nature genetics

Sipa1 is a candidate for underlying the metastasis efficiency modifier locus *Mtes1*

Yeong-Gwan Park[1,4,5], Xiaohong Zhao[1,5], Fabienne Lesueur[2,4], Douglas R Lowy[3], Mindy Lancaster[1], Paul Pharoah[2], Xiaolan Qian[3] & Kent W Hunter[1]

We previously identified loci in the mouse genome that substantially influence the metastatic efficiency of mammary tumors. Here, we present data supporting the idea that the signal transduction molecule, *Sipa1*, is a candidate for underlying the metastasis efficiency modifier locus *Mtes1*. Analysis of candidate genes identified a nonsynonymous amino acid polymorphism in Sipa1 that affects the Sipa1 Rap-GAP function. Spontaneous metastasis assays using cells ectopically expressing *Sipa1* or cells with knocked-down *Sipa1* expression showed that metastatic capacity was correlated with cellular *Sipa1* levels. We examined human expression data and found that they were consistent with the idea that *Sipa1* concentration has a role in metastasis. Taken together, these data suggest that the *Sipa1* polymorphism is one of the genetic polymorphisms underlying the *Mtes1* locus. This report is also the first demonstration, to our knowledge, of a constitutional genetic polymorphism affecting tumor metastasis.

The process of metastasis is important in the clinical management of cancer, because most cancer mortality is attributed to metastatic disease rather than the primary tumor[1]. In most cases, individuals with cancer with localized tumors have better prognoses than those with disseminated tumors. But even people with no evidence of tumor dissemination at primary diagnosis are at risk for metastatic disease. Approximately one-third of women with breast cancer who are sentinel lymph node–negative at the time of surgical resection of the primary breast tumor will later develop clinically detectable secondary tumors[2]. Early identification of these individuals might alter management of their disease and improve their prognosis.

To gain a better understanding of the many factors that modulate metastatic progression, our laboratory initiated an investigation into the effects of constitutional genetic polymorphism on metastatic efficiency. Using the polyoma middle-T (PyMT) transgene-induced mouse mammary tumor model[3], we showed that the genetic background on which a tumor arose influenced its ability to form pulmonary metastases[4]. Quantitative trait genetic mapping analysis identified a probable metastasis efficiency locus, called *Mtes1*, on proximal mouse chromosome 19 (ref. 5) in a 10-Mb region orthologous to human chromosome 11q12–13, which contains the metastasis suppressor gene *Brms1* (ref. 6). Extensive sequence analysis of mouse *Brms1* did not identify any polymorphisms associated with metastatic efficiency suppression[7], indicating that the causative polymorphism was probably associated with another linked gene.

To identify other potential candidates for underlying *Mtes1*, a multiple cross mapping (MCM) strategy was used. The MCM strategy[8] takes advantage of shared haplotypes among different inbred strains of mice used in genetic mapping studies to reduce the number of potential candidate genes[9] in a given candidate interval. Using this methodology, a medium-resolution map of the 10-Mb region of mouse chromosome 19 was constructed[10] across four inbred strains used to map the *Mtes1* locus. Two strains with high metastatic efficiency (AKR/J and FVB/NJ) and two with low metastatic efficiency (DBA/2J and NZB/B1NJ) were included in the analysis. Identification of five haplotypes blocks that segregated appropriately across the inbred strains reduced the number of high-priority candidate genes from ~500 to 23 (ref. 10), a more tractable number for further characterization.

Here we describe further analysis of the potential *Mtes1* candidate genes identified in previous studies. Using a combination of bioinformatics, sequence analysis and *in vitro* and *in vivo* experiments, we identified the signal-induced proliferation-associated gene 1 (*Sipa1*, also called *Spa1*; ref. 11) as a strong candidate for underlying the *Mtes1* locus.

RESULTS

Sequence analysis of candidate genes

We previously used a multiple cross and haplotypes mapping strategy[8,12] to identify regions of mouse chromosome 19 that were most likely to contain the gene underlying *Mtes1* (ref. 10). We prioritized

[1]Laboratory of Population Genetics, National Cancer Institute, Building 41, Room 702, 41 Library Drive, Bethesda, Maryland 20892, USA. [2]Department of Oncology, University of Cambridge, Hutchinson/MRC Research Centre, Hills Road, Cambridge, CB2 2XZ, UK. [3]Laboratory of Cellular Oncology, National Cancer Institute, Bethesda, Maryland 20892, USA. [4]Present addresses: Exam. Div. of Food & Biological Resources, Korean Intellectual Property Office, Room 903, Gov. Complex Daejeon Building 4, 130 Seonsaro, Seo-gu, Daejeon 302-701, Korea (Y.-G.P.); C.N.R.G., 2 rue Gaston Cremieux, CP 5721, 91057 Evry Cedex, France (F.L.). [5]These authors contributed equally to this work. Correspondence should be addressed to K.W.H. (hunterk@mail.nih.gov).

Received 3 May; accepted 11 July; published online 4 September 2005; doi:10.1038/ng1635

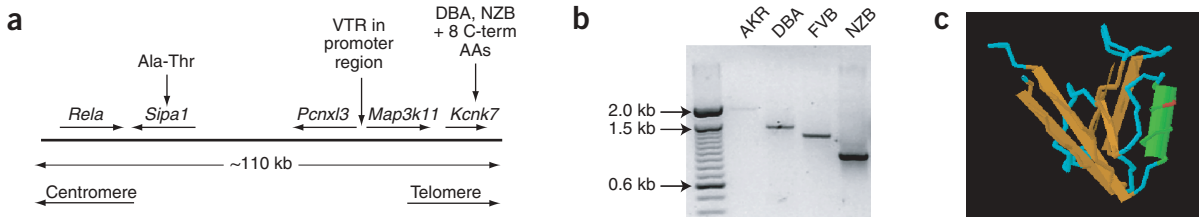

Figure 1 Analysis of the *Sipa1* candidate region. (**a**) Schematic representation of the *Mtes1* or *Sipa1* candidate haplotype interval. The chromosome is represented by the solid horizontal bar, with the centromere to the left and telomere to the right. The genes and their orientations are indicated by the arrows above the bar. Potentially important polymorphisms are depicted above the genes. (**b**) Size of the *Map3k11* VTR in four strains of mice as determined by PCR amplification. Size of the VTR does not correlate with metastatic potential, as the size of the repeat in the strain with low metastatic efficiency (DBA/2J) is intermediate between the sizes of the repeats in the two strains with high metastatic efficiency (AKR/J and FVB/NJ). (**c**) Three-dimensional model of the PDZ domain to identify the location of the polymorphism. The A739T polymorphism is represented by the red section of the wire diagram on the open face of the α-helical region.

