Oncogene (2004) 23, 4060–4067

www.nature.com/onc

Microarray expression profiling in melanoma reveals a *BRAF* mutation signature

Sandra Pavey[1], Peter Johansson[2], Leisl Packer[1], Jennifer Taylor[3], Mitchell Stark[1], Pamela M Pollock[4], Graeme J Walker[1], Glen M Boyle[1], Ursula Harper[4], Sarah-Jane Cozzi[1], Katherine Hansen[4], Laura Yudt[4], Chris Schmidt[1], Peter Hersey[5], Kay AO Ellem[1], Michael GE O'Rourke[6], Peter G Parsons[1], Paul Meltzer[4], Markus Ringnér[2] and Nicholas K Hayward*,[1]

[1]*Queensland Institute of Medical Research, 300 Herston Rd, Herston, Queensland 4006, Australia;* [2]*Department of Theoretical Physics, Complex Systems Division, Lund University, Sölvegatan 14A, Lund SE-223 62, Sweden;* [3]*Queensland Centre for Schizophrenia Research, The Park-Centre for Mental Health, Wolston Park Rd, Wacol, Queensland 4076, Australia;* [4]*Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, 50 South Drive, Building 50 Rm 5139, Bethesda, MD 20892, USA;* [5]*University of Newcastle, David Maddison Building, Cnr King and Watt Sts, Newcastle, New South Wales 2300, Australia;* [6]*Mater Misericordiae Hospital, Raymond Tce, South Brisbane, Queensland 4101, Australia*

We have used microarray gene expression profiling and machine learning to predict the presence of *BRAF* mutations in a panel of 61 melanoma cell lines. The *BRAF* gene was found to be mutated in 42 samples (69%) and intragenic mutations of the *NRAS* gene were detected in seven samples (11%). No cell line carried mutations of both genes. Using support vector machines, we have built a classifier that differentiates between melanoma cell lines based on *BRAF* mutation status. As few as 83 genes are able to discriminate between *BRAF* mutant and *BRAF* wild-type samples with clear separation observed using hierarchical clustering. Multidimensional scaling was used to visualize the relationship between a *BRAF* mutation signature and that of a generalized mitogen-activated protein kinase (MAPK) activation (either *BRAF* or *NRAS* mutation) in the context of the discriminating gene list. We observed that samples carrying *NRAS* mutations lie somewhere between those with or without *BRAF* mutations. These observations suggest that there are gene-specific mutation signals in addition to a common MAPK activation that result from the pleiotropic effects of either *BRAF* or *NRAS* on other signaling pathways, leading to measurably different transcriptional changes.
Oncogene (2004) **23,** 4060–4067. doi:10.1038/sj.onc.1207563
Published online 29 March 2004

Keywords: *BRAF*; melanoma; microarray; mitogen-activated protein kinase; mutation

Introduction

Constitutive activation of the receptor tyrosine kinase (RTK)/Ras/Raf/mitogen-activated protein kinase (MAPK) pathway is a frequent and early event in melanoma development (Cohen *et al*., 2002; Satyamoorthy *et al*., 2003). Recently, mutation of *BRAF* (v-raf murine sarcoma viral oncogene homolog B1) has been shown to be the primary mechanism by which this activation occurs (Davies *et al*., 2002). *BRAF* mutation is arguably the most critical step in the initiation of melanocytic neoplasia, but is insufficient to confer the malignant potential since mutations occur as often in benign melanocytic nevi as in invasive cutaneous melanomas (Pollock *et al*., 2003). Somatic *BRAF* mutations occur in 41–88% of melanomas and nevi (Brose *et al*., 2002; Davies *et al*., 2002; Dong *et al*., 2003; Gorden *et al*., 2003; Pollock *et al*., 2003; Satyamoorthy *et al*., 2003) and in a variety of other tumor types, including 36–69% of papillary thyroid cancers (Cohen *et al*., 2003; Fukushima *et al*., 2003; Kimura *et al*., 2003), 5–18% of colorectal carcinomas (Davies *et al*., 2002; Rajagopalan *et al*., 2002; Yuen *et al*., 2002) and 2–3% of lung cancers (Brose *et al*., 2002; Davies *et al*., 2002; Naoki *et al*., 2002; Cohen *et al*., 2003). All documented mutations to date have been found in the kinase domain of B-Raf, encoded by exons 11 and 15 of the *BRAF* gene (Brose *et al*., 2002; Davies *et al*., 2002; Naoki *et al*., 2002; Yuen *et al*., 2002). The majority of these mutations affect one critical amino acid, resulting in a valine to glutamic acid substitution at residue 599. The V599E substitution is thought to lead to constitutive kinase activity of B-Raf, potentially by mimicking the phosphorylation of the T598 and S601 residues that occurs during the normal activation of the kinase (Davies *et al*., 2002).

In some melanomas without *BRAF* mutation, the MAPK pathway is constitutively activated through mutation of *NRAS* (neuroblastoma RAS viral (v-ras) oncogene homolog) (van Elsas *et al*., 1996). *BRAF* and *NRAS* mutations appear to have the same effect in melanoma development since their occurrence in the same tumor is mutually exclusive (Cohen *et al*., 2002; Davies *et al*., 2002; Pollock *et al*., 2003; Satyamoorthy

*Correspondence: NK Hayward; E-mail: nickH@qimr.edu.au
Received 23 October 2003; revised 18 December 2003; accepted 22 January 2004; Published online 29 March 2004

et al., 2003). A similar situation has also been observed in thyroid (Kimura *et al*., 2003), lung (Brose *et al*., 2002; Davies *et al*., 2002; Naoki *et al*., 2002) and colon cancers (Davies *et al*., 2002; Rajagopalan *et al*., 2002; Yuen *et al*., 2002), where *BRAF* and *RAS* mutations are seldom found in the same tumor. In the few exceptional colon (Davies *et al*., 2002; Yuen *et al*., 2002) and lung cancers (Brose *et al*., 2002; Davies *et al*., 2002) in which both *BRAF* and *RAS* mutations occur, the mutations in *BRAF* never include the V599E change (Davies *et al*., 2002; Yuen *et al*., 2002), indicating that substitutions elsewhere in B-Raf may not have the same potency in activating the MAPK pathway.

