New Functional Signatures for Understanding Melanoma Biology from Tumor Cell Lineage-Specific Analysis

Florian Rambow, Bastien Job, Valérie Petit, Franck Gesbert, Véronique Delmas, Hannah Seberg, Guillaume Meurice, Eric Van Otterloo, Philippe Dessen, Caroline Robert, Daniel Gauheret, Robert A. Cornell, Alain Sarasin, and Lionel Larue
Supplemental Methods

Similarity core analysis (SCA)

Data transformation and context generation

For gene expression, all public data were obtained from the same platform (GPL570, Affymetrix Human Genome U133 Plus 2.0), so the corresponding tumor dataset resulted in 21,049 unique gene symbols. We further restricted this set to the 17,322 gene symbols common to the Agilent Whole Human Genome Microarray 4x44K platform for further analysis. For miRNA, data were obtained from three platforms, resulting in an intersection of 422 miRNAs.

PCA space computation and tumor projection

We limited the RSC size to 20 genes for three main reasons: (i) for reliable correlation evaluation, (ii) to limit the chances of picking up a false-positive gene in the RSCs selected due to the presence of true positive genes, and (iii) to ensure that the calculation could be performed in a reasonable amount of time.

As our tumor dataset represented eight diseases, we limited the number of components to the first eight, assuming the different types of cancer to be the major source of variance. Most of the variance was still covered: 85% of the total variance for more than 99.9% of the RSCs (data not shown).

Data retrieval, preprocessing and normalization of public datasets

Normalized public datasets for mRNA and miRNA (Table S1a) levels were downloaded from GEO (Gene Expression Omnibus, NCBI), with GEOquery v2.28.0 for R. For each dataset, we discarded probes with more than 50% missing values across samples and samples with more than 10% missing values across probes. The remaining missing values were imputed by K-nearest neighbor averaging. We discarded probes tagged as controls in their corresponding GPL annotation or with
no gene symbol or miRNA name. Probe-level intensities were log2-transformed and merged by median to mRNA- or miRNA-level. Preprocessed (level 3) TCGA data were acquired from the official FTP website (https://tcga-data.nci.nih.gov/tcga/). All differential feature expression analyses were performed as described in their respective figure legends.

**Variability of gene expression**
Gene expression variability was assessed with the interquartile range (IQR). Genes were ranked in descending order of IQR and the top 5,000, those with the most variable expression, were retained (Domcke et al., 2013).

**Sample classification**
Hierarchical clustering and graphical representations were performed in R, using Euclidean and Spearman distances for samples and features, respectively, and Ward's construction method for both. We evaluated the stability of each branchpoint by generating and independently clustering 1000 bootstrap gene set replicates. The percentage of times each branch was recovered was calculated with pvclust v1.2.2 with default parameters (Suzuki and Shimodaira, 2006). For 3 SDE miRNA clustering, the signature was too small for bootstrapping.

**Literature vector analysis**
The SCA-MEL-mRNA signature was analyzed for overrepresented terms by Literature Vector Analysis or LitVAn (Akavia et al., 2010). This method uses a manually curated database (NCBI Gene) to connect genes with terms from the complete text of more than 70,000 published scientific articles (http://litvan.bio.columbia.edu). It generates a ranked list of terms, indicating their significance, and a map linking the genes, significant terms and papers contributing to the score. Higher scores are assigned to words overrepresented in a subset of
documents with respect to the full corpus. The Inverse Document Frequency (IDF) assigns a score to each “term” (word) based on the proportion of documents in which it appears, with high scores for low coverage. Term Frequency (TF), is calculated for a subset of documents rather than the entire set. It is a direct count of the number of times the term appears in the subset. For each set of genes (module), the TF in associated papers is determined and compared with the null distribution, through the TF*IDF score (Salton et al., 1983). The SCA-MEL-miR signature was analyzed indirectly, because LitVAn does not recognize miRNA identifiers. We therefore performed LitVAn on sequence-based predicted (http://www.targetscan.org/) miRNA targets overlapping the SCA-MEL-mRNA signature.