five regions containing ~23 of the ~500 genes present in the original candidate interval for further characterization. We categorized these prioritized genes on the basis of their molecular function, analyzing genes of known function before attempting to analyze the more numerous genes of unknown function. We selected six genes in four of the five haplotype blocks for the first round of analysis. The fifth haplotype block did not contain any genes of interest according to existing knowledge. Three genes in the intervals (*Rhod*, *Map3k11* (also called *Mlk3* or *Kcnk7*) and *Sipa1*) were members of a signal transduction pathway associated with metastatic progression[13]. None of the prioritized genes were previously associated with metastasis. We also considered three additional genes to be high-priority candidates because of their roles in global gene expression (*Htatip* and *Rela*) or DNA damage repair function (*Ddb1*). Three of these six genes (*Rela*, *Map3k11* and *Sipa1*) map within the same haplotype block (**Fig. 1a**). Two other genes (*Pcnxl3* and *Kcnk7*) also lie in this haplotype block, but these genes are not part of pathways previously associated with metastatic progression or are transcribed exclusively in neuronal tissues and were therefore excluded from the initial analysis. We sequenced the exons, intron-exon boundaries and promoters, as well as 1–2 kb of the regions immediately upstream of the promoters, in the two strains of mice with high metastatic efficiency (AKR/J and FVB/NJ) and the two strains with low metastatic efficiency (DBA/2J and NZB/B1NJ) that were used to construct the haplotypes map[10]. Although we found intronic and silent polymorphisms in *Rela*, *Ddb1*, *Htatip* and *Rhod*, we observed no obvious functionally relevant polymorphisms in the protein-coding, promoter or upstream regions. Therefore, we excluded these genes from further consideration.

Sequence analysis of the four inbred strains identified a variable tandem repeat (VTR) in the upstream region of *Map3k11*, containing a variable number of the MMS41 minisatellite tandem repeats[14]. Because VTRs may be associated with alterations in transcriptional activity[15,16], we determined the structure of the VTR in the four inbred strains and its effect on the transcriptional and translational activity. PCR amplification across the VTR showed that the number of repeats did not segregate with metastatic efficiency in these four strains (**Fig. 1b**). In addition, quantitative RT-PCR and western-blot analyses did not detect substantial differences in either the mRNA levels or the protein levels between the strains (data not shown). *Map3k11* was therefore excluded from further analysis.

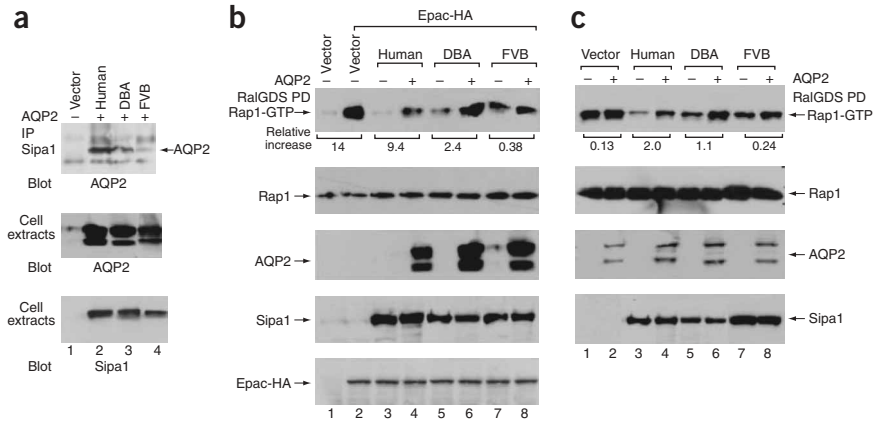

Figure 2 The A739T polymorphism affects complex formation with AQP2 and the Rap-GAP function of *Sipa1* in an AQP2-dependent manner. (**a**) Coimmunoprecipitation of AQP2 and Sipa1. AQP2 was transiently coexpressed with human SIPA1 or mouse Sipa1 from DBA or FVB in COS7 cells. In the upper panel, anti-Sipa1 immunocomplexes were immunoblotted with antibodies to AQP2. The middle panel shows the expression of AQP2 and Sipa1 in cell extracts, and the lower panel shows the expression of Sipa1 in cell extracts. (**b**,**c**) Influence of AQP2 on the Rap-GAP activity of Sipa1 in COS7 (**b**) and U373MG (**c**) cells. Both lines have very low levels of endogenous Sipa1 and AQP2. Human SIPA1 or mouse Sipa1 from DBA or FVB was expressed transiently in COS7 cells (**b**) or as stable clones in U373MG cells (**c**). Hemagglutinin (HA)-tagged Epac was added transiently to the COS7 cells to raise their level of GTP-Rap1. AQP2 was also transiently expressed in COS7 and U373MG cells. A preliminary experiment (not shown) indicated that transfection of *AQP2* alone into parental COS7 cells did not alter the endogenous level of GTP-Rap1. The upper panels show the *in vivo* level of GTP-Rap1, which was assayed by RalGDS pull-down, as an indicator of the relative Rap-GAP activity in the presence or absence of AQP2. The relative signal densities of GTP-Rap1 were quantified by Scion Image. The numbers under the upper panels refer to the relative increase in GTP-Rap1 for a particular *Sipa1* allele in the presence of AQP2. The relative increase is 1 less than the relative difference. Thus, a twofold difference is designated a onefold increase. The remaining panels are immunoblots of cell extracts that document the relative level of the indicated protein. The Rap1 immunoblots also serve as loading controls. IP, immunoprecipitation; PD, pull-down.

Analysis of *Sipa1* identified a polymorphism that results in substitution of an alanine (in strains NZB and DBA) with a threonine (in strains FVB and AKR) in a PDZ domain, a motif frequently implicated in protein-protein interactions[17]. We previously suggested that *Mtes1* was associated with the AKR/J allele rather than the DBA/2J allele[7]. Additional analysis[13] showed that this was an error and that the suppression was in fact associated with the DBA/2J allele. We modeled the polymorphism using the software package Cn3D[18] and found that the polymorphism occurs in the open face of an α helix (**Fig. 1c**). BLAST analysis[19] showed that the alanine was conserved between mouse and human, suggesting that the threonine substitution might have a functional consequence for protein-protein interactions. Because of the potential functional consequence of this polymorphism, we considered *Sipa1* to be a candidate for underlying *Mtes1* and subjected it to further characterization.

Functional analysis of *Sipa1* polymorphism

Sipa1 is a mitogen-inducible gene that encodes a GTPase-activating protein (GAP) specific for Rap1 and Rap2 GTPases, which are members of the superfamily of Ras-related proteins[20]. Sipa1 therefore negatively regulates Rap through its Rap-GAP activity, which catalyzes the hydrolysis of active GTP-Rap to inactive GDP-Rap. An interaction between the products of human *SIPA1* and *AQP2*, a gene implicated in genetically determined nephrogenic diabetes insipidus, was recently identified[21]. AQP2 is the first protein known to bind the PDZ domain of SIPA1, although the possible influence of this binding on the Rap-GAP activity of SIPA1 has not been examined.