Recently, microarray gene expression profiling has been used to develop a number of phenotypic models that predict the activity of various oncogenic signaling pathways, including those emanating from the activation of Ha-ras, c-*myc* and members of the E2F family of transcription factors (Huang *et al*., 2003). The models were extremely accurate in assigning the activation status of various oncogenic pathways after the infection of murine embryonic fibroblasts with oncogene-expressing adenoviruses. Similar discrimination was seen between mammary tumors that arose in mice carrying either *MYC* or *HRAS* transgenes driven by the MMTV promoter. These findings indicate that oncogene activation can lead to highly specific and lasting gene expression changes.

Supervised analysis methods are very powerful for classification and prediction of cancer gene expression profiles into predefined classes (Golub *et al*., 1999; Simon *et al*., 2003). In these methods, expression data from cancer samples, together with knowledge about which class each sample belongs to, are used to construct a classifier (prediction rule). The accuracy of the classifier is evaluated on independent samples that were neither used to select genes to include in the classifier nor to construct the prediction rule. Recently, supervised machine learning methods such as artificial neural networks and support vector machines (SVMs) have been used to classify cancer expression profiles (Furey *et al*., 2000; Khan *et al*., 2001). Here, we have used expression profiling and SVM learning as a tool to predict the presence of *BRAF* activating mutations in a panel of melanoma cell lines.

Results

Mutation data

Mutation status of *BRAF* and *NRAS* was determined for each cell line (see Supporting Table 2 in Supplementary Material at the following URL: http://www.qimr.edu.au/research/labs/nickh/Pavey-et-al-Supporting-Information.pdf). The following mutations were detected:

BRAF

Four amino-acid substitutions were detected in exon 15 and none were observed in exon 11. At nucleotide positions 1786 and 1787, a transition of a C>T and a T>C, respectively, led to a substitution at codon 596 (L596S). At nucleotide position 1786, a transversion of a C>G led to a substitution at codon 596 (L596V). At nucleotide positions 1795 and 1796, a transition of a G>A and transversion of a T>A, respectively, led to a substitution at codon 599 (V599K). At nucleotide position 1796, a transversion of T>A led to a substitution at codon 599 (V599E). In the panel of 61 cell lines, L596S, L596V and V599K each occurred once (1.6%). The V599E mutation occurred at a frequency of 69% (42/61).

NRAS

Two amino-acid substitutions were detected in exon 1. At nucleotide position 34, a transition of a G>A led to a substitution at codon 12 (G12S) and at nucleotide position 37, a transversion of a G>C led to substitution at codon 13 (G13R). Both mutations occurred once (1.6%). In exon 2, three amino-acid substitutions were detected affecting codon 61. At nucleotide position 181, a transversion of a C>A led to a Q61K substitution. At nucleotide position 182, a transversion of an A>T and a transition of an A>G led to Q61L and Q61R substitutions, respectively. Q61K, Q61L and Q61R mutations occurred at frequencies of 1.6% (1/61), 6.5% (4/61) and 3.3% (2/61), respectively. In one cell line, MM649, *NRAS* was homozygously deleted.

Supervised gene selection

The first pass of analysis used a supervised approach, based on a nonparametric method to determine differential gene expression between samples with *BRAF* or *NRAS* mutations and wild-type samples. We used all 61 cell lines in each analysis. Using the Mann–Whitney *U*-test, we expect 50 of the 5041 filtered clones to have a *P*-value of less than 0.01 by chance. The *BRAF* mutant versus *BRAF* wild-type supervised analysis yielded 135 clones from the filtered list with $P<0.01$ (see Supporting Table 3), and the *NRAS* mutant versus *NRAS* wild-type analysis yielded 48 clones (see Supporting Table 4). The overlap between these *BRAF* and *NRAS* lists was 19 clones (see Supporting Table 4). The combined genotype of having either *BRAF* or *NRAS* activating mutations versus wild type for both *NRAS* and *BRAF* yielded 37 clones at $P<0.01$.

Supervised classification

We used the receiver operating characteristic (ROC) curve area to measure the prediction performance on samples not used to train the classifier. For the SVM committee that discriminates cell lines according to *BRAF* mutation status, we got an area of 82% (Supporting Figure 1). Regardless of the number of samples in each class, a random classifier will on average result in an area of 50% (ideally the area is 100%). When we performed the same analysis with randomly permuted sample labels, we got better or equal