**Survival analysis**

Survival analysis was performed with the *survival v2.37-7* package for R (Therneau and Grambsch, 2000).

**Cell culture**

Melanoma cell lines (Table S1h) were cultured at 37°C in RPMI1640 (Gibco) supplemented with 10% fetal bovine serum, 1% penicillin/streptomycin, and 1% glutamine, under a humidified atmosphere containing 5% CO2.

**Microarray mRNA expression profiling**

Total RNA was isolated from melanoma cell lines with the miRNeasy Mini Kit (Qiagen). It was labeled with Cy5 and RNA from a pool of five normal human epidermal melanocyte (NHEM) samples was labeled with Cy3, according to the manufacturer’s instructions (Agilent Low Fluorescent Low Input Linear Amplification Kit). The RNA was hybridized in dual-color on the Agilent Whole Human Genome
4x44K Microarray (AMADID 014850). After hybridization and washing, slides were scanned on an Agilent G2505B DNA microarray scanner at a resolution of 5 µm and a color depth of 20 bits. Signal acquisition was performed with Feature Extraction software (Agilent) v10.2.1.3.

**Gene expression data preprocessing and normalization**

We extracted raw probe median intensity tracks for our melanoma cell lines. For each track, we discarded probes for which intensity measurements were saturated or pixel inconsistency was flagged. No further intensity-based flagging was applied. We identified two outliers for overall expression, which were discarded (WM793 and Quar). Replicated probe intensities were merged, using their median value. Datasets were limited to the 29,364 probes annotated with a HNGC HUGO gene symbol as of 2012/08, according to the “hgug4112a” database available in Bioconductor for R v3.0.2 (http://cran.r-project.org). Raw intensity data were log2-transformed and missing values were imputed by K-nearest neighbor averaging from the impute package v1.36.0 for R. Quantile normalization was then applied, with the limma package v3.18.13 for R.

**Differential expression analysis**

Differential expression (intra-melanoma): the Empirical Bayes (ComBat) (Johnson et al., 2007) method was applied to remove global differences between the two datasets GSE19234 (30 out 44 melanoma metastases, used in SCA) and GSE7127 (63 melanoma cell lines, used in SCA). Hierarchical clustering of the complete merged dataset displayed two clusters as observed with SCA. A differential expression test was then performed between melanomas of the two cluster types, using a Wilcoxon rank-sum test. Genes were ranked according to their increasing Benjamini-Hochberg FDR-adjusted p-value, and the first 100 genes considered for the signature.
Differential expression (melanoma versus others): a differential expression test using limma (Smyth et al., 2005) was performed on the set of 240 tumors used in SCA, comparing the expression of the 30 melanoma samples to the 210 other samples corresponding to all 7 other pathologies. Genes were ranked according to their increasing Benjamini-Hochberg FDR-adjusted p-value, and the first 100 genes considered for the signature.

DNA sequencing for BRAF and NRAS mutations

Genomic DNA was extracted from human melanoma cell lines and column-purified with the DNeasy Blood and Tissue kit (Qiagen). Genomic regions around the NRAS\textsuperscript{Q61} and BRAF\textsuperscript{V600} hotspots were amplified and amplicons were sequenced in both directions (Table S1h).

Wound scratch migration assay

Cells were added to six-well plates on day 0 and grown to confluence. On day 1, confluent cells were wounded by scratching with a 200µl yellow pipette tip and the medium was replaced. Cells were imaged at 30-minute intervals over a period of 15h, with a fully automated inverted microscope (Leica DM IRB, Leica Microsystems) in an environmental chamber containing a humidified 95% air/5% CO\textsubscript{2} atmosphere at 37°C (Life Imaging Services, Switzerland), operating under Metamorph software (Molecular Devices). The distance migrated was evaluated by measuring the size of the wound with ImageJ software (Gallagher et al., 2013). At least three independent assays were performed per cell line.