Using AQP2 as a marker protein to monitor binding with the PDZ domain of SIPA1 or Sipa1, we sought to determine whether the efficiency of the interaction might be influenced by the polymorphism identified in the PDZ domain of Sipa1. We transiently transfected COS7 cells, which express very low levels of endogenous AQP2 and Sipa1, with *AQP2* and with the DBA allele of *Sipa1*, the FVB allele of *Sipa1* or the human *SIPA1* allele (**Fig. 2a**). Like the DBA allele, the human allele encodes alanine at the codon corresponding to the polymorphic site in the PDZ domain of Sipa1. When extracts of the transfected cells were immunoprecipitated with an antibody to SIPA1 and then immunoblotted with an antibody to AQP2, the extracts containing the human allele or the *Sipa1* DBA allele gave a stronger signal than the extracts containing the FVB allele (**Fig. 2a**). Therefore, AQP2 binds the protein encoded by the DBA allele more efficiently than that encoded by the FVB allele.

We then examined whether the *Sipa1* polymorphism might influence its *in vivo* Rap-GAP activity, in the presence or absence of AQP2, in transiently transfected COS7 cells (**Fig. 2b**). Mock-transfected cells had low levels of GTP-Rap1 (**Fig. 2b**). Therefore, to be able to assess the Rap-GAP activity of Sipa1 in COS7 cells, we cotransfected the cells with *Sipa1* and *Epac*, which encodes a Rap-specific guanine nucleotide exchange factor, which should increase the level of GTP-Rap1 (**Fig. 2**). As expected, *Epac* added without *Sipa1* increased the level of GTP-Rap1 (**Fig. 2b**), but the level of GTP-Rap1 was reduced when *Sipa1* was added together with *Epac*. Under these conditions, the human *Sipa1* allele was more active than either of the mouse alleles in reducing the level of GTP-Rap1, and the DBA allele was somewhat more active than the FVB allele. We assessed the influence of *AQP2* on *in vivo* Rap-GAP activity by cotransfecting COS7 cells with *AQP2* and with *Epac* and a *Sipa1* allele. In cells expressing the human or DBA *Sipa1* allele, the presence of AQP2 was associated with roughly a ninefold or a two- to threefold increase, respectively, in the level of GTP-Rap1 compared with the corresponding control cells lacking AQP2 (**Fig. 2b**). By contrast, in cells expressing the FVB allele, *AQP2* induced less than a 0.5-fold increase.

To confirm that these results could be extended to another cell line, we made stable transfectants of the *Sipa1* alleles in U373MG, a human astrocytoma line (**Fig. 2c**). We chose this cell line because it contained high levels of endogenous GTP-Rap1 (**Fig. 2c**) and low levels of endogenous Sipa1 (data not shown). In cells stably transfected with *Sipa1*, each of the mouse alleles induced a modest reduction in GTP-Rap1, whereas the human allele induced a more substantial reduction. Transient transfection of these cells with *AQP2* resulted in increases in the levels of GTP-Rap1 of roughly twofold in the human allele transfectants and onefold in the DBA allele transfectants. By contrast, AQP2 resulted in levels that were only marginally increased (0.25-fold) in cells expressing the FVB allele.

Taken together, these experiments indicate that expression of *AQP2* induces a reduction in the Rap-GAP activity of Sipa1. The increased binding of *AQP2* to the DBA allele of *Sipa1* relative to the FVB allele results in a greater inhibition of Sipa1 activity, leading to higher concentrations of activated Rap1 (GTP-Rap1) in the DBA transfectants compared with the FVB transfectants (**Fig. 2b**,**c**).

Knock-down of *Sipa1* reduces metastatic capacity

To test whether reduction of Sipa1 concentrations reduced metastatic capacity, as predicted by the dominant suppressor hypothesis, we carried out RNA interference (RNAi) knock-down experiments. We were unable to establish cell lines that stably expressed short hairpin RNA (shRNA) with multiple independent single constructs that would knock down mRNA and protein levels for more than ~10 d (data not shown). Therefore, we transfected cells from the Mvt1 line, a highly metastatic mammary tumor cell line derived from a MMTV-myc MMTV-vegf double transgenic mouse on an FVB/N genetic background[22], with two shRNA constructs and isolated individual colonies, resulting in a cell line with a stable reduction of roughly fourfold in *Sipa1* mRNA and protein levels (**Fig. 3**). We implanted the shRNA knock-down and empty vector control cell lines subcutaneously into virgin FVB/NJ female mice, allowed tumors to grow for four weeks and then examined the lungs for metastases. Macroscopic pulmonary surface metastases were significantly ($P < 0.002$) reduced in the shRNA knock-down cell line (**Fig. 4a**), suggesting that *Sipa1* promotes metastasis. The reduction in pulmonary metastases could not be explained by a nonspecific reduction of the proliferative ability of the shRNA knock-down cells, as the primary tumors from the shRNA knock-down line were approximately two times larger than those from the control cell line (**Fig. 4b**). The increased growth potential of the *Sipa1* RNAi cell line was unexpected, as no substantial difference in

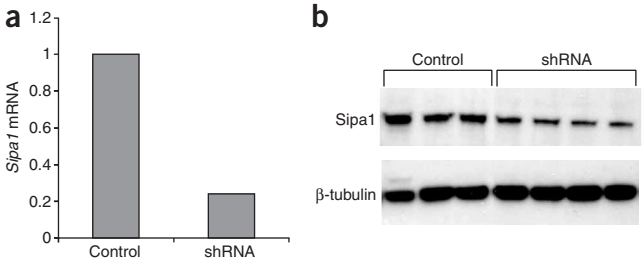

Figure 3 shRNA knock-down of *Sipa1*. (**a**) The relative abundance of *Sipa1* mRNA in the empty vector control cell line and the shRNA cell line. (**b**) Western-blot analysis showing the reduction of *Sipa1* protein levels in the shRNA cells compared with the control cells. β-tubulin was used as a loading control.

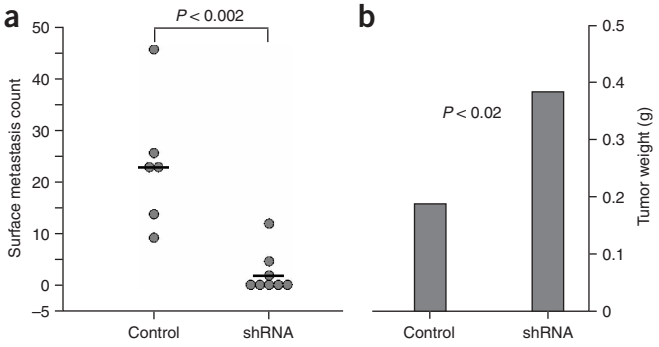

Figure 4 Results of shRNA analysis. (**a**) Scatter plot of the results of the shRNA experimental metastasis assay. (**b**) Graphical representation of the tumor weights of the control and shRNA cell lines in the experimental metastasis assay. Error bars represent s.d.

