Published in final edited form as:
Mol Cancer Res. 2008 May ; 6(5): 760–769. doi:10.1158/1541-7786.MCR-07-0344.

Gene Expression Changes in an Animal Melanoma Model Correlate with Aggressiveness of Human Melanoma Metastases

Lei Xu[1,#], **Steven S. Shen**[1,2,*], **Yujin Hoshida**[3], **Aravind Subramanian**[3], **Ken Ross**[3], **Jean-Philippe Brunet**[3], **Stephan N. Wagner**[5], **Sridhar Ramaswamy**[4], **Jill P. Mesirov**[3], and **Richard O. Hynes**[1]

[1]Howard Hughes Medical Institute, Center for Cancer Research, Cambridge, MA 02139

[2]Biomicro Center, Massachusetts Institute of Technology, Cambridge, MA 02139

[3]Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge, MA 20141

[4]Massachusetts General Hospital & Harvard Medical School, 185 Cambridge Street, Boston, MA 02114

[5]DIAID, Department of Dermatology, Medical University of Vienna, and Center for Molecular Medicine, Austrian Academy of Sciences, Wahringer Gurtel 18-20, A-1090 Vienna, Austria

Abstract

Metastasis is the deadliest phase of cancer progression. Experimental models using immunodeficient mice have been used to gain insights into the mechanisms of metastasis. We report here the identification of a “metastasis aggressiveness gene expression signature” derived using human melanoma cells selected based on their metastatic potentials in a xenotransplant metastasis model. Comparison with expression data from human melanoma patients shows that this metastasis gene signature correlates with the aggressiveness of melanoma metastases in human patients. Many genes encoding secreted and membrane proteins are included in the signature, suggesting the importance of tumor-microenvironment interactions during metastasis.

Keywords

Melanoma; Metastasis; Gene Signature; Animal Model; Microarray

Introduction

Metastasis is the dispersal of cancer cells from their primary loci to distant organs and accounts for more than 90% of deaths in cancer patients. The mechanisms of metastasis remain incompletely understood (1-3). Metastasis is a rare event, as shown by both clinical and animal studies. Clinicians often find a large number of circulating tumor cells in cancer patients but not many detectable metastases (4,5). The process of metastasis has also been studied in animal metastasis models. In these models, a pool of poorly metastatic human tumor cells, is injected into immunodeficient mice and the resulting metastases are isolated and cultured in vitro as cell lines (6-10). These cell lines often show enhanced metastatic ability when reinjected into immunodeficient mice. Studies using such models have found that most of the injected cancer

Requests for reprints: Richard O. Hynes, 40 Ames St. E17-227, Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139. Phone: 617-253-6422; Fax: 617-253-8357; rohynes@mit.edu..
[#]Current address: Department of Biomedical Genetics, University of Rochester Medical Center, Rochester, NY 14642
[*]Current address: Helicos BioSciences Corporation, One Kendall Square, Building 700, Cambridge, 02139

cells are able to disseminate into different organs but only a small proportion of them grow as detectable metastases, consistent with the notion that metastasis is a relatively rare event. Consequently, metastases have been postulated to result from small populations of cancer cells within a primary tumor, which are able to enter and survive in the circulation and then exit the circulation and grow in a distant organ (4,5,11).

Recently, microarray analyses have provided valuable tools to dissect further the mechanisms of cancer progression and to improve cancer treatment. Gene expression patterns have been used to classify the subtypes of primary tumors and to predict the clinical outcome of their treatments (12-14). Among these studies, several have shown that primary tumors and their metastases have similarities in gene expression profiles (15-17), and van't Veer et al. and Ramaswamy et al discovered that some primary tumors contain gene signatures that can predict their propensity to metastasize (13,14). Those observations raise the possibility that the ability to metastasize is determined early in primary tumor development and does not require further selections among primary tumor cells. They suggest an alternative to the view that metastases arise from rare populations within the primary tumor.

These two views of metastasis are not mutually exclusive (18). It is entirely possible that different primary tumors can be, either *ab initio* or as a consequence of their progression/evolution, either of good or poor prognosis and for these properties to be reflected in their gene expression profiles. That does not exclude the possibility that further alterations in gene expression can be either contributory to, or necessary for, effective spread and growth of metastases and, indeed, data are available showing the existence of gene expression signatures characteristic of metastases (7,9,14). The challenge is to determine gene expression signatures that contribute to various aspects of tumor progression — predisposition to metastasis, actual metastasis, aggressiveness of metastases, etc., and to relate those signatures to clinical data and outcomes.

We describe here the derivation of a series of metastatic human melanoma cell lines from poorly metastatic parental lines and their analysis in xenotransplant tumor and metastasis models. Genes differentially expressed between tumor samples derived from highly metastatic derivatives and from their poorly metastatic parental lines were identified. Expression of this gene expression signature in human metastases was found to correlate with poor survival of melanoma patients with metastases. Among these genes, many are secreted or membrane proteins, suggesting the importance of interactions between tumor cells and their microenvironment in the aggression of metastases.

Materials and Methods

Derivation of Metastatic Melanoma Cell lines

The highly metastatic human melanoma cell lines were derived from two poorly metastatic parental lines as described (8). The two parental A375 lines were obtained either from ATCC (#CRL-1619, for Set A cells), or as a gift from Dr. Isaiah Fidler (Texas MD Anderson Cancer Center, for Set F cells). Briefly, 500,000 A375 cells were injected intravenously into nude mice (Cby.Cg-*Foxn1*^{nu}, Jackson Laboratory, Maine). Two months later, individual lung metastases (believed to be clones) from different mice were harvested and amplified *in vitro* as independent cell lines. These cell lines were reinjected into mice for a second round of selection. MEA2 cell line was reinjected for a third round of selection. In total, four cell lines were derived from the Set A parental line and seven were derived from the Set F parental line (Figure 1A and B). An SM cell line was derived in a similar manner in Dr. Fidler's laboratory (19) and was also included in the array analyses along with the other Set F cell lines.

To test the metastatic properties of derived cell lines, 200,000 Set A cells or 500,000 Set F cells were injected intravenously into nude mice and lung metastases were counted under a dissecting microscope two months later. These lung metastases were also harvested for microarray analyses. Those from the same mouse were pooled and processed as one sample for micoarray analyses. To generate subcutaneous tumors, 500,000 cells were injected subcutaneously into the right flank of immunodeficient mice and the tumors were harvested four weeks after injection. Each subcutaneous tumor was processed separately and used as one sample for microarray analyses. All the samples included in the array analyses and their corresponding cell lines were listed in Supplementary Document S1.