Matrigel invasion assay

Transwell inserts were coated with 50µg of Matrigel (BD Biosciences) for 1.5h at 37°C. We added 10\textsuperscript{5} melanoma cells in serum-free RPMI medium to transwell inserts (8 µm pore size, BD Biosciences). RPMI medium containing 10% fetal bovine serum
was added to the external compartment. The inserts were then incubated for 48 hours at 37°C. Non-invading cells remaining on the upper surface of the chamber were removed by scrubbing with a cotton-tipped swab. The inserts/cells were stained with crystal violet before lysing the cells on the outer side of the insert. Three randomly selected images were acquired per well and cells were counted with ImageJ. Experiments included biological triplicates and technical duplicates.

Xenografts
Xenograft experiments were performed as previously described (Grille et al., 2003). All animals were housed in specific pathogen-free conditions at Institut Curie, in accordance with French and European Union law. This work was approved by the ethics committee (P2.LL.029.07). For each melanoma line, we injected $4 \times 10^6$ cells into the right flank of six-week-old female athymic nude mice (Swiss nude, Charles River and NMRI nude, Janvier). Mice were kept in individually ventilated cages for up to 100 days postinjection. Tumor volume was measured weekly with Vernier calipers ($\text{volume} = \text{length} \times \text{width}^2$). If the condition of the mouse deteriorated (e.g., weight loss), it was killed and the tumors were removed. All surviving mice were killed on D100 and their tumors removed. The data presented are the mean tumor volumes for three mice per cell line.

Cell proliferation assay
Melanoma cells were incubated with 10µM 5-bromo-2'-deoxyuridine (BrdU) (Sigma) for 2h. They were then harvested, washed and fixed in 70% (v/v) ethanol. The permeabilized cells were treated with 2M HCl for 20min and neutralized by incubation with 0.1M sodium borate for 2min. Cells were incubated with FITC-labeled mouse anti-BrdU antibody (BD, Pharmingen) for 45min. Cells were washed, treated with RNaseA (10mg/ml) at 37°C for 20min and stained with propidium iodide.
(1mg/ml) (Sigma). Flow cytometry was performed on a FACScalibur (Becton Dickinson) and data were analyzed with CellQuest and Winmdii.

**RNA isolation, cDNA synthesis, and quantitative RT-PCR**

Total RNA was isolated from melanoma cell lines with the miRNeasy Mini Kit (Qiagen) and used for cDNA synthesis with the miScript II Reverse Transcription Kit (Qiagen). Real-time PCR was carried out with iTaq Universal Sybrgreen Supermix (Biorad). Quantitative RT-PCR was performed under the following thermocycler (AbiPrism 7900) conditions, in a final reaction volume of 25μl: 95°C for 1.5min, and 40 cycles of 95°C for 30s, 60°C for 60s, with a final melting curve analysis. Relative expression was determined by the comparative ΔΔCt method. This method involves comparison of target gene (CAPN3, TRIM63, MITF-M) expression with the expression of an endogenous control (TBP) within a sample, to normalize expression (Table S1i). We chose an appropriate reference sample from each group of samples. Each sample was then compared with the nominal reference sample, to determine the relative expression of the target gene with respect to this reference sample.

**RNA silencing**

501mel melanoma cells were transfected with siRNAs against CAPN3 and TRIM63 (ThermoScientific), at a final concentration of 100nM (Table S1i), with Lipofectamine. After 24h of RNA silencing, phenotypic assays were performed.

**Microarray miRNA expression profiling**

Total RNA was isolated from melanoma cell lines with the miRNeasy Mini Kit (Qiagen). Each sample was prepared according to the Agilent miRNA Microarray System protocol. Total RNA (100ng) was dephosphorylated with calf intestine alkaline phosphatase (GE Healthcare Europe GmbH), denatured with dimethyl
sulfoxide, and labeled with pCp-Cy3, using T4 RNA ligase (GE Healthcare Europe GmbH). The labeled RNAs were hybridized to Agilent Unrestricted Human miRNA microarrays (V3) for 20h at 55°C, with rotation. Slides were scanned with an Agilent 2565 AB DNA microarray scanner, with high dynamic range settings, as recommend by the manufacturer. Agilent Feature Extraction software (10.5.5.1) was used for data extraction. After quality control (QC) analysis, 2 melanoma cell lines (Quar and WM793) were excluded from the panel of 23.