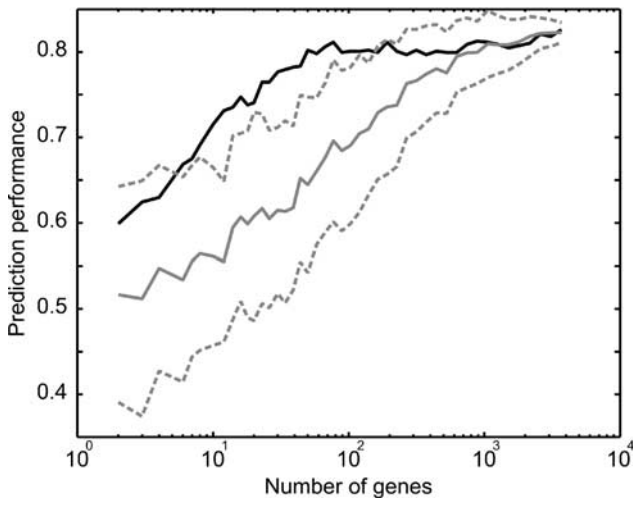

Figure 1 Prediction performance from SVM classification of *BRAF* mutation status. SVM prediction performance, as measured by the ROC area, of *BRAF* status using varying numbers of top-ranked genes as input to the SVMs. Black curve – ROC area as a function of the number of top-ranked genes used; gray curve – results obtained when selecting the same number of genes randomly from the filtered data set; dotted gray curves – one standard deviation from the average random result

performance only 29 times out of 10 000 replicates ($P = 0.0029$), which strongly suggests that there was no overfitting in our classification procedure. Hence, there is a strong correlation between gene expression profiles and *BRAF* status that can be used to predict significantly the status of samples not used to train the classifier.

Next, we ranked the genes (refer to Supporting Table 3) and built a new classifier using only the N top-ranked genes (see Materials and methods). In addition, we built a classifier based on N randomly selected genes. Performing this for different values of N, we got a significantly better performance using genes from the ranking than from random selections (Figure 1). The difference was most significant when we used a small number of genes. This conclusion is expected since choosing more random genes increases the number of selected top-ranked genes. Hence, it is probable that we had some overlap between the 100 genes selected by random and the 100 top-ranked genes. Using the top 80 genes, we get a performance of similar quality as when using all the genes. Thus, to get a list of *BRAF* discriminatory genes, we selected genes that were ranked in the top 80 by at least 25% of the SVMs, which resulted in a total of 83 genes (Figure 2 and Supporting Table 5).

Hierarchical clustering using these 83 *BRAF* discriminatory genes was performed in both the sample and clone dimensions (Figure 2). This provided clear clustering of the cell lines carrying *BRAF* activating mutations. The relationships of the genotype classes are further illustrated in a multidimensional scaling (MDS) visualization (Figure 3). This plot again demonstrated clear discrimination between the samples carrying *BRAF* activating mutations to samples wild type for *BRAF*, while allowing observation of the samples carrying the *NRAS* mutation as lying somewhere between the *BRAF* wild-type and the *BRAF* mutated samples.

Quantitative RT–PCR (qRT–PCR)

To assess the reliability of the array hybridization results, transcript levels of nine differentially expressed genes were measured using qRT–PCR analysis. Intra- and interassay variation was 2.9 and 6.8%, respectively, and the qRT–PCR duplicate assays had a coefficient of variation less than 0.05. The concordance between the qRT–PCR and the microarray expression levels was determined (Figure 4) for each of the genes validated (see Supporting Tables 6a–i for raw data) as follows: for each gene expression ratios determined by microarray analysis and qRT–PCR were grouped into three 'bins', defined as upregulated genes (>2.0-fold expression ratio), genes with equal expression (within 0.5- to 2.0-fold) and downregulated genes (<0.5-fold). When microarray and qRT–PCR expression ratios were in the same 'bin', the methods were regarded as concordant. The two methods were highly concordant, with an average of 75% concordance between genes upregulated in association with a *BRAF* mutation, and 79% for genes downregulated in *BRAF* mutant samples. The lack of concordance between a small proportion of the samples may be due to a number of possible factors, including minor divergence between replicate spots on the microarray, variation in distribution and intensity of pixels within each spot or lack of dynamic range across expression levels in microarray data in comparison to the qRT–PCR expression range. Fold changes in transcript levels were generally more compressed using microarrays, in agreement with previous reports (Rajeevan *et al*., 2001; Chuaqui *et al*., 2002).

Discussion

Mutation status of *BRAF* and *NRAS* was determined for 61 melanoma cell lines. *BRAF* mutations were detected in 44 samples (72%). All mutations occurred in exon 15 and all but three resulted in a V599E substitution. *NRAS* activating mutations were found in nine samples and another cell line had a homozygous deletion of this gene. No cell line with a *BRAF* mutation also carried an intragenic mutation of *NRAS*, in keeping with previous reports that have found *RAS* and *BRAF* mutations to be almost mutually exclusive in a variety of cancer types (Brose *et al*., 2002; Cohen *et al*., 2002; Davies *et al*., 2002; Naoki *et al*., 2002; Yuen *et al*., 2002; Kimura *et al*., 2003; Pollock *et al*., 2003; Satyamoorthy *et al*., 2003).

Using SVMs, we have built a classifier that based on gene expression profiles discriminates between melanoma cell lines according to whether they carry mutations in *BRAF*. As few as 83 genes are able to discriminate between *BRAF* mutant and *BRAF* wild-type cell lines.