RNA Preparation and Data Collection

Each cell line was injected into at least three different mice to obtain subcutaneous tumors or lung metastases for microarray analyses. As described above, tumor(s) from the same mouse were pooled and processed as one sample for the array analyses. RNA was extracted from the tumors using Qiagen RNeasy Midi Kit according to the manufacturers' instructions. cRNA was prepared according to the GeneChip Expression Analysis Technical Manual (Affymetrix, Santa Clara, CA), hybridized onto HU133A chips (Affymetrix), and scanned by a GeneArray@ 2500 Scanner (Affymetrix).

The quality of raw microarray profiles was generally assessed using measurements of overall microarray fluorescence intensity (e.g. mean, variance), the distribution of feature or spot intensities, and the proportion of total genes showing significant signal. 32 set A and 39 set F data sets passed these quality control criteria and were used for subsequent data analysis, such as normalization, expression marker selection and gene set enrichment analysis (GSEA). The data have been deposited in NCBIs Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo) and are accessible through GEO Series accession number GSE7929 (set A) and GSE7956 (set F). They were normalized using RMAexpress software (http://stat-www.berkeley.edu/users/bolstad/RMAExpress/RMAExpress.html) before being further analyzed.

83 fresh melanoma biopsies from patients undergoing surgery were collected from 1992 to 2001 as a part of the diagnostic work-up or therapeutic strategy. Immediately after surgery, half of each specimen was fixed in formalin and processed for routine histology, and the other half was immediately snap-frozen and stored in liquid nitrogen until use for RNA extraction. Histopathological diagnosis of each tissue specimen was performed independently by two histopathologists. All patient material has been collected and used according to the approval by the institutional ethics committee and written informed consent in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki. Total cellular RNA was prepared by guanidinium thiocyanate extraction and cesium chloride centrifugation and purified from remaining melanin with the Qiagen Rneasy Fibrous Tissue Mini Kit (Quiagen, Hilden, Germany). cRNA was prepared and profiled as described above. The data have been deposited in NCBIs Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo) and are accessible through GEO Series accession number GSE8401.

Microarray Analyses

Marker selections were performed using the GenePattern software (http://www.broad.mit.edu/cancer/software/genepattern/) (20). Set A data were first preprocessed using the PreProcessDataset module (after which 9108 probe sets are left for further analyses) and then marker genes were selected using the ComparativeMarkerSelection module. Genes differentially regulated in tumors from set A metastatic cells were selected based on their fold change (>3) and their adjusted p-value (max $T < 0.01$). They include 185 probe sets, which correspond with 150 nonredundant genes. The GeneOntology analyses for

their cellular distribution and the pathways they are involved were performed using DAVID/EASE software (http://david.abcc.ncifcrf.gov/tools.jsp).

S-plus software (Insightful) was used to generate Kaplan-Meier survival curves and perform the log-rank tests for assessment of statistical significance between each pair of curves. All the Fisher's exact tests were performed using the web-based calculation tools (http://www.physics.csbsju.edu/stats/).

Gene Set Enrichment Analysis (GSEA)

For the analyses shown in Figure 2, the 100, 200, or 500 most up- or down-regulated genes in tumors from each set of metastatic derivatives were selected using the ClassNeighborhood module in the GenePattern software. Signal-to-noise ratio was used to calculate statistics and the cut-off p-values were assigned based upon 1000 random permutation tests. The selected marker genes from one data set were subsequently used as gene sets to measure their enrichment in the other data set by GSEA (see ref. 22 and http://www.broad.mit.edu/gsea/). A normalized enrichment score (NES) was calculated based on the size of the gene set and its enrichment score. A nominal p-value was calculated after permutation testing of the microarray samples and a false discovery rate (FDR) (21) was calculated to correct for multiple hypothesis testing. Generally, a gene set is considered significantly enriched when its p-value is less than 0.05 and FDR score is less than 0.25 (22).

For the analysis shown in Figure 5, the ~22,000 probe sets on HU133A Affymetrix DNA chips were ranked by their Cox scores, which evaluate the correlation of their expression values with the poor survival of patients that developed melanoma metastases. This pre-ranked list of probe sets was used as the template and up-regulated or down-regulated genes within the 150-gene signature were used as gene sets to perform GSEA and measure their correlation with poor survival of patients with metastases.

Nearest Template Prediction Method

The nearest template prediction is a variation of the k-means clustering. First, we defined the "templates" of the "metastasis" and "non-metastasis" patterns, which are equivalent to the centroid in the k-means method. They include the 185 probe sets of our signature. In the templates, only the information on the direction of gene expression change was retained, i.e., in the "metastasis" template, values for the up-regulated probe sets (99 in total) were set to one, and the values for the down-regulated probe sets (86 in total) were set to zero, and vice versa for the "non-metastasis" template.

The expression values of each of the 185 probe sets were normalized across all the samples, with a mean equal to 0 and standard deviation equal to 1. The distance of each sample to either of the templates was assessed by Pearson correlation coefficient. The samples were then separated into two major groups first based on which template they were closer to, then by the significance of their proximity. The significance of the proximity was evaluated based on an empirical null distribution of the correlation coefficient generated by randomly picking the same number of genes from the entire microarray data for each sample (n=1000). A nominal p-value was computed using the rank of the observed correlation coefficient in the null distribution. The nominal p-value was corrected for multiple hypotheses testing using the false discovery rate (FDR). An FDR < 0.05 was regarded as significant. Samples closer to the "metastasis" template and with FDR < 0.05 were grouped as class 1, samples closer to the "non-metastasis" template and with FDR < 0.05 were grouped as class 3, and the rest of samples were grouped as class 2.