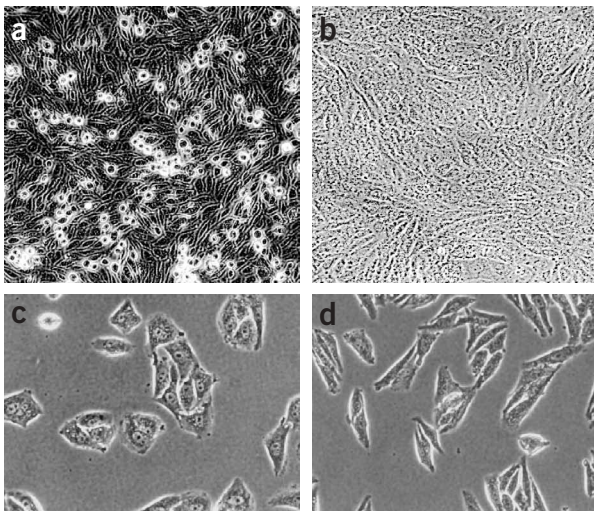

Figure 5 Morphology of the shRNA cell line. Photomicrographs of the growth properties and morphology of the empty vector (**a**,**c**) versus the shRNA (**b**,**d**) cell lines. The control Mvt1 cell line produces many rounded cells when grown to confluence compared with the flat sheet observed in the shRNA knock-down cell line. The shRNA knock-down cell line also has a more spindle shaped morphology than the more metastatic empty vector control.

growth potential between the FVB and (DBA × FVB) F_1 hybrids was observed in the original breeding study[4,5], suggesting that other modifiers might mask this phenotype in the *in vivo* experiment[4,5]. To determine whether the effects of the RNAi experiments might be due to off-target effects, we examined gene expression data for evidence of interferon induction. We observed no obvious off-target effects (**Supplementary Note** and **Supplementary Table 1** online).

To determine whether there might be compensation for loss of *Sipa1* in the cell, we determined the relative intracellular concentrations of the three Rap1 GTPase mRNAs (*Sipa1*, *Sipa1l1* (also called *E6TP1*; ref. 23) and *Rap1ga1* (ref. 24)) by quantitative RT-PCR. We observed equivalent amounts of *Sipa1* and *Rap1ga1* mRNAs and six- to eightfold more *Sipa1l1* mRNA (data not shown; **Supplementary Note**).

shRNA alters cell morphology and adhesive properties

Overexpression of *Sipa1* induces detachment of cells from the matrix[11]. Therefore, we examined the RNAi cells to determine whether suppression of *Sipa1* levels resulted in increased cell adhesion. Mvt1 control cells transfected with empty vector and grown to confluence consisted of loose sheets of cells with round cells above the monolayer (**Fig. 5a**). *Sipa1* knock-down cells, in contrast, formed a solid monolayer without rounded or loosely attached cells (**Fig. 5b**), suggestive of strong cellular adherence. High-power magnification of the cells also showed a substantial difference in cellular morphology between the clones: Mvt1 control cells had an epithelial morphology, whereas *Sipa1* knock-down cells were more spindle-shaped (**Fig. 5c**,**d**).

Sipa1 overexpression increases metastatic capacity

If reduction of *Sipa1* activity, by either a dominant suppressor allele or a reduction in intracellular level, suppresses metastasis, then overexpression of *Sipa1* should enhance metastatic potential. To test this hypothesis, we generated cell lines that overexpress *Sipa1*. If the reduced rate of metastasis observed in *Sipa1* knock-down cell lines is attributable to specific effects on *Sipa1* rather than to nonspecific shRNA-induced interferon induction[25] or off-target effects[26], then overexpression of *Sipa1* should enhance the metastatic potential of Mvt1 cells. To test this, we transfected the Mvt1 cell line with a construct encoding the *Sipa1* FVB allele with an epitope tag, isolated a clone expressing this construct and implanted it subcutaneously into FVB female mice. We collected lungs after 4 weeks and counted the surface lesions. The cell line ectopically overexpressing *Sipa1* developed roughly twice as many surface pulmonary metastases as cell line transfected with empty vector (36.6 versus 16.6; $P = 0.0004$; **Fig. 6**).

Sipa1 in human metastatic progression

To determine whether *Sipa1* has a role in human metastatic progression, we carried out a meta-analysis of human gene expression data. We compared the relative expression of *Sipa1* in metastatic versus nonmetastatic tumors using the Oncomine website. Data were available from prostate[27,28], lung[29,30] and a meta-analysis of a variety of solid tumors[31]. Consistent with the data in the mouse breast cancer model, overexpression of *Sipa1* was associated with metastatic progression in human prostate cancer ($P = 0.001$; **Fig. 7**).

DISCUSSION

Many genes have been associated with metastatic progression. Most were identified by differential expression between the primary tumors and the disseminated lesions, mediated by loss of heterozygosity[32],

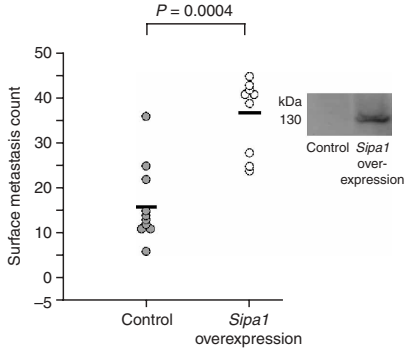

Figure 6 Scatter plot of the lung surface metastasis counts of mice implanted with cells ectopically overexpressing *Sipa1*. The western blot shows expression of Sipa1 tagged with the V5-His6 epitope.

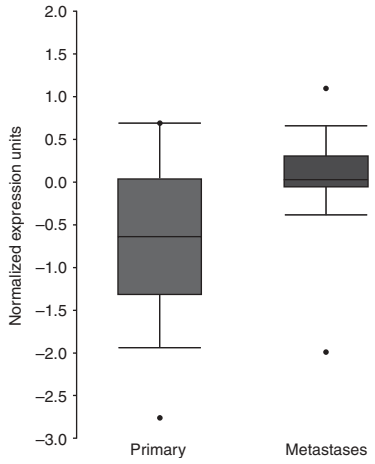

Figure 7 Graphical results of the Oncomine meta-analysis. The Oncomine website was used to analyze whether *Sipa1* was differentially expressed in metastatic versus nonmetastatic tumors.