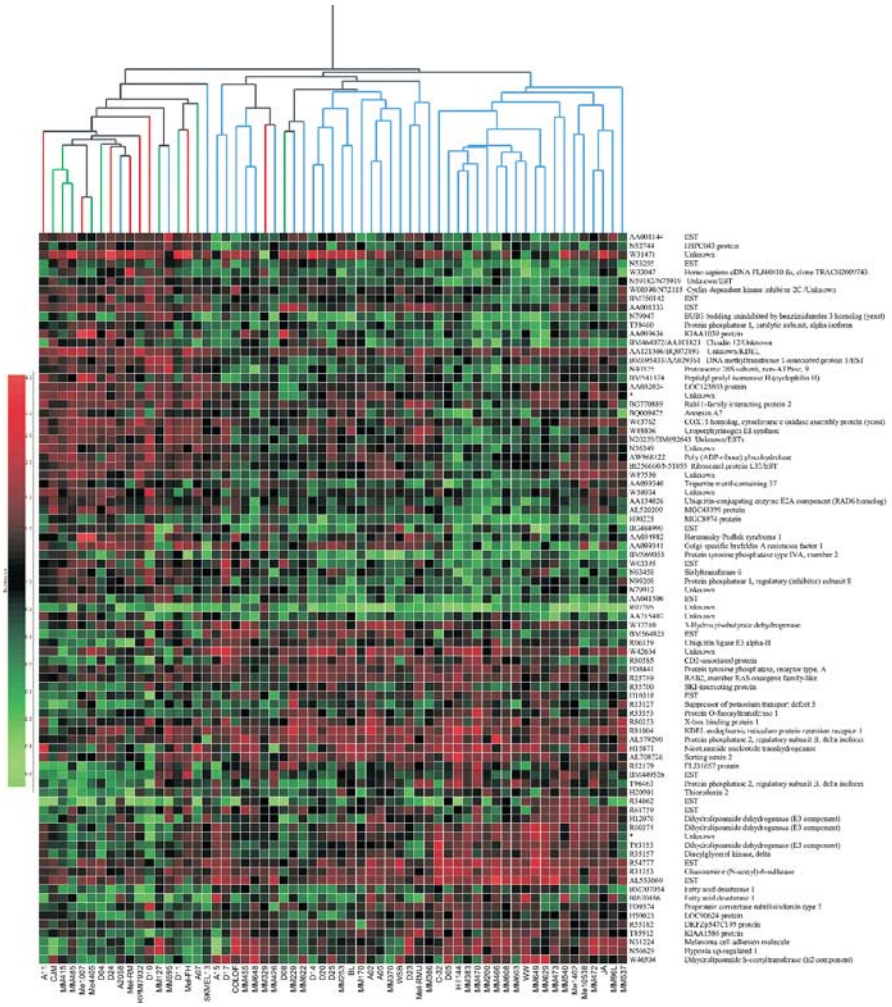

Figure 2 Hierarchical clustering of 61 melanoma cell lines and genes, using the *BRAF* discriminatory genes ($n = 83$). Spearman's correlation was used to cluster samples and genes based on centralized data. Expression ratios (see color scale bar) used to color the dendrogram were derived from normalized values. Branches pertaining to individual cell lines carrying a *BRAF* mutation are colored blue (*BRAF* mutant/*NRAS* wild type), cell lines carrying an *NRAS* mutation colored green (*NRAS* mutant/*BRAF* wild type) and cell lines that are wild type at both loci colored red. Asterisks in the GenBank accession column refer to probes where no data were obtained during sequence validation by the array manufacturer (refer to http://www.microarray.ca/support/glists.html)

Hierarchical clustering using these discriminatory genes gives good separation of the samples (Figure 2). Initially, we considered there might be a common MAPK activation signature (resulting from either BRAF or NRAS mutation); however, we found no overabundance of discriminatory genes for the combined group of samples having either *BRAF* or *NRAS* mutations. Furthermore, we built SVMs for discriminating samples with mutation of either *BRAF* or *NRAS* from samples being wild type for both *BRAF* and *NRAS*, and obtained results comparable to random predictions. Moreover, using MDS, we found clear separation of *BRAF* mutant samples and samples wild type for both *BRAF*/*NRAS*, and observed a tendency that *NRAS* mutant samples generally clustered between these two groups. This observation suggests that there may be some genes that specifically discriminate between the three genotypic classes, but more samples are required to establish a specific *NRAS* mutation signature. Nonetheless, our findings suggest that the transcriptional consequences resulting from mutation of *BRAF* or *NRAS* are different, presumably through their differential capacity to receive input signals and transduce them through various effectors. Indeed, since all cell lines were grown in the presence of serum at the time of RNA extraction (hence the MAPK pathway would be expected to be constitutively activated in every line), the genes that discriminate *BRAF* or *NRAS* mutant cells are independent of this common MAPK activation. This notion implies that some of the genes on the *BRAF* discriminating gene list may not necessarily be the direct targets of the transcription factors (e.g. Elk-1) that are ultimately activated by MAPKs. This hypothesis has important ramifications for the development on new melanoma treatments, as it would open up the possibility of identifying novel therapeutic targets outside of the MAPK pathway that could be used to treat melanomas carrying *BRAF* mutations. In a highly

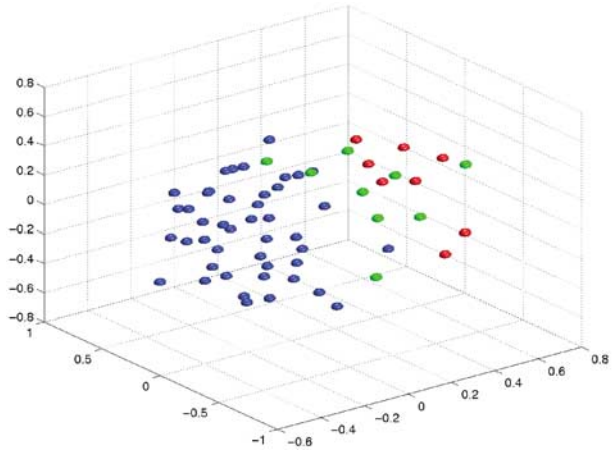

Figure 3 MDS plot using the *BRAF* discriminatory genes ($n = 83$). Each spot represents an individual sample, with cell lines carrying a *BRAF* mutation colored blue (*BRAF* mutant/*NRAS* wild type), cell lines carrying an *NRAS* mutation colored green (*NRAS* mutant/*BRAF* wild type) and cell lines that are wild type at both loci colored red