Results

Derivation of two groups of highly metastatic cell lines from poorly metastatic human melanoma cells

To study the mechanisms of metastasis, we took advantage of an experimental metastasis assay to derive several melanoma cell lines from two poorly metastatic parental lines (Figure 1A and B) (8) . The parental lines were either from ATCC or from Dr. Fidler (U. of Texas MD Anderson Cancer Center). These are both A375 melanoma cells, but have been cultured separately for decades. Nevertheless, they both remain poorly metastatic when tested *in vivo* and give rise to few lung metastases when injected into the circulation of immunodeficient mice. These few lung metastases were isolated and amplified *in vitro* as cell lines (MA-1, MB-1, MC-1 and MEA-1, MEC-1 and MED-1) (Figure 1A and B). The derivatives from the ATCC parental line were denoted as set A cells, and those from the Fidler parental line were denoted as set F cells. These cells were reinjected into mice for a second round of selection to generate MA-2 and MC-2 from set A and MEA-2, MEC-2 and MED-2 from set F (Figure 1 A and B). The MEA-2 cell line was injected once more into immunodeficient mice to generate cell line MEA-3 (Figure 1B). The derived cell lines exhibit increased metastatic ability as compared with their parental lines when tested by intravenous injection (Figure 1C).

Tumors from the two groups of metastatic derivatives show similar expression profiles

We performed microarray analyses to identify genes that are up- or down-regulated in the tumor samples from highly metastatic cells compared with those from the parental lines. Each cell line was injected into immunodeficient mice either intravenously or subcutaneously. The lung metastases or subcutaneous tumor from each mouse were processed as one biologically independent sample in the microarray analyses after their RNA was extracted and hybridized onto human oligonucleotide microarrays (see Materials and Methods). Genes differentially expressed between the tumor samples from metastatic variants and those from their respective parental lines were identified using GenePattern software. The expression levels of some of them (~30 genes) were validated by real-time PCR and found to be largely consistent with the results from array analyses. The top and bottom 50 differentially expressed genes are shown in Figure 2A. Gene Set Enrichment Analysis (GSEA) was performed to examine the similarities between the expression profiles (22). GSEA uses an algorithm that measures the cumulative enrichment of one set of genes (called a gene set) in a ranked second gene list from an array comparison. In our analysis, the most up-regulated or down-regulated 100, 200, or 500 genes from the tumor samples of one set of metastatic cells were used as gene sets to measure their enrichments in the tumor samples from the other set. A normalized enrichment score (NES) was assigned to each gene set and the statistical significance of its enrichment was measured by p-values and FDR scores (see Materials and Methods). The results showed that genes altered in the tumor samples from one set of cells were significantly altered in the other ($p < 0.05$ and FDR < 0.25; Figure 2B), indicating that similar genes or pathways are regulated in the tumor samples from the two independent sets of derived human metastatic melanoma cells, although the precise order of genes differed between the two sets. However, although the gene lists from the two sets of tumors were related, the gene expression values for set A samples were more consistent and the set A data set has been used in the subsequent analyses.

Metastasis genes from set A tumors are differentially expressed in human melanoma metastases and associated with poor survival

We next tested whether the up- or down-regulated genes in tumor samples from our metastatic variants correlate with the clinical outcome of human melanoma patients. We selected probe sets that are up- or down-regulated in the tumors from Set A metastatic cells by at least three-fold and with a maxT value < 0.01 (see Materials and Methods). 185 probe sets (99 up-regulated and 86 down-regulated) were selected, which correspond with 150 non-redundant genes (74

up-regulated and 76 down-regulated; Supplementary Document S3). These probe sets were used as templates for the following nearest template prediction method (for details, see Materials and Methods). Briefly, the expression values of these probe sets in array data from 52 human melanoma metastases were extracted and normalized with a mean equal to 0 and a standard deviation equal to 1. Their distances from the templates were calculated using Pearson correlation coefficient and FDR scores were calculated to measure the statistical significance. The human metastasis samples were separated into three major classes based on their distances from the templates and their false discovery rates (FDRs) (Figure 3A). Class 1 and Class 3 metastases samples show opposite expression patterns of the 185 probe sets and their FDR scores are less than 0.05. Class 1 samples showed similar expression patterns to the tumor samples from our Set A metastatic derivatives, whereas class 3 samples show similar expression patterns to those from our Set A parental line. Class 2 samples have FDR scores of over 0.05 and their expression pattern of the 185 probe sets are in between class 1 and class 3 samples. The survival data from patients with the three classes of metastases were used to generate Kaplan-Meier survival curves (Figure 3B, Supplementary Document S4). The patients with class 1 metastases had shorter survival than the patients with class 2 or class 3 metastases. The difference in survival probability between patients with class 1 and class 3 metastases was found to be significant by a log rank test ($p = 0.003$). Since class 3 metastases include a few lymph node metastases, which contain high percentages of contaminating lymphocytes, we tested whether the survival difference between class 1 patients and class 3 patients still exist when the lymph node metastases were excluded from our analyses. The results showed that the survival probability of class 1 patients still differed significantly from that of class 3 patients when lymph node metastases were omitted from the analyses (Supplementary Figure S5). These data suggest that the set of 150 genes differentially regulated in the tumors from our set A metastatic cells correlates with the aggressiveness of human melanoma metastases.

Interestingly, expression levels of these 150 genes also separated 31 human primary melanomas into three classes (Figure 3C). However, since the nearest template prediction method describes the relative distribution among the input samples, this classification could be specific to primary melanomas. To test whether the three classes of primary melanomas correspond with the three classes of metastases as described in Figure 3A and C, we mixed the primary melanoma samples with metastases samples and regrouped them using the nearest template prediction method. We found that most of the class 1 primary melanomas grouped with class 1 metastases, and most of the class 3 primary melanomas grouped with class 3 metastases (Figure 4, bottom bars), suggesting that the gene expression profile characteristic of aggressive human melanoma metastases (see Figure 3A, B) is already represented in a subclass of primary melanomas (Figure 3C). However, the patients with the class 1 and class 3 of primary melanomas did not differ significantly in their survival ($p = 0.182$, log rank test, Supplementary Figure S6). They also did not differ in their Clark levels when the proportion of level III samples in all class 1 samples was compared with that in all class 3 samples (Figure 3C, colored squares, $p = 0.38$ Fisher's exact test), or their Breslow thicknesses when the proportion of samples with thickness of less than 1 mm in all class 1 samples was compared with that in all class 3 samples (Figure 3C, colored hexagons, $p = 0.32$, Fisher's exact test). They also did not differ in their propensity to develop metastases (Figure 3C, black dots, $p = 0.224$, Fisher's exact test). This lack of significant difference between these two classes of primary melanomas may well be due to the small number of patients that are available for the analyses (see Discussion).