miRNA target prediction

Two criteria were used to predict miRNA targets. First, the miRNA target was predicted by a sequence-based approach, using the miRNA Data Integration Portal (mirDIP, significant interaction between miR and target when the standardized score was above 50 [0 = least confident and 100 = most confident]), which includes 12 microRNA prediction datasets from six microRNA prediction databases (Shirdel et al., 2011). Second, mRNA and miRNA levels had to be inversely correlated. Spearman's rank correlation analysis was applied to mRNA and miRNA levels for human melanoma cell lines (n=21). We considered mRNA/miRNA pairs to be anticorrelated if the correlation coefficient was below -0.4 and p<0.05.

References

Akavia, U.D., Litvin, O., Kim, J., Sanchez-Garcia, F., Kotliar, D., Causton, H.C., Pochanard, P., Mozes, E., Garraway, L.A., and Pe'er, D. (2010). An integrated approach to uncover drivers of cancer. Cell 143, 1005-1017.
Bogunovic, D., O'Neill, D.W., Belitskaya-Levy, I., Vacic, V., Yu, Y.L., Adams, S., Darvishian, F., Berman, R., Shapiro, R., Pavlick, A.C., et al. (2009). Immune profile and mitotic index of metastatic melanoma lesions enhance clinical staging in predicting patient survival. Proceedings of the National Academy of Sciences of the United States of America 106, 20429-20434.
Domcke, S., Sinha, R., Levine, D.A., Sander, C., and Schultz, N. (2013). Evaluating cell lines as tumour models by comparison of genomic profiles. Nat Commun 4, 2126.
Gallagher, S.J., Rambow, F., Kumasaka, M., Champeval, D., Bellacosa, A., Delmas, V., and Larue, L. (2013). Beta-catenin inhibits melanocyte migration but induces melanoma metastasis. Oncogene 32, 2230-2238.
Gillet, J.P., Calcagno, A.M., Varma, S., Marino, M., Green, L.J., Vora, M.I., Patel, C., Orina, J.N., Eliseeva, T.A., Singal, V., et al. (2011). Redefining the relevance of established cancer cell lines to the study of mechanisms of clinical anti-cancer drug resistance. Proceedings of the National Academy of Sciences of the United States of America 108, 18708-18713.

Grille, S.J., Bellacosa, A., Upson, J., Klein-Szanto, A.J., van Roy, F., Lee-Kwon, W., Donowitz, M., Tsichlis, P.N., and Larue, L. (2003). The protein kinase Akt induces epithelial mesenchymal transition and promotes enhanced motility and invasiveness of squamous cell carcinoma lines. Cancer Res 63, 2172-2178.

Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 8, 118-127.

Jonsson, G., Busch, C., Knappskog, S., Geisler, J., Miletic, H., Ringner, M., Lillegaard, J.R., Borg, A., and Lonning, P.E. (2010). Gene expression profiling-based identification of molecular subtypes in stage IV melanomas with different clinical outcome. Clin Cancer Res 16, 3356-3367.

Mann, G.J., Pupo, G.M., Campain, A.E., Carter, C.D., Schramm, S.J., Pianova, S., Gerega, S.K., De Silva, C., Lai, K., Wilmott, J.S., et al. (2013). BRAF mutation, NRAS mutation, and the absence of an immune-related expressed gene profile predict poor outcome in patients with stage III melanoma. The Journal of investigative dermatology 133, 509-517.

Salton, G., Buckley, C., and Fox, E.A. (1983). Automatic query formulations in information retrieval. J Am Soc Inf Sci 34, 262-280.

Segura, M.F., Belitskaya-Levy, I., Rose, A.E., Zakrzewski, J., Gaziel, A., Hanniford, D., Darvishian, F., Berman, R.S., Shapiro, R.L., Pavlick, A.C., et al. (2010). Melanoma MicroRNA signature predicts post-recurrence survival. Clinical cancer research : an official journal of the American Association for Cancer Research 16, 1577-1586.