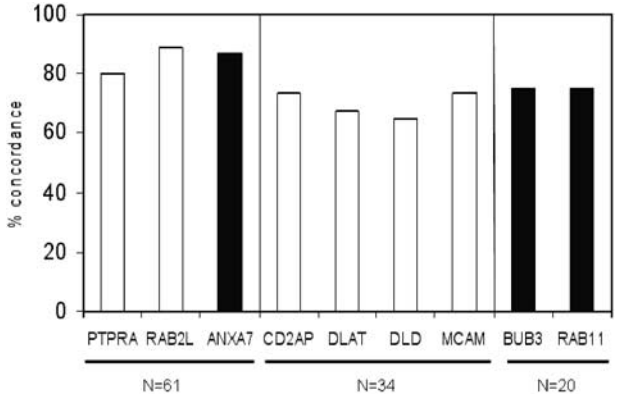

Figure 4 Concordance between gene expression levels measured by microarrays and qRT–PCR. Gene expression levels obtained using microarrays were confirmed by qRT–PCR for nine different transcripts. Concordance was deemed to occur if the gene expression ratios (relative to the reference sample) were assessed to be within the same 'bin', namely, upregulated (>2.0-fold upregulated); roughly equal (within 0.5- to 2.0-fold); or downregulated (<0.5-fold), by both methods. Samples for which expression ratios fell into separate bins for each method were regarded to be nonconcordant. The percentage of samples concordant between the two methods are shown for each gene. Genes with a higher average expression in *BRAF* mutant samples compared to wild-type samples are denoted by open bars and genes that have a lower average expression in *BRAF* mutant samples compared to wild-type cell lines are shown as solid bars

analogous situation to that we have described here, Huang *et al*. (2003) built expression models to predict the activation status of E2F1, E2F2 and E2F3, and showed that the models could readily discriminate between the activation of these three very closely related transcription factors.

Of the 83 *BRAF* discriminatory genes, 42 have known function and the remainder encode hypothetical proteins or are simply ESTs. Notably, five of the known genes encode phosphatases, enzymes with key functions in regulating signal transduction pathways. PTPRA for example, is a member of the protein tyrosine phosphatase (PTP) family involved in regulating cell cycle transition from G2 phase to mitosis and has been shown to dephosphorylate and activate Src family tyrosine kinases (Mustelin and Hunter, 2002). PTPRA has also been implicated in the regulation of integrin signaling, cell adhesion and proliferation (Zheng *et al*., 1992; Harder *et al*., 1998; Zheng and Shalloway, 2001). The higher expression levels seen in *BRAF* mutant samples supports a role for PTPRA in melanoma cell proliferation.

While space prohibits the discussion of all named genes on the discriminating list, brief summaries of few key genes that have biological relevance to melanoma follow. ANXA7 is a member of the annexin family of Ca^{2+}-dependent phospholipid-binding proteins and has a postulated role in suppressing prostate cancer (Srivastava *et al*., 2001). We found reduced *ANXA7* mRNA expression in *BRAF* mutant samples, supporting a similar tumor suppressor role for *ANXA7* in melanoma. The related ANX1 and ANX6 have also been assigned tumor suppressor roles in other cancer models (Bastian, 1997), including a loss of ANX6 expression during progression from benign to malignant melanoma (Francia *et al*., 1996).

The gene encoding melanoma cell adhesion molecule (MCAM/MUC18/CD146) was found to be expressed at higher levels in *BRAF* mutant samples. MCAM functions as a Ca^{2+}-independent cell adhesion molecule involved in homotypic and heterotypic adhesion between melanoma cells and endothelial cells, respectively (Johnson *et al*., 1997; Shih *et al*., 1997). Our data are consistent with higher expression of MCAM being associated with increased tumor growth and metastatic potential of melanoma cells (Luca *et al*., 1993; Xie *et al*., 1997).

The SKI protein has been implicated as a key regulator of melanoma tumor progression (Medrano, 2003). Ski-interacting protein (SKIP), together with SKI, interacts with pRb, resulting in the repression of pRb-induced cell cycle arrest (Prathapam *et al*., 2002). We found generally increased SKIP expression in *BRAF* mutant samples, suggesting the possibility of abrogated pRb activity with concomitant cell cycle progression in *BRAF* mutant melanomas.

The genes encoding the E2 (DLAT) and E3 (DLD) components of pyruvate dehydrogenase were in the top 83 ranked discriminating genes. Both genes showed increased mRNA expression in *BRAF* mutant samples, which may reflect altered energy production in melanoma cells carrying these mutations.

A number of microarray studies in various types of cancer have identified gene expression patterns indicative of the mutational activation of oncogenic pathways or inactivation of tumor suppressor pathways. Typical examples include signatures underlying germline *BRCA1* or *BRCA2* mutations in breast (Hedenfalk *et al*., 2001) and ovarian cancer (Jazaeri *et al*., 2002), as

well as somatic mutations of *TP53* in breast cancer (Sorlie *et al*., 2001), and a variety of mutations/translocations in T-cell acute lymphoblastic (Ferrando *et al*., 2002) or acute myeloid leukemia (Schoch *et al*., 2002). The work we have presented in this study has led to the identification of an expression signature that predicts *BRAF* mutation status in melanoma. While this finding points to underlying structure in global gene expression profiles, it is only through further analysis of the individual genes that discriminate between mutated and wild-type samples that we may hope to better understand the molecular events controlling melanoma development. Importantly, some of the genes on the *BRAF* discriminating gene list may prove to encode useful new therapeutic targets to treat melanomas carrying *BRAF* mutations.