The correlation between our 150 genes and the aggressiveness of human melanoma metastases was also tested by GSEA. We first pre-ranked the ~22,000 probe sets on the HU133A DNA chip based on their Cox-ranking scores, measuring their correlations with the survival of patients carrying melanoma metastases. Then the 185 probe sets that correspond with our 150

genes were split into two gene sets: up-regulated genes in one set and down-regulated genes in the other. The enrichments of these two gene sets in the above pre-ranked probe set list were measured by GSEA. The results showed that the up-regulated probe sets were significantly enriched ($p < 0.05$ and FDR < 0.25) (Figure 5A and B, Supplementary Document S7), and 56 of them contributed positively to the enrichment (Supplementary Document S7). These data suggest that up-regulated genes among the 150 correlate with poor survival of patients carrying melanoma metastases. This is consistent with our previous finding using the nearest template prediction method and suggests that the up-regulated genes are most significantly correlated with aggressiveness in melanoma metastases

Strikingly, among these 150 genes, many are externally exposed proteins, such as secreted or extracellular matrix proteins or plasma membrane proteins (Supplementary Document S1 and Figure 6B): these two categories represent about 56% of the 74 up-regulated genes and 42% of the 76 down-regulated genes compared with ~15 % in the total genes unselected by fold changes and maxT value (which correspond to 9108 probe sets on HU133A chip) (Figure 6A, Supplementary Document S8). KEGG pathway analyses using DAVID/EASE software showed that the 150 genes appear to be involved mostly in processes such as ECM-receptor interaction, focal adhesion, TGF-beta signaling pathway, as well as glycerolipid metabolism and arginine and proline metabolism (Supplementary Document S8). For further discussion of the potential implications for metastasis of this gene signature, see the Discussion.

Discussion

In this paper, we report the derivation of highly metastatic human melanoma cell lines from poorly metastatic parental lines, using an animal metastasis model. We subsequently identified a “metastasis aggressiveness gene signature” by comparing the gene expression patterns of tumor samples from the highly metastatic derivatives with those from their parental lines. By comparisons with gene expression data from human clinical samples, we found that expression of this “metastasis gene signature” in human melanoma metastases correlates with poor survival of the corresponding patients. The signature is able to segregate melanoma-bearing patients into three groups, one of which has a significantly lower survival probability. That suggests that the signature provides an indication of “aggressiveness” of the melanoma metastases rather than of metastasis per se, similar to the lung metastasis signature reported by Minn. et al. (10). This result has been confirmed by alternative methods, such as GSEA and hierarchical clustering (data not shown). Interestingly, our gene signature is also able to separate primary tumors into the same three classes, a result reminiscent of some other gene signatures that have been reported (13,14,17). Given the size of the sample of primary tumors, we were unable, as yet, to demonstrate a predictive role for the gene signature when detected in primary tumors. Only six patients’ primary tumors expressed the signature (see Figure 3C, class 1), and, in those six patients, only one had reported metastases and two died — insufficient for statistical analysis. It will be of interest to test this gene signature on independent data sets from larger numbers of patients to test whether it has useful predictive value. It is clear that not all melanoma metastases express the 150-gene signature that we have described and the data suggest that those that do express it have a poorer prognosis, so the signature may well have prognostic value for patients with diagnosed and biopsied metastases. Those possibilities need to be tested against independent sets of melanoma samples.

Turning next to the implications of our results for understanding the cellular mechanisms of metastatic spread, several issues need to be discussed. Since the 150-gene signature is expressed in some primary tumors, it falls into the category of expression profiles that preexist in the bulk of (some) primary tumors, although it is unclear whether the profile originates in the cells of origin of the tumor, is a consequence of the nature of the initial oncogenic transformation event or arose during progression/development of the primary tumor. It is also

unclear whether this signature is necessary or sufficient for metastasis, although the data do suggest that it correlates with, and may contribute to, the “aggressiveness” of the metastases. It is entirely possible that other sets of genes are necessary for metastasis even in the context of cells expressing the 150-gene signature and it is clear (from the existence of metastases that do not express this signature) that it is not the only gene expression state characteristic of metastasis.

It is of some interest that the 150-gene signature that we report here is significantly enriched in secreted and membrane proteins that could be involved in tumor-microenvironment interactions contributing to the progression of metastases. Among them, the 56 genes that contribute positively to the enrichment of the 74 up-regulated genes in aggressive melanoma metastases are likely good candidates for further investigations. Many of them have been implicated previously in melanoma progression - examples include endothelin receptor (23), ERBB (24), Frizzled homolog 7 (25). All these genes can be tested directly in the animal metastasis model for their roles in the interactions between tumor cells and their microenvironment in metastasis. They should also be good markers for diagnosis and prognosis of malignant melanoma, since they are secreted and thus may be present in the plasma of cancer patients and easily detected by methods such as ELISA or mass spectrometry (26).

In conclusion, the findings we report here should contribute to the understanding, diagnosis and treatment of malignant melanoma, which currently remains essentially untreatable.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial support: NIH (RO1-CA17007), the National Cancer Institute’s Integrative Cancer Biology Program (U54-CA112967), the Virginia and Daniel K Ludwig Fund for Cancer Research and the Howard Hughes Medical Institute (to R.O.H.). Grant No. APP19722FW from the FWF - Fonds zur Förderung der wissenschaftlichen Forschung (to S.N.W.).