Shirdel, E.A., Xie, W., Mak, T.W., and Jurisica, I. (2011). NAViGaTing the micronome—using multiple microRNA prediction databases to identify signalling pathway-associated microRNAs. PLoS ONE 6, e17429.

Smyth, G.K., Michaud, J., and Scott, H.S. (2005). Use of within-array replicate spots for assessing differential expression in microarray experiments. Bioinformatics 21, 2067-2075.

Suzuki, R., and Shimodaira, H. (2006). Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics 22, 1540-1542.

Therneau, T., and Grambsch, P. (2000). Modeling Survival Data: Extending the Cox Model, Springer edn.

Widmer, D.S., Cheng, P.F., Eichhoff, O.M., Belloni, B.C., Zipser, M.C., Schlegel, N.C., Javelaud, D., Mauviel, A., Dummer, R., and Hoek, K.S. (2012). Systematic classification of melanoma cells by phenotype-specific gene expression mapping. Pigment cell & melanoma research 25, 343-353.
Figure legends

Figure S1: Alternative melanoma clusterings and SCA details (related to Figure 1)

Expression-based clustering of 63 melanoma cell lines and 44 melanomas (a-c). a) All 17,322 exploitable mRNAs on the microarray were used for unsupervised clustering. Melanoma cell lines (black) are clearly separated from melanomas (blue). b) Clustering with the 5,000 most variable genes results also in a separation of cell lines and tumors (Table S1j). c) Clustering with a supervised gene list including 244 multi-drug resistance (MDR) related genes separates distinctly cell lines from tumors (Gillet et al., 2011). d) Selection criteria for RSC in correlation space. As example, the graphical representation of the correlation space corresponding to RSC #3463, limited to the first two dimensions, is shown. 30 tumors (smaller dots) from 8 pathologies (different colors) are represented with their respective centroid (bigger dots), as well as 8 melanoma cell lines (crosses). Red dashed circle represents the outline of the selection criterion (coordinate of the melanoma cell line two times closer to the melanoma centroid than to any other centroid, the second closer one in this example being Ewing sarcoma). Convex hull is represented for each pathology as an evaluation of their area in the space. e) Signature code definition.

Figure S2: Clustering of melanoma cell lines using using the entire SCA signature (related to Figure 2)

Twenty three independent melanoma cell lines which did not contribute to SCA, were clustered using the 100 core genes. Confidence in clusters are indicated with bootstrap values (approximately unbiased p-value, au). The major classes are
independent of their mutational status for BRAFV600E (white B), NRASQ61K,L,R (white N) or wild type for BRAF/NRAS (white W). Note that one cluster of cell line (Mull, Gerlach and SK29) belonged to group 1 using SCA-MEL-mRNA-55UP and belonged to the other group using SCA-MEL-mRNA-100.

Figure S3: Implications of the SCA-MEL-mRNA-28SDE signature regarding BRAFi and MEKi sensitivity (related to Figure 5)

a) Differential expression of melanoma-overexpressed genes. From the 55 core mRNAs, generally overexpressed in melanoma, 28 are significantly differentially expressed (SDE) between the two main groups of melanoma cell lines. 26 of 28 genes are overexpressed in non-aggressive (cluster 1) compared to aggressive (cluster 2) cell lines.
(b-e) SCA-MEL-mRNA-28SDE signature predicts MEKi and BRAFi sensitivity.

b) 40 melanoma cell lines (CCLE) were clustered according to the expression of the SCA-MEL-mRNA-28SDE.

c) Melanoma cell lines expressing strongly the SCA-MEL-mRNA-28SDE (group B) are significantly more sensitive to the MEKi (AZD6244) according to a Mann-Whitney test.

d) 29 melanoma cell lines mutated for BRAFV600E (CCLE) were clustered according to the expression of SCA-MEL-mRNA-28SDE signature.