epigenetic modulation[33] or other somatic events. These events are thought to influence metastatic progression by increasing the activity of metastasis-promoting genes or downregulating or abrogating the inhibitory effects of metastasis-suppressing factors. The mouse has been an integral component in the identification of these genes. Introduction of candidate metastasis-associated genes into cell culture followed by introduction of the cells into the lateral tail vein or ectopic or orthotopic implantation is used routinely to assess the effects of particular genes in the establishment of pulmonary metastases. It is believed that the lungs are the primary site of metastasis in mice, unlike humans, because tumor cells are unable to pass through the pulmonary capillary beds owing to the narrow diameter of the pulmonary capillaries. Metastasis to other sites, replicating the diverse metastatic target organs in the human population, can be achieved by introduction of cells into the vasculature in postpulmonary sites (*e.g.*, intracardiac, intrasplenic or portal vein injection[34]). These strategies have yielded a wealth of information regarding loss or gain of gene functions and the cellular processes that modulate and accompany metastatic progression. Somatic genomic changes resulting in increased or decreased expression of genes are not, however, the only ways to modulate metastasis-associated gene functions. Polymorphism may also influence gene function. The interaction of functional polymorphisms probably underlies most of the nonhereditary cancer susceptibility in the human population. It should not be unexpected, therefore, that similar susceptibility exists for all aspects of tumorigenesis, including metastatic progression.

This study represents an ongoing effort in our laboratory to apply the MCM strategy in parallel with conventional quantitative trait locus precision mapping efforts to identify and characterize genes that modify metastatic efficiency (**Supplementary Note**). Using this strategy, we identified a polymorphism in the signal transduction gene *Sipa1* as a candidate for underlying the metastasis efficiency modifier locus *Mtes1*. Biochemical analysis indicated that the polymorphism, located in the PDZ domain of Sipa1, can influence the known Rap-GAP function of Sipa1, and experimental manipulation of cellular *Sipa1* mRNA levels affected the ability of a highly metastatic mammary tumor cell line to colonize the lung. These data suggest that the relative functional concentration of *Sipa1*, as determined by its overall protein concentration or its availability to inactivate Rap1, modulates metastatic progression. These data also predict that further reduction of functional *Sipa1* activity in cells homozygous with respect to the DBA allele, compared with the (DBA × FVB) F_1 cells that we used, would further reduce metastatic capacity. But it is not possible to carry out these experiments unambiguously; the limitations of meiotic recombination preclude the specific isolation of the *Sipa1* DBA allele free of surrounding chromosomal DNA in an otherwise homozygous FVB background. Because there is evidence for at least one other potential candidate gene near *Sipa1* contributing to the *Mtes1* locus, the results of a congenic experiment could not unambiguously be ascribed to the *Sipa1* polymorphism. The ideal experiment would be a knock-in, but the lack of FVB-derived embryonic stem cells precludes carrying out this experiment.

The observation that *Sipa1* modulation was effective in a cell line derived from a transgenic mouse other than the PyMT transgenic mouse used in the previous genetic modifier mapping studies[4,5] suggests that the postulated role of *Sipa1* in metastasis is not limited to the particular oncogenic events induced by the PyMT antigen. It remains to be determined how broadly this gene may function in metastatic progression. Our data and the meta-analysis from the Oncomine website suggest that *Sipa1* may be important in more than a single cancer type.

The role of *Sipa1* in human populations will also require additional studies to resolve. A number of polymorphisms in *SIPA1* are present in the public databases, including two nonsynonymous missense mutations near the RAP-GAP domain. Association studies in the Anglian breast cancer cohort[35,36] identified a possible weak, but not statistically significant, risk of distant metastases in this cohort (data not shown). But *SIPA1* has not yet been entirely resequenced, and so additional haplotypes may need to be explored to determine whether polymorphisms in *SIPA1* have a role in human breast cancer. The Oncomine data also suggest a possible role in prostate cancer, and studies are underway to validate this result and to examine other solid tumor types.

The precise mechanism by which *Sipa1* modulates metastatic efficiency is not known, but our results provide some clues. *Sipa1* was originally cloned as a mitogen-inducible protein[37] that was subsequently shown to be a negative regulator of Rap1 by serving as a GAP for Rap1 (ref. 20). *Sipa1* affects cellular adhesion[11], primarily related to its effects on Rap1, which has been implicated in maintaining the integrity of polarized epithelial[38] and intercellular adherens junctions[39]. Consistent with these activities, we found that knockdown of *Sipa1* increased the adhesive properties of Mvt1 cells and eliminated the rounded cells observed in the untransfected cells and empty vector controls (**Fig. 5**). Notably, the RNAi knock-down cells also acquired a more spindle-shaped morphology, as compared with the more cuboidal morphology of the control cells, similar to morphological changes observed in other metastasis-suppressed cell systems[40]. Rap1 is also implicated in maintaining the integrity of polarized epithelial[38] and maintenance of intercellular adherens junctions[39]. Part of *Sipa1*'s role in metastasis may therefore be potentiating tumor cells escaping from the primary tumor by modulating cell morphology and the strength of intercellular contacts. In this scenario, the FVB *Sipa1* allele would be more active than the DBA allele in promoting lung metastases in the PyMT model, presumably through regulation of Rap1. One might anticipate that modulating cellular adhesion could result in different histologies between tumors with high and low metastatic efficiency. But examination of the tumors by hematoxylin and eosin staining showed no obvious histological differences (R. Cardiff, personal communication), suggesting either that changes in adhesion or cell morphology are not the primary

mechanism of *Sipa1* metastatic suppression or that the alterations in adhesion or cell morphology do not substantially influence structures at this level of resolution.

In addition to activities that promote metastases, *Sipa1* may have effects that curtail cell growth and tumor formation, just as Rap1 has activities that may stimulate or inhibit growth[41]. *Sipa1* affects cell cycle progression[37,42] and interacts with a bromodomain protein, Brd4; alterations in the relative ratio of these two proteins disrupted normal cell cycle proliferation[42]. In some contexts, the negative growth functions of Sipa1 may predominate. For example, *Sipa1* homozygous knockout mice are viable but eventually develop a myeloproliferative stem cell disorder[43]. The different allelic variants may therefore differentially affect the ability of disseminated cells to enter the cell cycle and proliferate into macroscopic lesions. We have not yet explored this possible effect of the *Sipa1* polymorphism.

Our strategy coupled MCM with existing knowledge of gene function to prioritize genes for analysis. Although this approach seems to have been successful in identifying *Sipa1* as either underlying the *Mtes1* locus or forming a component of this locus, there is no reason to believe that genes of unknown function or known genes in pathways not previously associated with metastasis are not also potential candidates. Our goal in using this method was to eliminate the most obvious and easily testable genes before tackling the more difficult collection of genes of unknown function or functions not thought to be cancer-related. Using other methods to prioritize genes (*e.g.*, gene expression[12]) would probably identify additional candidate genes. Recent efforts in our laboratory to investigate the *Mtes1* locus further, by this and other methods, have identified a second candidate gene of unknown function whose expression correlates with metastatic potential that may contribute to the *Mtes1* locus. We are currently investigating the role of this uncharacterized gene in metastatic efficiency.