References

1. Chambers AF, Groom AC, MacDonald IC. Dissemination and growth of cancer cells in metastatic sites. Nat Rev Cancer 2002;2(8):563–72. [PubMed: 12154349]
2. Steeg PS. Tumor metastasis: mechanistic insights and clinical challenges. Nat Med 2006;12(8):895–904. [PubMed: 16892035]
3. Gupta GP, Massague J. Cancer metastasis: building a framework. Cell 2006;127(4):679–95. [PubMed: 17110329]
4. Fidler IJ. The pathogenesis of cancer metastasis: the ‘seed and soil’ hypothesis revisited. Nat Rev Cancer 2003;3(6):453–8. [PubMed: 12778135]
5. Tarin D, Price JE, Kettlewell MG, Souter RG, Vass AC, Crossley B. Mechanisms of human tumor metastasis studied in patients with peritoneovenous shunts. Cancer Res 1984;44(8):3584–92. [PubMed: 6744281]
6. Fidler IJ. Selection of Successive Tumour Lines for Metastases. Nature New Biology 1973;242:148–9.
7. Clark EA, Golub TR, Lander ES, Hynes RO. Genomic analysis of metastasis reveals an essential role for RhoC. Nature 2000;406(6795):532–5. [PubMed: 10952316]
8. Xu L, Begum S, Hearn JD, Hynes RO. GPR56, an atypical G protein-coupled receptor, binds tissue transglutaminase, TG2, and inhibits melanoma tumor growth and metastasis. Proc Natl Acad Sci U S A 2006;103(24):9023–8. [PubMed: 16757564]

9. Kang Y, Siegel PM, Shu W, et al. A multigenic program mediating breast cancer metastasis to bone. Cancer Cell 2003;3(6):537–49. [PubMed: 12842083]
10. Minn AJ, Gupta GP, Siegel PM, et al. Genes that mediate breast cancer metastasis to lung. Nature 2005;436(7050):518–24. [PubMed: 16049480]
11. MacDonald IC, Groom AC, Chambers AF. Cancer spread and micrometastasis development: quantitative approaches for in vivo models. Bioessays 2002;24(10):885–93. [PubMed: 12325121]
12. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98(19):10869–74. [PubMed: 11553815]
13. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415(6871):530–6. [PubMed: 11823860]
14. Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. Nat Genet 2003;33(1):49–54. [PubMed: 12469122]
15. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. Nature 2000;406 (6797):747–52. [PubMed: 10963602]
16. Weigelt B, Glas AM, Wessels LF, Witteveen AT, Peterse JL, van't Veer LJ. Gene expression profiles of primary breast tumors maintained in distant metastases. Proc Natl Acad Sci U S A 2003;100(26): 15901–5. [PubMed: 14665696]
17. Weigelt B, Hu Z, He X, et al. Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer. Cancer Res 2005;65(20):9155–8. [PubMed: 16230372]
18. Hynes RO. Metastatic potential: generic predisposition of the primary tumor or rare, metastatic variants-or both? Cell 2003;113(7):821–3. [PubMed: 12837240]
19. Kozlowski J, Hart IR, Fidler IJ, Hanna N. A human melanoma line heterogeneous with respect to metastatic capacity in athymic nude mice. J. Natl. Cancer Inst 1984;72(4):913–7. [PubMed: 6584666]
20. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. GenePattern 2.0. Nat Genet 2006;38 (5):500–1. [PubMed: 16642009]
21. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. Stat Med 1990;9(7):811–8. [PubMed: 2218183]
22. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102(43): 15545–50. [PubMed: 16199517]
23. Lahav R. Endothelin receptor B is required for the expansion of melanocyte precursors and malignant melanoma. Int J Dev Biol 2005;49(23):173–80. [PubMed: 15906230]
24. Bodey B, Bodey B Jr, Groger AM, Luck JV, Siegel SET, Taylor CR, Kaiser HE. Clinical and prognostic significance of the expression of the c-erbB-2 and c-erbB-3 oncoproteins in primary and metastatic malignant melanomas and breast carcinomas. Anticancer Res 1997;17(2B):1319–30. [PubMed: 9137492]
25. Weeraratna AT. A Wnt-er wonderland-the complexity of Wnt signaling in melanoma. Cancer Metastasis Rev 2005;24(2):237–50. [PubMed: 15986134]
26. Wulfkuhle JD, Liotta LA, Petricoin EF. Proteomic applications for the early detection of cancer. Nat Rev Cancer 2003;3(4):267–75. [PubMed: 12671665]

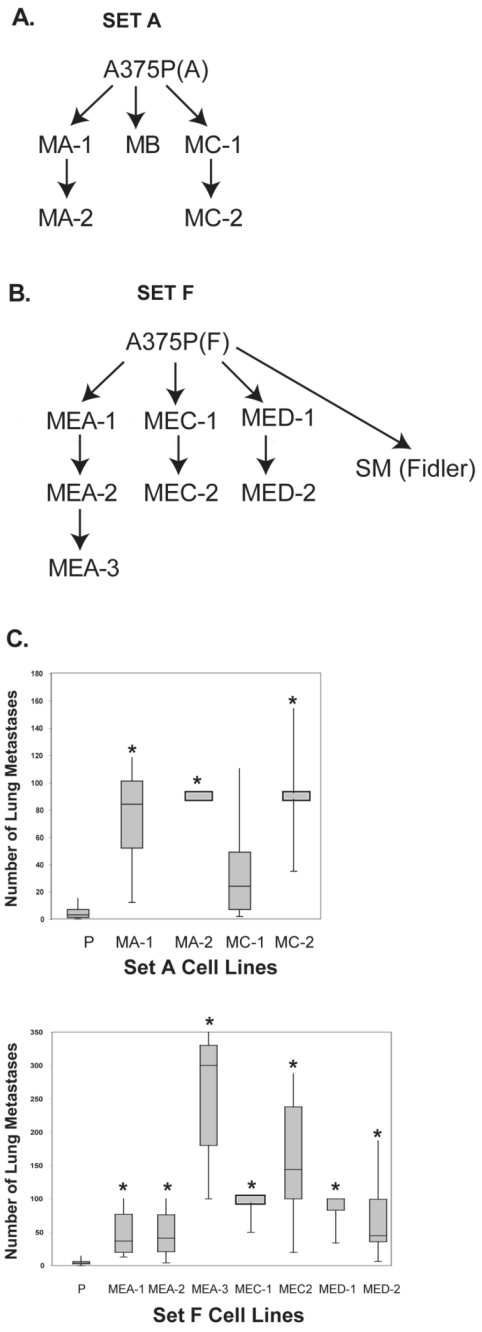
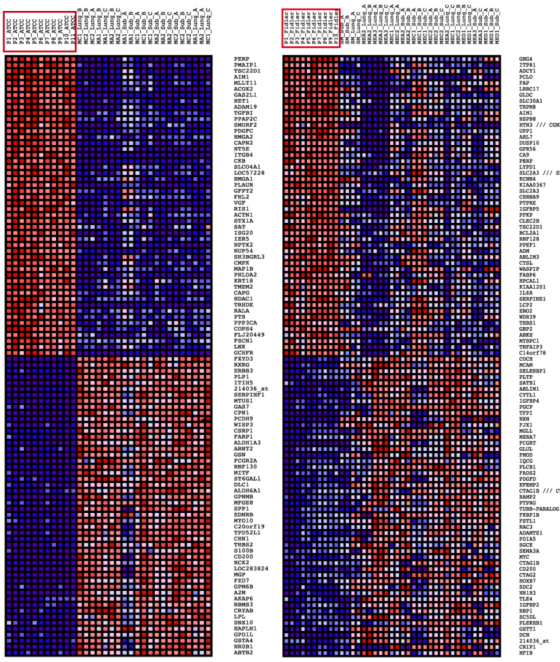


Figure 1. Derivation of Highly Metastatic Melanoma Cell Lines

A, B) Lineages of derived cell lines. Two related poorly metastatic A375P melanoma cell lines were injected into immunodeficient mice and individual lung metastases were isolated and cultured *in vitro* as independent cell lines (see Materials and Methods).