Materials and methods

Cell culture and RNA extraction

A panel of 61 melanoma cell lines derived from cutaneous melanomas or nodal metastases were used. Of these, 38 cell lines have been described previously (Castellano *et al*., 1997). Of the remaining lines, the series A2–A15 and D4–D25 were established by Dr Christopher Schmidt, Professor Kay Ellem, Professor Michael O'Rourke and co-workers; ME1007, ME1402, ME4405, ME10538, Mel-FH, Mel-RM and Mel-RMU were established by Professor Peter Hersey and co-workers, and MM470, MM537 and MM629 were established by Dr Peter Parsons and co-workers. All cell lines were cultured in RPMI1640 in the presence of 10% fetal bovine serum from the same batch. Total RNA was extracted using Qiagen RNeasy Midi-kits from cells in log phase growth at 70% confluency lysed directly on the plate. Cell lysates were stored at -70°C until extraction, which was carried out as per the manufacturer's instructions (further information is available as Supporting Text 1).

Genotyping of cell lines

Since all *BRAF* mutations to date have been reported to occur in exons 11 and 15 (Brose *et al*., 2002; Davies *et al*., 2002; Naoki *et al*., 2002; Yuen *et al*., 2002), each line was screened for variants in these exons by PCR sequencing. *NRAS* was also screened for mutations in codons 12, 13 and 61, which have been found previously to activate the potential of *NRAS* to transform cultured cells (Schleger *et al*., 2000) and have been found in a variety of human tumors including melanomas (van Elsas *et al*., 1996). For further information refer to Supporting Text 2.

Microarray probe preparation, hybridization and scanning

Each sample was cohybridized on the arrays together with that of a common reference cell line, MM329, derived from a primary melanoma and which is wild type for *BRAF*, *NRAS* and *CDKN2A*. Probes were prepared using 40 μg of RNA for test samples and 50 μg of reference RNA. RNA was reverse transcribed into fluorescently labeled cDNA by direct dye incorporation, using Cy5-dUTP in the test samples, and Cy3-dUTP in the reference. Each sample was hybridized to commercially available cDNA arrays printed on glass slides by the Microarray Centre, University Health Network, Ontario, Canada (http://www.microarrays.ca). The slides were Human 19K Arrays (v2.0) containing 19 008 human ESTs, derived by PCR amplification of inserts, representing 18 107 separate cDNAs spotted in duplicate across two slides. Details of clone identity and sequence verification are available at http://www.microarray.ca/support/glists.html. Hybridization was carried out at 42°C for 16–18 h, and the slides were washed according to the manufacturer's protocol. The chips were scanned by a GMS418 confocal scanner (Affymetrix/GeneticMicrosystems) with SoftMax Pro software to obtain raw images. Refer to Supporting Text 3 for further information.

Microarray data analysis

Expression profiles from the 61 cell lines were used in each analysis. Raw images were imported into ImaGene v4.2 (BioDiscovery), and mean pixel intensities were extracted and spots with poor/absent signal were flagged. For each clone, the logarithm of the ratio between the intensity in the sample (red) channel and the reference (green) channel was averaged over the duplicates and used as the expression value for the clone. As saturated and low-intensity data tend to be noise dominated, we used quality control criteria that required clones to have all four intensities (red and green for both duplicates) between 50 and 64 000 fluorescence units. Of 19 200 clones in duplicate, 5041 survived this filter across all 61 samples. The data were centralized sample by sample such that the average expression value for a sample was zero. MDS analysis was performed as described by Khan *et al*. (2001) in three dimensions using Euclidean distance measures. Hierarchical clustering was performed on data centralized such that the average expression for each gene was zero using GeneSpring v5.0 (Silicon Genetics, Redwood City, CA, USA) with default settings. Data analysis incorporating mutation status and expression data from each cell line were undertaken by supervised analysis methods (outlined below).

Supervised gene selection

The filtered set of clones was investigated for clones that displayed statistically significant differences between the two *BRAF* genotype groups (*BRAF* wild type and *BRAF* mutant) and the two *NRAS* genotype groups (*NRAS* wild type and *NRAS* mutant). For this purpose, a supervised approach using the Mann–Whitney *U*-statistic was used to generate a list of clones that satisfied statistical significance between genotype groups to a *P*-value of less than 0.01. The Mann–Whitney *U*-statistic has been demonstrated to be robust and conservative (low Type I error) in its application to the identification of discriminatory genes from expression data (Troyanskaya *et al*., 2002).

Supervised classification

We used linear maximal-margin SVMs (Cristianini and Shawe-Taylor, 2000) to classify the samples according to mutational status. SVMs were trained in a threefold crossvalidation scheme, in which samples were randomly split into three groups, and two groups were used for training an SVM and the remaining group was used for validation. This was repeated three times such that each group (and consequently each sample) was used for validation once. A committee of SVMs was created by repeating this entire procedure 10 times. Hence, for each sample there were 10 SVMs for which the sample was not used in the training. The average of the outputs from these 10 SVMs was used as prediction output for the sample.

We used the ROC curve area (Hanley and McNeil, 1982) to measure the prediction performance of the SVM committee.

As we used linear maximal-margin SVMs that have no user-tunable parameters, the risk of overfitting in our cross-validation procedure was small. Nevertheless, in order to rule out overfitting and to validate the significance of the performance of the committee, we performed a random permutation test. We randomly relabeled the samples keeping the class proportions, and with these new labels performed the full crossvalidation procedure described above. This was carried out for 10 000 random sample labelings and an empirical probability distribution of the ROC curve area with random labels was generated. Using this probability distribution, the actual ROC area was assigned a *P*-value corresponding to the probability to obtain this prediction performance or better under the null hypothesis of gene expression patterns randomly associated with the classes.

Next, we ranked the genes using the Mann–Whitney statistic and investigated how many genes were needed to get good performance. For each SVM, genes were ranked based on a Mann–Whitney test applied only to the subset of samples used when training the SVM. Since we have a total of 30 SVMs, this results in 30 ranks assigned to each gene, one for each SVM. To achieve a consensus gene ranking, we used the 25th percentile of these 30 ranks.