In conclusion, our data implicate *Sipa1* as a strong candidate for underlying *Mtes1*. To the best of our knowledge, this is the first demonstration that a constitutional genetic polymorphism can influence the metastatic process and strongly supports the idea that genetic background is an important determinant in metastatic progression[44–46]. Identification of *Sipa1* as a component of the metastatic cascade provides insight into the molecular mechanisms that function during tumor progression and provides an additional target to perturb in our efforts to control or eradicate disseminated disease, or one that might eventually serve as a marker for primary tumors at high risk of metastasis.

METHODS

Primers. The sequences of all oligonucleotide primers used in this study are available in **Supplementary Table 2** online.

Sequence analysis. We designed sequencing primers using the Primer 3 software package (S. Rozen and H. Skaletsky, Whitehead Institute of Biomedical Research, Cambridge, Massachusetts). We designed primers in intronic sequences to flank exons of interest where possible. We generated PCR products under standard amplification conditions and purified them with Qiagen PCR purification kits. We carried out double-strand sequencing with a Perkin Elmer BigDye Dye Terminator sequence kit. Sequencing was done on a Perkin Elmer 3100 Automated Fluorescent Sequencer. We compiled and analyzed sequences with the computer software packages PHRED and PHRAP[47].

Quantitative RT-PCR. We transcribed mRNAs into cDNA using ThermoScript RT-PCR System (Invitrogen) in accordance with the manufacturer's protocol. We used SYBR Green Quantitative PCR to detect *Sipa1* mRNA levels using an ABI PRISM 7900HT Sequence Detection System. We normalized mRNA levels to *Gapdh* mRNA levels using primers purchased from Applied Biosystems.

Construction of plasmids. We obtained the FVB *Sipa1* construct (IMAGE Clone ID 5324326) from Invitrogen. We constructed pcDNA-fl-sipa1-FVB-neo by inserting the IMAGE cDNA into the vector pcDNA3.1/V5-His TOPO (Invitrogen) and constructed pcDNA-fl-sipa1-DBA-neo by replacing the FVB *Xba*I-*Bsr*GI fragment with the corresponding fragment from the DBA cDNA. We confirmed the sequences by sequencing with BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems).

We constructed pcDNA-3.1-V5/His-puro and pcDNA-sf-sipa1-FVB-puro plasmids by digesting pcDNA-3.1-V5/His (Invitrogen), pcDNA-fl-sipa1-DBA-neo and pcDNA-fl-sipa1-FVB-neo plasmids with *Nsi*I and *Bst*Z17I to replace the neomycin-resistance gene with the analogous *Ssp*I and *Nsi*I fragment from the puromycin-resistance gene from pSuper.retro (Oligoengine).

Complex formation between AQP2 and Sipa1. We transiently transfected COS7 cells, using lipofectamine (Invitrogen) in accordance with the manufacturer's instructions, with pcDNA3 vector alone or with pcDNA3-AQP2 (from ATCC) and pSRα-Sipa1 expressing human *SIPA1*, pcDNA-fl-sipa1-DBA-neo expressing DBA *Sipa1* or pcDNA-fl-sipa1-FVB-neo expressing FVB *Sipa1*. Each dish received the same total amount of DNA. Two days after transfection, we lysed cells with Golden Lysis Buffer containing 20 mM Tris (pH 7.9), 137 mM NaCl, 5 mM EDTA, 1 mM EGTA, 10 mM NaF, 10% glycerol, 1 mM sodium pyrophosphate, 1 mM leupeptin, 1 mM PMSF and 10 μg ml^{-1} aprotinin. We immunoprecipitated cell extracts with monoclonal antibody to Sipa1 and added protein A/G (PIERCE) with overnight rotation at 4 °C. We washed the immune complexes once with Golden Lysis Buffer, once with high salt HNTG (20 mM Hepes buffer, 500 mM NaCl, 0.1% Triton-X100 and 10% glycerol), and twice with low salt of HNTG (20 mM Hepes buffer, 150 mM NaCl, 0.1% Triton-X100 and 10% glycerol). We then analyzed the immune complexes by immunoblotting with antibody to AQP2 (Santa Cruz). We also analyzed cell extracts from transfectants for protein expression by immunoblotting with the antibody to AQP2 and a monoclonal antibody to Sipa1 (BD BioSciences). For each blot, we used horseradish peroxidase–conjugated antibodies to rabbit, mouse or goat immunoglobulin G for the second reaction at a 1:10,000 dilution. We visualized immune complexes by enhanced chemiluminescence with an ECL Kit (Amersham).

RalGDS pull-down assay. We transiently transfected COS7 cells as described above, except that we also added a plasmid encoding hemagglutinin-tagged Epac to elevate the level of GTP-Rap1. Alternatively, we also transfected an astrocytoma cell line, U373MG, which has high levels of endogenous GTP-Rap1, with the *Sipa1* alleles to establish stable expressing clones. We then transiently transfected these clones with the AQP2-expressing plasmid or vector control. Two days after transfection, we processed cells using a Rap1 activation kit (Upstate Biotech. Inc.) in accordance with the manufacturer's instructions. GTP-Rap1 protein was pulled down by RalGDS beads, washed three times and subjected to gel analysis and immunoblotting with an antibody to Rap1 (Santa Cruz). We also analyzed cell extracts from transfectants as above for protein expression by immunoblotting with antibody to AQP2, monoclonal antibody to Sipa1 or antibody to hemagglutinin (Convance).

Establishment of stable cell lines expressing allelic variants of *Sipa1*. We used the Mvt1 cell line[22], derived from a primary mammary tumor in an MMTV-vegf MMTV-myc doubly transgenic mouse, to generate stable cell lines expressing the FVB form of *Sipa1*. The plasmids pcDNA-3.1-puro, pcDNA-fl-sipa1-DBA-puro or pcDNA-fl-sipa1-FVB-puro were linearized by *Pme*I digestion and used to transfect Mvt1 cells using Polyfect Transfection Reagent (Qiagen). Twenty-four hours after transfection, we selected the cells in medium containing 10 μg ml^{-1} puromycin, transferred them to 96-well plates and selected individual clones by limiting dilution. We screened colonies by western blotting against V5 antibody to identify clones expressing full-length *Sipa1*.

Sipa1 shRNA assays. We generated two different shRNA constructs against Sipa1 by cloning oligonucleotides into the pSuper.Retro.puro vector (Oligoengine). We transfected cells with these plasmids using FuGene (Roche) in accordance with the manufacturer's suggested protocol and selected cells by adding puromycin to the medium to a final concentration of 10 μg ml^{-1}.