e) Melanoma cell lines expressing strongly the SCA-MEL-mRNA-28SDE signature (group B) are significantly more sensitive to the BRAFi (PLX4720) according to a Mann-Whitney test.

f) Hierarchical clustering of tumor samples using mRNA Hoek’s invasive signature. Unsupervised clustering of 240 tumor samples from 8 different cancer types that contributed to SCA based on 45 invasive genes (Widmer et al., 2012).
**Figure S4:** SCA approach using miRNA expression data and related clusterings (related to Figure 4)

miRNA expression-based clustering of 51 melanoma cell lines and 21 melanomas (a,b). a) All 422 exploitable miRNAs on the microarray were used for unsupervised clustering. Melanoma cell lines (black) are clearly separated from melanomas (blue).
b) Clustering with the 100 most variable miRNAs results also in a separation of cell lines and tumors (Table S1k).
c) SCA flowchart for miRNA expression. miRNA expression data of 8 different tumor types (Ma = Melanoma, Co = Colon, Ov = Ovary, Br = Breast, Lu = Lung, Li = Liver, Ew = Ewing, Gl = Glioblastoma) were acquired and preprocessed to generate Randomized Selected Contexts (RSC) or “packs” of 20 arbitrarily chosen miRNAs. SCA extracts cancer-type specific miRNA expression patterns which allow an *in vitro/in vivo* comparison by measuring the proximity of cancer cell lines (melanoma) with tumors (melanoma) versus other tumor types in PCA-based correlation spaces corresponding to 1 million RSCs. Screening of these spaces was performed according to a minimal distance criterion between melanoma cell lines and the melanoma centroid. To prioritize miRNAs contributing most to melanoma specificity, miRNA enrichment among the selected spaces was calculated and resulted in a core of 51 miRNAs (SCA-miRNA-signature).

**Figure S5:** Integration of SCA-mRNA and SCA-miRNA core genes and SCA-miR related survival analysis (related to Figure 5)

(a-c) miRNA/mRNA signature of the invasive melanoma phenotype.
a) Inversed correlation of miRNA and mRNA expression levels among 21 melanoma cell lines. Anticorrelation of expression is criterion 1 (see Table S1l).
b) Sequence-based miRNA target prediction approach using mirDIP. A representative example for sequence-based prediction is shown (Table S1l). Sequence-based prediction is criterion 2.
c) Regulatory network of melanoma-specific miRNAs and mRNAs after integration of both SCA signatures (criteria 1 and 2). Overexpression of miR-221-3p and miR-10a-5p and underexpression of miR-211-5p are linked to an invasive phenotype in vitro. EMT-transcription factors FOSL1 and TWIST1 were shown to induce the levels of the miR-221-3p and miR-10a-5p.

(d-g) Validation of survival analysis by external melanoma data sets. External melanoma data sets from different microarray platforms were analyzed to validate the link between survival and expression of the 28SDE signature.

d) Clustering of 57 melanoma metastaseses (Jonsson et al., 2010) using the 28 SDE signature (25 of 28 genes were recognized) generated two main clusters of samples/patients.

e) Distant metastasis-free survival (DMFS) was significantly reduced in cluster 1 patients (n=30) overexpressing the 28 SDE signature (p=0.0205 Mantel-Cox test).

f) Clustering of 44 melanoma metastaseses (Bogunovic et al., 2009) using the 28 SDE signature generated two main clusters of samples/patients.

g) Overall Survival (time since diagnosis) was significantly reduced in cluster 1 patients (n=29) overexpressing the 28SDE signature (p=0.0362 Mantel-Cox test).

(h-k) Known melanoma mRNA signatures and survival analysis. Published melanoma gene expression signatures were applied to publically available melanoma samples.

h) Melanoma metastases (TCGA portal) were clustered based on the expression of 97 "melanoma phenotype specific expression" (MPSE) genes (Widmer et al., 2012). Two principal clusters of samples/patients were obtained.

i) Overall survival of the 2 major groups of patients were not significantly altered (p=0.4433, Mantel-Cox test). 97 MPSE genes are not associated with melanoma survival.
j) Equally, a second melanoma signature which consists of 46 genes mainly linked to the immune system (Mann et al., 2013) was applied to the TCGA data. Two principal clusters of patients were obtained.

k) Overall survival of cluster 2 patients (n=94), generally underexpressing the 46 genes, was significantly reduced (p=0.0196, Mantel-Cox test).