To check the significance of the gene ranking the cross-validation procedure was redone using only the top *N* genes from the rankings. Here, we used the individual gene ranking for each of the 30 SVMs. Thus, the validation samples were not used in the selection of genes to use in the training and there was no information leak. We did this classification for different numbers of top-ranked genes in steps from using only one gene to using all 5041 genes. In addition, we checked the performance of the crossvalidation when we randomly selected *N* genes for each SVM committee. For each *N* we did this random selection 100 times.

Quantitative RT–PCR

To further confirm the validity of the microarray expression data, the mRNA levels of nine unique transcripts selected from the 83 highest ranking genes from the SVM consensus gene list were assessed by qRT–PCR. Selections were based on the potential roles of the genes in melanocyte biology, the MAPK pathway or cell cycle regulation (see Supporting Table 1). To obtain an appropriate control, we looked for genes that showed minimal variation across the reference and control channels, that is, within 0.7- to 1.4-fold of the reference value in all test samples. Only eight ESTs satisfied this criterion. Of these, two encoded GAPDH, a common historical control in RT–PCR experiments. The reference cell line MM329 was used to establish the qRT–PCR efficiencies of each gene (Pfaffl, 2001). Briefly, the same RNA samples extracted for the microarray experiments were used in the qRT–PCR experiments. cDNA was made using Superscript III reverse transcriptase (Invitrogen). Subsequent PCR reactions were carried out on a Corbett RotorGene 3000 (Corbett Research, Australia) using a QuantiTect SYBR® Green PCR kit (Qiagen, Germany). Test cell lines and the reference cell line were amplified in parallel reactions using specific primers (for primer sequences, see Supporting Table 1 and for qRT–PCR conditions see Supporting Text 4). To confirm the accuracy and reproducibility of qRT—PCR, the intra-assay precision was determined in 10 repeats within one run. Interassay variation was investigated in 10 different experimental runs. Specificity of PCR products obtained was characterized by melting curve analysis. Gel electrophoresis and DNA sequencing was carried out on PCR products for each primer set to confirm identity.

Acknowledgements

We thank Patrik Edén and Javed Khan for valuable assistance, and Cathy Davern and Michelle Down for culturing some of the melanoma cell lines. We also thank Yidong Chen, NHGRI, for access to the software used to generate the MDS figure. This work was supported by the National Health and Medical Research Council of Australia Grant Number 199600, the Swedish Research Council and the Knut and Alice Wallenberg Foundation through the Swegene consortium.

References

Bastian BC. (1997). *Cell Mol. Life Sci.*, **53,** 554–556.

Brose MS, Volpe P, Feldman M, Kumar M, Rishi I, Gerrero R, Einhorn E, Herlyn M, Minna J, Nicholson A, Roth JA, Albelda SM, Davies H, Cox C, Brignell G, Stephens P, Futreal PA, Wooster R, Stratton MR and Weber BL. (2002). *Cancer Res.*, **62,** 6997–7000.

Castellano M, Pollock PM, Walters MK, Sparrow LE, Down LM, Gabrielli BG, Parsons PG and Hayward NK. (1997). *Cancer Res.*, **57,** 4868–4875.

Chuaqui RF, Bonner RF, Best CJ, Gillespie JW, Flaig MJ, Hewitt SM, Phillips JL, Krizman DB, Tangrea MA, Ahram M, Linehan WM, Knezevic V and Emmert-Buck MR. (2002). *Nat. Genet.*, **32,** 509–514.

Cohen C, Zavala-Pompa A, Sequeira JH, Shoji M, Sexton DG, Cotsonis G, Cerimele F, Govindarajan B, Macaron N and Arbiser JL. (2002). *Clin. Cancer Res.*, **8,** 3728–3733.

Cohen Y, Xing M, Mambo E, Guo Z, Wu G, Trink B, Beller U, Westra WH, Ladenson PW and Sidransky D. (2003). *J. Natl. Cancer Inst.*, **95,** 625–627.

Cristianini N and Shawe-Taylor J. (2000). *An Introduction to Support Vector Machines: and Other Kernel-based Learning Methods*. Cambridge University Press: Cambridge.

Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, Teague J, Woffendin H, Garnett MJ, Bottomley W, Davis N, Dicks E, Ewing R, Floyd Y, Gray K, Hall S, Hawes R, Hughes J, Kosmidou V, Menzies A, Mould C, Parker A, Stevens C, Watt S, Hooper S, Wilson R, Jayatilake H, Gusterson BA, Cooper C, Shipley J, Hargrave D, Pritchard-Jones K, Maitland N, Chenevix-Trench G, Riggins GJ, Bigner DD, Palmieri G, Cossu A, Flanagan A, Nicholson A, Ho JW, Leung SY, Yuen ST, Weber BL, Seigler HF, Darrow TL, Paterson H, Marais R, Marshall CJ, Wooster R, Stratton MR and Futreal PA. (2002). *Nature*, **417,** 949–954.

Dong J, Phelps RG, Qiao R, Yao S, Benard O, Ronai Z and Aaronson SA. (2003). *Cancer Res.*, **63,** 3883–3885.

Ferrando AA, Neuberg DS, Staunton J, Loh ML, Huard C, Raimondi SC, Behm FG, Pui CH, Downing JR, Gilliland DG, Lander ES, Golub TR and Look AT. (2002). *Cancer Cell*, **1,** 75–87.

Francia G, Mitchell SD, Moss SE, Hanby AM, Marshall JF and Hart IR. (1996). *Cancer Res.*, **56,** 3855–3858.