C) These derived cell lines are more metastatic than the parental lines. For set A cell lines, 2×10^5 cells were injected into immunodeficient mice ($n > 5$) and lung metastases were counted after one month. For Set F cell lines, 5×10^5 cells were injected and lung metastases were counted after two months. *: p value < 0.05.

A.



B.

Signatures from Data Set A are Enriched in Data Set F

NAME	ES	NES	NOM p-val	FDR q-val	FWER p-val	Number of Enriched Genes
Down_100	-0.76	-1.62	0.009	0.009	0.020	63
Down_200	-0.73	-1.59	0.028	0.017	0.036	116
Down_500	-0.65	-1.50	0.081	0.052	0.115	234
Up_100	0.77	1.51	0.05	0.046	0.106	77
Up_200	0.78	1.61	0.009	0.013	0.028	141
Up_500	0.72	1.60	0.024	0.013	0.033	339

Signatures from Data Set F are Enriched in Data Set A

NAME	ES	NES	NOM p-val	FDR q-val	FWER p-val	Number of Enriched Genes
Down_100	-0.71	-1.64	0.008	0.007	0.018	64
Down_200	-0.72	-1.67	0.005	0.008	0.015	131
Down_500	-0.69	-1.65	0.005	0.008	0.012	296
Up_100	0.81	1.79	0.002	0.003	0.004	62
Up_200	0.71	1.7	0.017	0.005	0.011	94
Up_500	0.69	1.66	0.027	0.005	0.011	227

Figure 2. Gene Set Enrichment Analysis Shows that Data Set A and F Contain Similar Patterns of Gene Expression

A. Expression patterns of the top and bottom 50 differentially expressed genes in set A and set F data identified by conventional marker selection method (class neighborhood in GenePattern software). Samples from the poorly metastatic parental lines are marked by red boxes. Sample names and gene symbols on each heatmap are also listed in Supplementary Document S2 for reference.

B. Gene set enrichment analysis (see Materials and Methods) shows that the top-ranked subsets of genes in set A are significantly enriched in samples from metastatic variants from Set F, and vice versa. ES: Enrichment Score; NES: Normalized Enrichment Score; FDR: false-discovery rate; Number of Enriched Genes : the number of genes in one gene subset that contribute positively to its enrichment in the other expression data set ; Gene subsets analyzed are as follows; Up_100, _200, or 500: top 100 or 200 or 500 up-regulated genes in samples from highly metastatic variants; Down_100, _200, or _500: top 100, 200, or 500 down-regulated genes in samples from highly metastatic variants. P values < 0.05 and FDR score < 0.25 are considered to reflect relatedness between a gene set and the comparison gene list. Note that all but one of the comparisons meet these criteria.

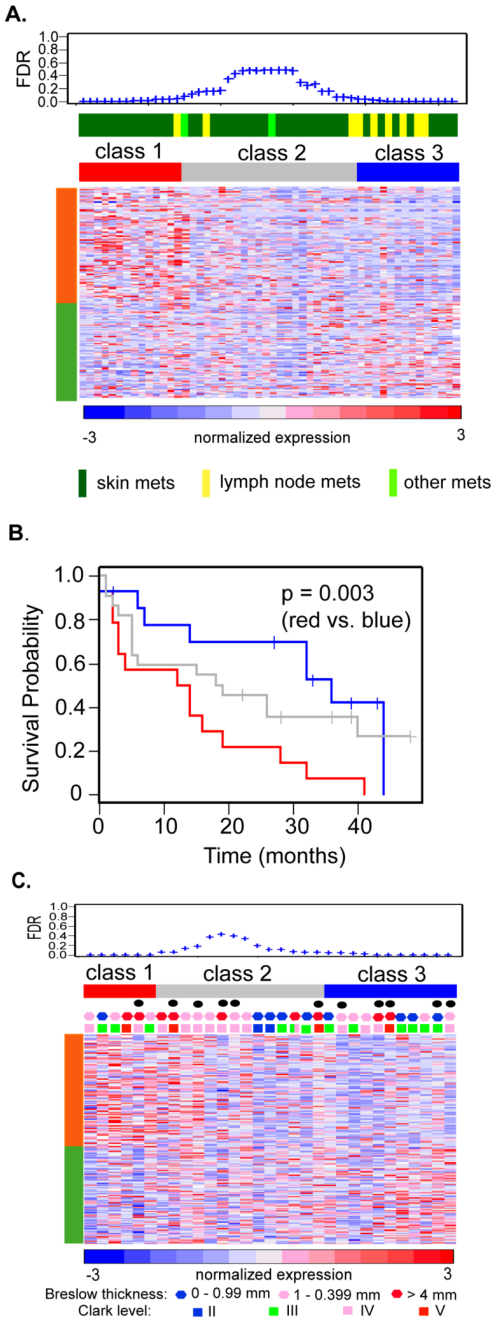


Figure 3. Genes Up-regulated in the Metastatic Derivatives from Set A Correlate with Metastatic Death in Human Melanoma Patients

A. The 185 probe sets (150 genes; see text) were used as the template for the nearest template prediction (for details, see Materials and Methods). They include 99 up-regulated probe sets (indicated by an orange bar) and 86 down-regulated probe sets (indicated by a green bar). Expression values of the 185 probe sets were extracted from expression profiling data from human melanoma metastases and were used to calculate the distance of each sample from the template. The human melanoma metastases were separated into three classes based on their distance and FDR value. Class 1 metastases (red bar) express high levels of the up-regulated genes in the signature, FDR < 0.05; class 3 metastases (blue bar) express high levels of the

down-regulated genes in the signature, FDR < 0.05; and class 2 metastases (grey bar) express intermediate levels of the signature genes, FDR > 0.05.