(l-m) Known melanoma miRNA signature and survival analysis. A published melanoma miRNA expression signatures was applied to publically available melanoma samples.

I) Melanoma metastases (TCGA portal) were clustered based on the expression of 18 miRNAs (Segura et al., 2010). Two principal clusters of samples/patients were obtained.

m) Overall survival of cluster 2 patients (n=135), generally underexpressing the 18 miRNAs, was reduced (p=0.0606, Mantel-Cox test).

**Figure S6: Standard approaches to detect lineage specific gene expression (related to Figure 1)**

Comparison of standard approaches with SCA to capture tumor lineage specific gene expression.

a) Venn diagram shows commonly identified genes with different approaches. Top 100 ranked genes lists of three different approaches (SCA, Limma, Batch removal) were compared. The SCA approach identifies largely unique genes, which would have been missed by standard approaches.

b) Clustering of 240 tumor samples (8 cancer types) and 63 melanoma cell lines using the top 100 Limma genes does not lead to a clean separation of melanoma samples from the rest.

c) Clustering of 240 tumor samples (8 cancer types) and 63 melanoma cell lines using the top 100 Batch removal genes does not lead to a clean separation of melanoma samples from the rest.
Table

Table S1:

a) Information about the transcriptomic data sets analyzed: mRNA and miRNA level (related to Figure 1 and Figure 4)
b) SCA-mRNA signature: gene information (related to Figure 1)
c) SCA-mRNA signature: LitVAn literature analysis – terms and scores (related to Figure 1)
d) LitVAn literature analysis – terms and genes (related to Figure 1)
e) SCA-miRNA signature: miR information (related to Figure 4)
f) SCA-miRNA signature: LitVAn literature analysis (related to Figure 4)
g) Transcription factor target genes (related to Figure 3)
h) Melanoma cell line information (related to Figure 2)
i) qRT-PCR primer and siRNA details (related to Experimental Procedures)
j) mRNA list showing highest variation in expression: 5000 most variably expressed mRNAs (related to Figure S1a)
k) miRNA list showing highest variation in expression: 100 most variably expressed miRNAs (related to Figure S4b)
l) mRNA and miRNA integration (related to Figure 5 and Figure S5a,b,c)
17,322 mRNAs
5,000 most variable mRNAs
244 MDR related mRNAs

Figure S1

SCA = Similarity-Core-Analysis
MEL = Melanoma
UP = Overexpressed in melanoma cell lines and melanomas compared to non-melanoma samples
DWN = Underexpressed in melanoma cell lines and melanomas compared to non-melanoma samples
SDE = Significantly differentially expressed between 2 groups of melanoma cell lines obtained by unsupervised clustering using previous signature
Figure S2

Rambow et al.
Figure S4

Data Acquisition and preprocessing
- Retrieval of miRNA expression profiles
- Aggregation in a unique data set
- Scaling

Data Transformation
- Generation of one million packs containing the expression levels of 20 randomly selected miRNAs

Pack selection
- Tumors to SVD Eigensamples
- Similarity and specificity to melanoma

miRNA selection
- miRNA enrichment
- Core generation

all miRNAs (422)

100 most variable miRNAs

Data Acquisition and preprocessing
of Ma, Co, Ov, Br, Lu, Li, Ew, Gl
- Retrieval of miRNA expression profiles
- Aggregation in a unique data set
- Scaling

Data Transformation
- Generation of one million packs containing the expression levels of 20 randomly selected miRNAs

Pack selection
- Tumors to SVD Eigensamples
- Similarity and specificity to melanoma

miRNA selection
- miRNA enrichment
- Core generation

Figure S4

Rambow et al.
**Figure S5**

Rambow et al.
Figure S6

Rambow et al.