4a and ras in melanoma susceptibility in vivo. Genes Dev., *11:* 2822–2834, 1997.
26. Kannan, K., Sharpless, N. E., Xu, J., O'Hagan, R. C., Bosenberg, M., and Chin, L. Components of the Rb pathway are critical targets of UV mutagenesis in a murine melanoma model. Proc. Natl. Acad. Sci. USA, *21:* 1221–1225, 2003.
27. Cortes, C., and Vapnik, V. Support-vector networks. Machine Learning, *20:* 273–297, 1995.
28. Vapnik, V. N. (ed.). Statistical Learning Theory. New York: Wiley, 1998.
29. Collobert, R., and Benigo, S. SVMTorch: support vector machines for large-scale regression problems. J. Machine Learn. Res., *1:* 143–160, 2001.
30. Han, T., Ozer, H., Sadamori, N., Emrich, L., Gomez, G. A., Henderson, E. S., Bloom, M. L., and Sandberg, A. A. Prognostic importance of cytogenetic abnormalities in patients with chronic lymphocytic leukemia. N. Engl. J. Med., *310:* 288–292, 1984.
31. Golomb, H. M., and Rowley, J. D. Significance of cytogenetic abnormalities in acute leukemias. Hum. Pathol., *12:* 515–522, 1981.