Fukushima T, Suzuki S, Mashiko M, Ohtake T, Endo Y, Takebayashi Y, Sekikawa K, Hagiwara K and Takenoshita S. (2003). *Oncogene*, **22,** 6455–6457.

Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M and Haussler D. (2000). *Bioinformatics*, **16,** 906–914.

Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD and Lander ES. (1999). *Science*, **286,** 531–537.

Gorden A, Iman O, Weiming G, He D, Huang W, Davidson A, Houghton AN, Busam K and Polsky D. (2003). *Cancer Res.*, **63,** 3955–3957.
Hanley JA and McNeil BJ. (1982). *Radiology*, **143,** 29–36.
Harder K, Moller N, Peacock J and Jirik F. (1998). *J. Biol. Chem.*, **273,** 31890–31900.
Hedenfalk I, Duggan D, Chen Y, Radmacher M, Bittner M, Simon R, Meltzer P, Gusterson B, Esteller M, Kallioniemi OP, Wilfond B, Borg A and Trent J. (2001). *N. Engl. J. Med.*, **344,** 539–548.
Huang E, Ishida S, Pittman J, Dressman H, Bild A, Kloos M, D'Amico M, Pestell RG, West M and Nevins JR. (2003). *Nat. Genet.*, **34,** 226–230.
Jazaeri AA, Yee CJ, Sotiriou C, Brantley KR, Boyd J and Liu ET. (2002). *J. Natl. Cancer Inst.*, **94,** 990–1000.
Johnson J, Bar-Eli M, Jansen B and Markhof E. (1997). *Int. J. Cancer*, **73,** 769–774.
Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu CR, Peterson C and Meltzer PS. (2001). *Nat. Med.*, **7,** 673–679.
Kimura E, Nikiforova M, Zhu Z, Knauf J, Nikiforov Y and Fagin J. (2003). *Cancer Res.*, **63,** 1454–1457.
Luca M, Hunt B, Bucana C, Johnson J, Fidler I and Bar-Eli M. (1993). *Melanoma Res.*, **3,** 35–41.
Medrano EE. (2003). *Oncogene*, **22,** 3123–3129.
Mustelin T and Hunter T. (2002). *Sci. STKE*, **115,** PE3.
Naoki K, Chen TH, Richards WG, Sugarbaker DJ and Meyerson M. (2002). *Cancer Res.*, **62,** 7001–7003.
Pfaffl MW. (2001). *Nucleic Acids Res.*, **29,** e45.
Pollock PM, Harper UL, Hansen KS, Yudt LM, Stark M, Robbins CM, Moses TY, Hostetter G, Wagner U, Kakareka J, Salem G, Pohida T, Heenan P, Duray P, Kallioniemi O, Hayward NK, Trent JM and Meltzer PS. (2003). *Nat. Genet.*, **33,** 19–20.
Prathapam T, Kuhne C and Banks L. (2002). *Nucleic Acids Res.*, **30,** 5261–5268.
Rajagopalan H, Bardelli A, Lengauer C, Kinzler KW, Vogelstein B and Velculescu VE. (2002). *Nature*, **418,** 934.
Rajeevan MS, Ranamukhaarachchi DG, Vernon SD and Unger ER. (2001). *Methods*, **25,** 443–451.
Satyamoorthy K, Li G, Gerrero MR, Brose MS, Volpe P, Weber BL, Van Belle P, Elder DE and Herlyn M. (2003). *Cancer Res.*, **63,** 756–759.
Schleger C, Heck R and Steinberg P. (2000). *Mol. Carcinog.*, **28,** 31–41.
Schoch C, Kohlmann A, Schnittger S, Brors B, Dugas M, Mergenthaler S, Kern W, Hiddemann W, Eils R and Haferlach T. (2002). *Proc. Natl. Acad. Sci. USA*, **99,** 10008–10013.
Shih I, Speicher D, Hsu M, Levine E and Herlyn M. (1997). *Cancer Res.*, **57,** 3835–3840.
Simon R, Radmacher MD, Dobbin K and McShane LM. (2003). *J. Natl. Cancer Inst.*, **95,** 14–18.
Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lonning P and Borresen-Dale AL. (2001). *Proc. Natl. Acad. Sci. USA*, **98,** 10869–10874.
Srivastava M, Bubendorf L, Srikantan V, Fossom L, Nolan L, Glasman M, Leighton X, Fehrle W, Pittaluga S, Raffeld M, Koivisto P, Willi N, Gasser TC, Kononen J, Sauter G, Kallioniemi OP, Srivastava S and Pollard HB. (2001). *Proc. Natl. Acad. Sci. USA*, **98,** 4575–4580.
Troyanskaya OG, Garber ME, Brown PO, Botstein D and Altman RB. (2002). *Bioinformatics*, **18,** 1454–1461.
van Elsas A, Zerp SF, van der Flier S, Kruse KM, Aarnoudse C, Hayward NK, Ruiter DJ and Schrier PI. (1996). *Am. J. Pathol.*, **149,** 883–893.
Xie S, Luca M, Huang S, Gutman M, Reich R, Johnson J and Bar-Eli M. (1997). *Cancer Res.*, **57,** 2295–2303.
Yuen ST, Davies H, Chan TL, Ho JW, Bignell GR, Cox C, Stephens P, Edkins S, Tsui WW, Chan AS, Futreal PA, Stratton MR, Wooster R and Leung SY. (2002). *Cancer Res.*, **62,** 6451–6455.
Zheng X and Shalloway D. (2001). *EMBO J.*, **20,** 6037–6049.
Zheng X, Wang Y and Pallen C. (1992). *Nature*, **359,** 336–339.

Supplementary material can be viewed at the following URL: http://www.qimr.edu.au/research/labs/nickh/Pavey-et-al-Supporting-Information.pdf