B. Kaplan-Meier survival curves were generated based on the correlation between the survival of patients and the classes of metastases they carry. The difference in survival probability between patients with the class 1 and class 3 metastases was found to be significant by a log-rank test. C. Similar analyses were performed as in A, but using expression data from primary human melanomas. The 185 probe sets (with the up-regulated ones indicated by an orange bar and down-regulated ones indicated by a green bar) separated the primary melanomas into three classes (red, grey and blue bars, respectively). Those melanomas that gave rise to metastases are labeled by black dots. Their Breslow thicknesses and Clark levels are labeled by colored hexagons and squares, respectively. The two extreme classes (class 1 and 3) of primary melanomas do not differ in their abilities to develop metastases ($p = 0.224$, Fisher's exact test), or their Breslow thicknesses ($p = 0.32$ if the proportion of samples with thickness of less than 1 mm was compared between the two classes, Fisher's exact test), or their Clark levels ($p = 0.38$ if the proportion of level III samples was compared between the two classes, Fisher's exact test). They also did not differ significantly in survival probability ($p = 0.106$, log-rank test), possibly because of the small number of patients available for analysis (see Discussion).

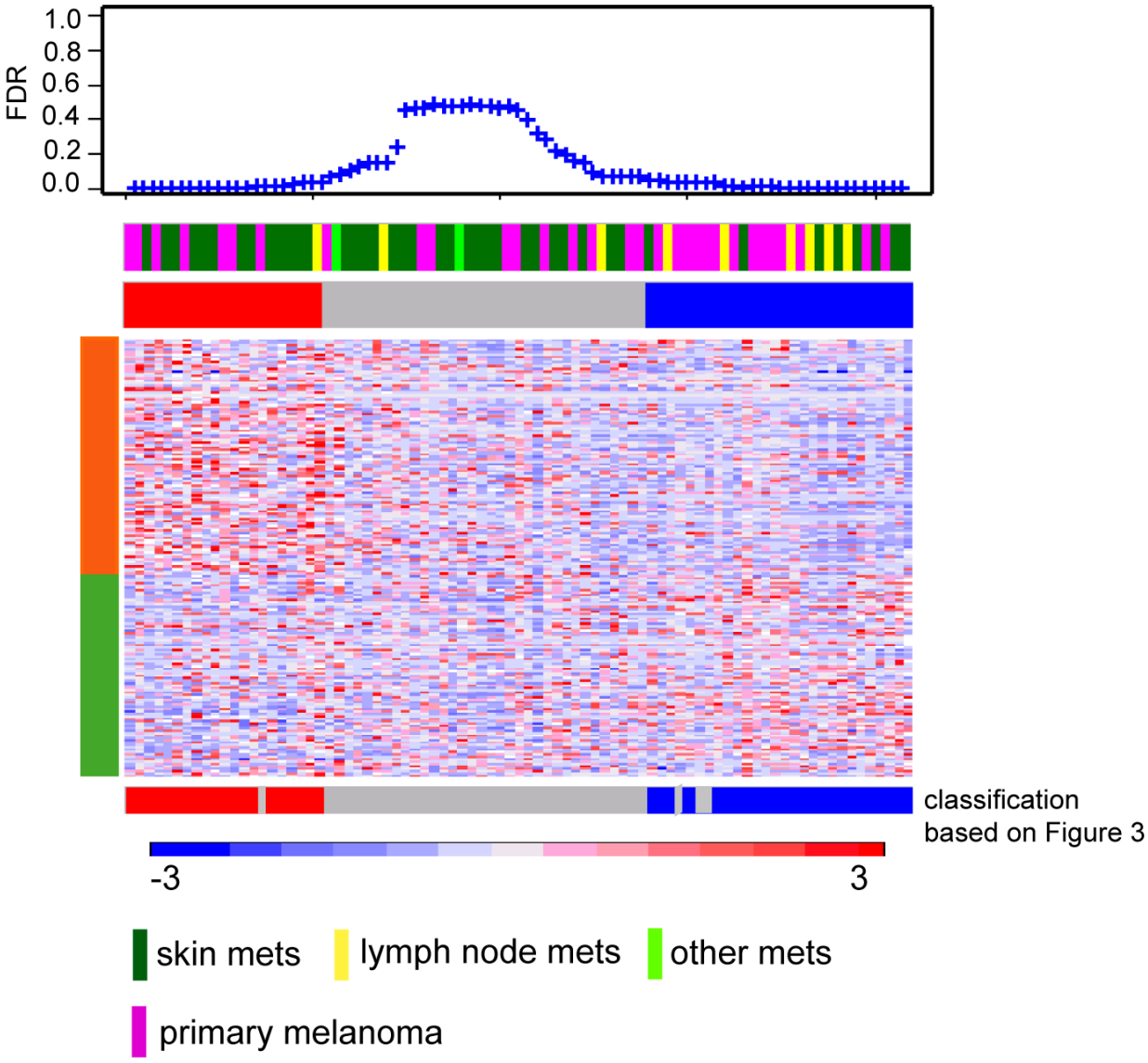


Figure 4. The three classes of primary melanomas and metastases show similar expression patterns of the 150 genes

The expression values of the 150 genes (see Text) were extracted from the mixed primary melanoma and metastasis samples and separated these samples into three classes by the nearest template prediction method as described in Figure 3A. These classes were indicated by a red, a grey, or a blue bar on the top of the heatmap. Each class of primary melanomas and metastasis samples were assigned with the same class numbers as in Figure 3, which are indicated by the red, grey, and blue bars at the bottom of the heat map.