Pulmonary metastasis assays. We cultured cell lines in Dulbecco's modified Eagle Medium (Cellgro) containing 5 μg ml^{-1} puromycin and 50 μg ml^{-1} gentamycin and supplemented with 10% fetal bovine serum. Two days before injection, cells were passaged and permitted to grow to 80–90% confluence. We then washed the cells with phosphate-buffered saline, treated them with trypsin, collected them, washed them twice with cold phosphate-buffered saline, counted them in a hemocytometer and resuspended them at a concentration of 2×10^6 cells per ml. We injected 100 μl of cells subcutaneously in the vicinity of the fourth mammary gland into 6-week-old virgin FVB/NJ female mice. Four weeks later, we killed the mice by anesthetic overdose. We dissected and weighed tumors at the injection site. We isolated lungs and counted the surface metastases using a dissecting microscope. These experiments were done in compliance with the National Cancer Institute's Animal Care and Use Committee Guidelines.

Electrophoresis and western blotting. We treated cells with trypsin, collected them and lysed cell pellets in cell lysis buffer (M-PER Mammalian Protein Extraction Reagent (PIERCE) with Halt Protease Inhibitor cocktail kit (PIERCE)). We carried out SDS-PAGE for 60–90 min at 100 V using the XCell SureLock Mini-Cell (Invitrogen) with NuPAGE Novex Bis-Tris Gels (Invitrogen). We transferred proteins to Immobilon-P membranes (Millipore) and immunoblotted them against V5 to detect exogenous Sipa1 protein or against antibody to Sipa1 to detect both endogenous and exogenous protein.

Quantitative western blotting. We extracted protein from tumor samples using the T-PER Tissue Protein Extraction Reagent (Pierce Biotechnology, Inc.) with Halt Protease Inhibitor Cocktail (Pierce Biotechnology, Inc.) and determined protein concentrations by the BCA assay (Pierce Biotechnology, Inc.). We separated protein samples of each group on SDS-PAGE precast gels (Invitrogen) and then transferred them to Immobilon-P membranes (Millipore). We carried out western blotting against rabbit antibody to Spa-1 (1:1,000; a gift from M. Hattori, Kyoto University) as previously described[48]. We detected the antigen using an ECL-Plus Kit (Amersham) in accordance with the manufacturer's protocol. We quantified protein signals using the one-dimensional quantification function of ImageQuant QL software (Amersham). We calculated the signal of the protein band by integrating local intensity and subtracting local background intensity. We used GAPDH as a loading control.

RNA extraction and processing for Affymetrix GeneChip analysis. We extracted total RNA as described above using TRIzol Reagent (Life Technologies, Inc.) in accordance with the standard protocol. We determined the quantity and quality of the RNA using an Agilent Technologies 2100 Bioanalyzer (Bio Sizing Software version A.02.01, Agilent Technologies) or a GeneQuant Pro (Amersham Biosciences Corp.). We purified samples containing high-quality total RNA with ratios between 1.8 and 2.1 using the RNeasy Mini Kit (Qiagen). An on-column digest was done as part of this purification step using RNase-Free DNase Set (Qiagen). We then pooled purified total RNA for each strain to produce a uniform sample containing 8 μg.

We synthesized double-stranded cDNA from this preparation using the SuperScript Choice System for cDNA Synthesis (Invitrogen) in accordance with to the protocol for Affymetrix GeneChip Eukaryotic Target Preparation. We purified the double-stranded cDNA using the GeneChip Sample Cleanup Module (Qiagen). We synthesized biotin-labeled cRNA by *in vitro* transcription of the purified template cDNA using the Enzo BioArray High Yield RNA Transcript Labeling Kit (T7) (Enzo Life Sciences, Inc.). We purified the cRNA target using the GeneChip Sample Cleanup Module (Qiagen). We prepared hybridization cocktails from each fragmentation reaction in accordance with the Affymetrix GeneChip protocol, applied the hybridization cocktails to Affymetrix Murine Genome Moe430A GeneChip Arrays, processed them on the Affymetrix Fluidics Station 400 and then analyzed them using the RMA function of the BRB-Array Tools software version 3.6 (ref. 49).

URLs. The software package Cn3D is available at http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml. The Oncomine website is http://141.214.6.50/oncomine/main/index.jsp.

Accession codes. BIND identifiers (http://bind.ca): 315307, 315308.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank B. Ponder and the members of the Laboratory of Population Genetics for discussions and A. Papageorge and V. Bliskovsky for technical assistance and discussions. This research was supported in part by the Intramural Research Program of the US National Institutes of Health, the US National Cancer Institute and the Center for Cancer Research.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturegenetics/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Liotta, L.A. & Stetler-Stevenson, W.G. *Principles of Molecular Cell Biology of Cancer: Cancer Metastasis*, 134–149 (J.B. Lippincott Co., Philadelphia, 1993).
2. Heimann, R., Lan, F., McBride, R. & Hellman, S. Separating favorable from unfavorable prognostic markers in breast cancer: the role of E-cadherin. *Cancer Res.* **60**, 298–304 (2000).
3. Guy, C.T., Cardiff, R.D. & Muller, W.J. Induction of mammary tumors by expression of polyomavirus middle T oncogene: A transgenic mouse model for metastatic disease. *Mol. Cell. Biol.* **12**, 954–961 (1992).
4. Lifsted, T. *et al.* Identification of inbred mouse strains harboring genetic modifiers of mammary tumor age of onset and metastatic progression. *Int. J. Cancer* **77**, 640–644 (1998).
5. Hunter, K.W. *et al.* Predisposition to efficient mammary tumor metastatic progression is linked to the breast cancer metastasis suppressor gene *Brms1*. *Cancer Res.* **61**, 8866–8872 (2001).
6. Seraj, M.J., Samant, R.S., Verderame, M.F. & Welch, D.R. Functional evidence for a novel human breast carcinoma metastasis suppressor, BRMS1, encoded at chromosome 11q13. *Cancer Res.* **60**, 2764–2769 (2000).
7. Park, Y.G. *et al.* Comparative sequence analysis in eight inbred strains of the metastasis modifier QTL candidate gene *Brms1*. *Mamm. Genome* **13**, 289–292 (2002).
8. Hitzemann, R. *et al.* Multiple cross mapping (MCM) markedly improves the localization of a QTL for ethanol-induced activation. *Genes Brain Behav.* **1**, 214–222 (2002).
9. Wiltshire, T. *et al.* Genome-wide single-nucleotide polymorphism analysis defines haplotype patterns in mouse. *Proc. Natl. Acad. Sci. USA* **100**, 3380–3385 (2003).
10. Park, Y.G., Clifford, R., Buetow, K.H. & Hunter, K.W. Multiple cross and inbred strain haplotype mapping of complex-trait candidate genes. *Genome Res.* **13**, 118–121 (2003).
11. Tsukamoto, N., Hattori, M., Yang, H., Bos, J.L. & Minato, N. Rap1 GTPase-activating protein SPA-1 negatively regulates cell adhesion. *J. Biol. Chem.* **274**, 18463–18469 (1999).
12. Hitzemann, R. *et al.* A strategy for the integration of QTL, gene expression, and sequence analyses. *Mamm. Genome* **14**, 733–747 (2003).
13. Steeg, P.S. Metastasis suppressors alter the signal transduction of cancer cells. *Nat. Rev. Cancer* **3**, 55–63 (2003).
14. Bois, P. *et al.* Isolation and characterization of mouse minisatellites. *Genomics* **50**, 317–330 (1998).
15. Nakamura, Y., Koyama, K. & Matsushima, M. VNTR (variable number of tandem repeat) sequences as transcriptional, translational, or functional regulators. *J. Hum. Genet.* **43**, 149–152 (1998).
16. Bailly, S., Israel, N., Fay, M., Gougerot-Pocidalo, M.A. & Duff, G.W. An intronic polymorphic repeat sequence modulates interleukin-1 alpha gene regulation. *Mol. Immunol.* **33**, 999–1006 (1996).
17. van Ham, M. & Hendriks, W. PDZ domains-glue and guide. *Mol. Biol. Rep.* **30**, 69–82 (2003).
18. Hogue, C.W. Cn3D: a new generation of three-dimensional molecular structure viewer. *Trends Biochem. Sci.* **22**, 314–316 (1997).
19. Altschul, S.F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
20. Kurachi, H. *et al.* Human *SPA-1* gene product selectively expressed in lymphoid tissues is a specific GTPase-activating protein for Rap1 and Rap2. Segregate expression profiles from a *rap1GAP* gene product. *J. Biol. Chem.* **272**, 28081–28088 (1997).
21. Noda, Y. *et al.* Aquaporin-2 trafficking is regulated by PDZ-domain containing protein SPA-1. *FEBS Lett.* **568**, 139–145 (2004).
22. Pei, X.F. *et al.* Explant-cell culture of primary mammary tumors from MMTV-c-Myc transgenic mice. *In Vitro Cell. Dev. Biol. Anim.* **40**, 14–21 (2004).
23. Gao, Q., Srinivasan, S., Boyer, S.N., Wazer, D.E. & Band, V. The E6 oncoproteins of high-risk papillomaviruses bind to a novel putative GAP protein, E6TP1, and target it for degradation. *Mol. Cell. Biol.* **19**, 733–744 (1999).
24. Gould, K.A. *et al.* Genetic evaluation of candidate genes for the *Mom1* modifier of intestinal neoplasia in mice. *Genetics* **144**, 1777–1785 (1996).
25. Sledz, C.A., Holko, M., de Veer, M.J., Silverman, R.H. & Williams, B.R. Activation of the interferon system by short-interfering RNAs. *Nat. Cell Biol.* **5**, 834–839 (2003).
26. Snove, O., Jr. & Holen, T. Many commonly used siRNAs risk off-target activity. *Biochem. Biophys. Res. Commun.* **319**, 256–263 (2004).

27. Dhanasekaran, S.M. *et al.* Delineation of prognostic biomarkers in prostate cancer. *Nature* **412**, 822–826 (2001).
28. LaTulippe, E. *et al.* Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastatic disease. *Cancer Res.* **62**, 4499–4506 (2002).
29. Bhattacharjee, A. *et al.* Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc. Natl. Acad. Sci. USA* **98**, 13790–13795 (2001).
30. Garber, M.E. *et al.* Diversity of gene expression in adenocarcinoma of the lung. *Proc. Natl. Acad. Sci. USA* **98**, 13784–13789 (2001).
31. Ramaswamy, S., Ross, K.N., Lander, E.S. & Golub, T.R. A molecular signature of metastasis in primary solid tumors. *Nat. Genet.* **33**, 49–54 (2003).
32. Miele, M.E. *et al.* Metastasis suppressed, but tumorigenicity and local invasiveness unaffected, in the human melanoma cell line MelJuSo after introduction of human chromosomes 1 or 6. *Mol. Carcinog.* **15**, 284–299 (1996).
33. Sekita, N. *et al.* Epigenetic regulation of the *KAI1* metastasis suppressor gene in human prostate cancer cell lines. *Jpn. J. Cancer Res.* **92**, 947–951 (2001).
34. Khanna, C. & Hunter, K. Modeling metastasis *in vivo*. *Carcinogenesis* **26**, 513–523 (2005).
35. Pharoah, P.D. *et al.* Polygenic susceptibility to breast cancer and implications for prevention. *Nat. Genet.* **31**, 33–36 (2002).
36. Day, N. *et al.* EPIC-Norfolk: study design and characteristics of the cohort. European prospective investigation of cancer. *Br. J. Cancer* **80** Suppl 1, 95–103 (1999).
37. Hattori, M. *et al.* Molecular cloning of a novel mitogen-inducible nuclear protein with a Ran GTPase-activating domain that affects cell cycle progression. *Mol. Cell. Biol.* **15**, 552–560 (1995).
38. Ohba, Y. *et al.* Requirement for C3G-dependent Rap1 activation for cell adhesion and embryogenesis. *EMBO J.* **20**, 3333–3341 (2001).
39. Yajnik, V. *et al.* DOCK4, a GTPase activator, is disrupted during tumorigenesis. *Cell* **112**, 673–684 (2003).
40. Seftor, E.A. *et al.* Molecular determinants of human uveal melanoma invasion and metastasis. *Clin. Exp. Metastasis* **19**, 233–246 (2002).
41. Stork, P.J. Does Rap1 deserve a bad Rap? *Trends Biochem. Sci.* **28**, 267–275 (2003).
42. Farina, A. *et al.* Bromodomain protein Brd4 binds to GTPase-activating SPA-1, modulating its activity and subcellular localization. *Mol. Cell. Biol.* **24**, 9059–9069 (2004).
43. Ishida, D. *et al.* Myeloproliferative stem cell disorders by deregulated Rap1 activation in SPA-1-deficient mice. *Cancer Cell* **4**, 55–65 (2003).
44. Hunter, K., Welch, D.R. & Liu, E.T. Genetic background is an important determinant of metastatic potential. *Nat. Genet.* **34**, 23–24 (2003).
45. Hunter, K.W. Allelic diversity in the host genetic background may be an important determinant in tumor metastatic dissemination. *Cancer Lett.* **200**, 97–105 (2003).
46. Hunter, K.W. Host genetics and tumour metastasis. *Br. J. Cancer* **90**, 752–755 (2004).
47. Gordon, D., Abajian, C. & Green, P. Consed: a graphical tool for sequence finishing. *Genome Res.* **8**, 195–202 (1998).
48. Le Voyer, T. *et al.* An epistatic interaction controls the latency of a transgene-induced mammary tumor. *Mamm. Genome* **11**, 883–889 (2000).
49. Hedenfalk, I. *et al.* Gene-expression profiles in hereditary breast cancer. *N. Engl. J. Med.* **344**, 539–548 (2001).