A

NAME	ES	NES	NOM p-val	FDR q-val	FWER p-val
UP	0.54	2.2	0.00	0.00	0.00
DOWN	0.21	0.86	0.73	0.74	0.79

B

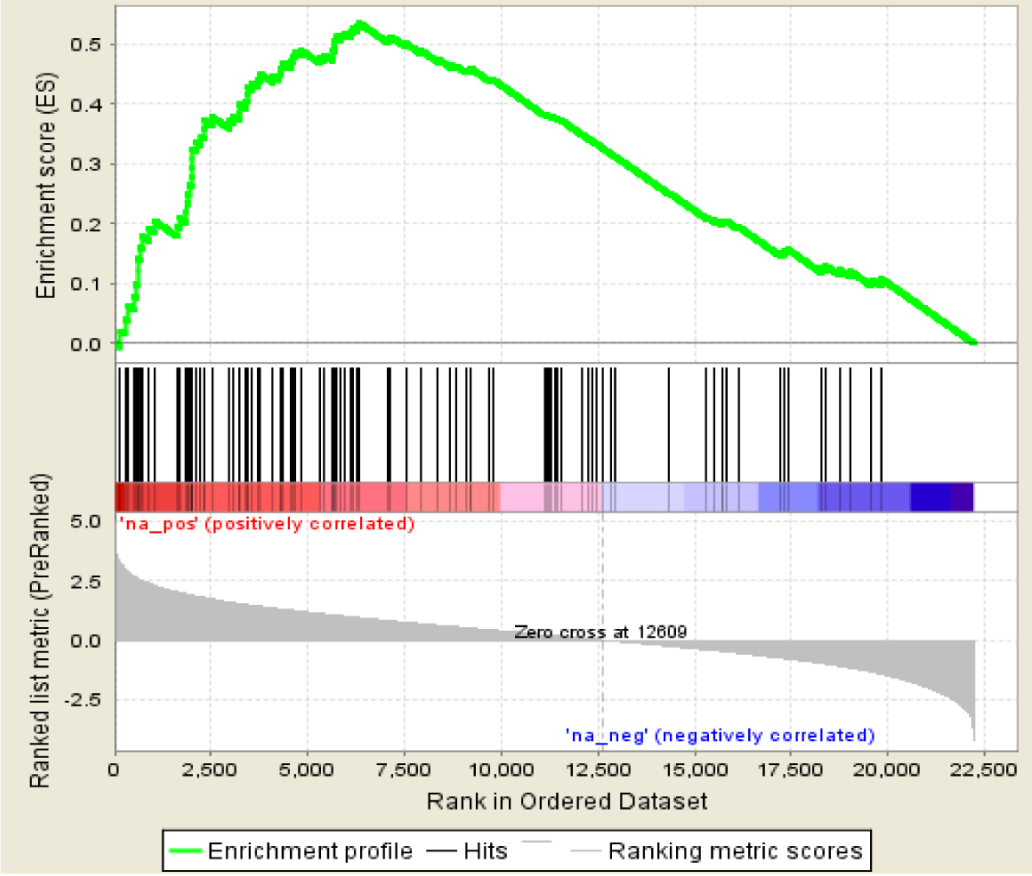


Figure 5. Gene Set Enrichment Analysis shows that up-regulated genes of the 150 correlate significantly with poor survival of patients carrying metastases

The ~22,000 probe sets on HU133A Affymetrix DNA chips were ranked by Cox scores based on their correlation with the survival of the melanoma metastasis-bearing patients. This pre-ranked list of probe sets was used as the template to assess the enrichments of the 99 up-regulated probe sets and 86 down-regulated probe sets from our signature. The up-regulated probe sets (74 genes) were found to be significantly enriched and to correlate with poor survival of the metastasis-positive patients. A. Various scores to assess the enrichment of each gene set in the pre-ranked list of probe sets are shown. The definition of each score is described in the legend of Figure 2. A gene set with an FDR score < 0.25 is considered significantly enriched. B. The enrichment of the up-regulated gene set is shown schematically. The X-axis of the curve for enrichment scores includes the ~22,000 probe sets on the HU133A chip, with those correlating best with poor survival on the left and those correlating best with good survival on

the right (shown in the Ranked List Metric below the curve). Each probe in the gene set is shown as a vertical line underneath the X-axis of the curve and the cumulative enrichment score is plotted as the green curve reaching a maximum enrichment at a score of 0.54. The rank order and the contribution of each of the 74 up-regulated genes to enrichment (shown as Core Enrichment) are listed in Supplementary Document S7.

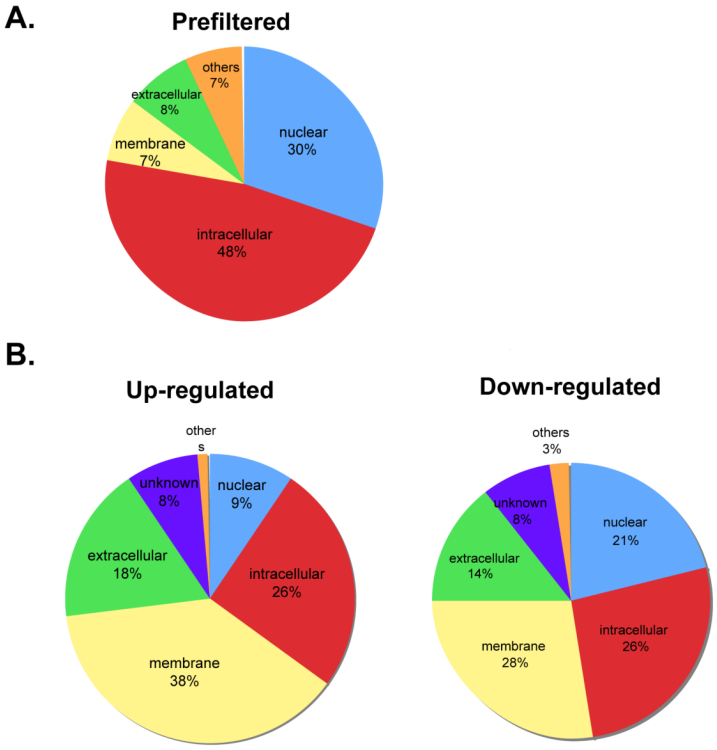


Figure 6. The 150 genes include many genes encoding secreted proteins

A. Before applying the selection criteria of a fold change >3 and maxT < 0.01, all the genes that passed the preprocessing step (corresponding to 9108 probe sets in total; see Materials and Methods) were categorized based on their known or predicted cellular distributions in the GeneOntology Database.

B. The 150 genes were categorized based on their known or predicted cellular distributions in GeneOntology Database. External proteins, including secreted extracellular and membrane proteins, represent 56% of the up-regulated genes and 42% of the down-regulated genes.