The EMBO Journal (2011) 30, 4500–4514 | © 2011 European Molecular Biology Organization | All Rights Reserved 0261-4189/11
www.embojournal.org

THE EMBO JOURNAL

Signalling pathway for RKIP and Let-7 regulates and predicts metastatic breast cancer

Jieun Yun[1,5], Casey A Frankenberger[1], Wen-Liang Kuo[1], Mirjam C Boelens[2], Eva M Eves[1], Nancy Cheng[1], Han Liang[3,6], Wen-Hsiung Li[3], Hemant Ishwaran[4], Andy J Minn[2,7,*] and Marsha Rich Rosner[1,7,*]

[1]Ben May Department for Cancer Research, University of Chicago, IL, USA, [2]Department of Radiation Oncology, Abramson Family Cancer Research Institute, University of Pennsylvania, Philadelphia, PA, USA, [3]Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA and [4]Department of Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH, USA

Tumour metastasis suppressors are inhibitors of metastasis but their mechanisms of action are generally not understood. We previously showed that the suppressor Raf kinase inhibitory protein (RKIP) inhibits breast tumour metastasis in part via let-7. Here, we demonstrate an integrated approach combining statistical analysis of breast tumour gene expression data and experimental validation to extend the signalling pathway for RKIP. We show that RKIP inhibits let-7 targets (HMGA2, BACH1) that in turn upregulate bone metastasis genes (MMP1, OPN, CXCR4). Our results reveal BACH1 as a novel let-7-regulated transcription factor that induces matrix metalloproteinase1 (MMP1) expression and promotes metastasis. An RKIP pathway metastasis signature (designated RPMS) derived from the complete signalling cascade predicts high metastatic risk better than the individual genes. These results highlight a powerful approach for identifying signalling pathways downstream of a key metastasis suppressor and indicate that analysis of genes in the context of their signalling environment is critical for understanding their predictive and therapeutic potential.
The EMBO Journal (2011) **30,** 4500–4514. doi:10.1038/emboj.2011.312; Published online 26 August 2011
Subject Categories: RNA; molecular biology of disease
Keywords: BACH1; gene signature; Let-7; metastasis; RKIP

*Corresponding authors. AJ Minn, Department of Radiation Oncology, Abramson Family Cancer Research Institute, University of Pennsylvania, 421 Curie Boulevard, BRB II/III, Room 510, Philadelphia, PA 19104, USA. Tel.: +1 215 746 2694; Fax: +1 215 746 5511;
E-mail: andyminn@exchange.upenn.edu or
MR Rosner, Ben May Department for Cancer Research, University of Chicago, 929 East 57th Street, W421C, Chicago, IL 60637, USA. Tel.: +1 773 702 0380; Fax: +1 773 702 4476;
E-mail: m-rosner@uchicago.edu
[5]Present address: Bio-evaluation Center, Korea Research Institute of Bioscience and Biotechnology, Ochang, Chungbuk 363-883, Korea
[6]Present address: Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA
[7]These authors contributed equally to this work

Received: 12 March 2011; accepted: 2 August 2011; published online: 26 August 2011

Introduction

Metastasis is a complex multistep process that leads to distant spread of cancer and is responsible for the majority of breast cancer-related mortality (Massague, 2007). However, with the multiplicity of potential tumourigenic genes revealed by new sequencing technologies, it is difficult to identify key downstream mediators of metastasis. Within any one tumour type, there is a spectrum of relatively rare mutant genes rather than a key driving oncogene. Furthermore, the genes that have been identified tend to cluster into discrete signalling pathways, suggesting the importance of understanding genes in the context of the signalling environment. Therefore, any method that enhances our ability to identify signalling pathways compromised in metastatic patients would be an improvement on current approaches.

To facilitate the discovery of metastasis mediators, recent experimentally driven approaches have led to signatures based upon specific oncogene activation (Bild *et al*, 2006), organ-specific metastasis (Kang *et al*, 2003; Minn *et al*, 2005), or well-characterized signalling pathways serving novel roles in metastatic disease (Nguyen *et al*, 2009). Nonetheless, even with the identification of mechanistically meaningful signatures, the genes that comprise them often have ambiguous biological relationships to one another.

One strategy for identifying key metastasis signalling pathways is to elucidate the mechanism of action of metastasis suppressor genes (reviewed in Bodenstine and Welch, 2008). Raf kinase inhibitory protein (RKIP; also PEBP1) is a metastasis suppressor that regulates growth and differentiation in a variety of organisms (reviewed in Granovsky and Rosner, 2008; Zeng *et al*, 2008). RKIP plays a key role as an inhibitor of MAP kinase (MAPK), G-protein-coupled receptor and NFκB signalling cascades (Yeung *et al*, 1999, 2001; Corbit *et al*, 2003; Lorenz *et al*, 2003; Trakul *et al*, 2005), and loss of RKIP promotes genomic instability through suppression of the spindle checkpoint (Eves *et al*, 2006). RKIP expression is decreased in many solid tumours and RKIP suppresses metastasis but not primary tumour growth in prostate and breast tumour xenografts (Fu *et al*, 2003, 2006; Schuierer *et al*, 2004; Hagan *et al*, 2005; Akaishi *et al*, 2006; Al-Mulla *et al*, 2006; Chatterjee *et al*, 2008; Dangi-Garimella *et al*, 2009). We recently identified a signalling pathway by which RKIP inhibits breast cancer invasion, intravasation and bone metastasis in a xenograft murine model (Dangi-Garimella *et al*, 2009). The mechanism involves inhibition of MAPK, leading to suppression of Myc, decreased LIN28 transcription, and enhanced processing of microRNA let-7. The present study extends this signalling cascade by identifying biologically and clinically relevant metastatic regulators that function downstream of let-7 to promote metastasis.

The goal of this study was to establish a mechanistically driven and clinically relevant metastasis signalling pathway for breast cancer. We utilized a bioinformatics-driven

 ©2011 European Molecular Biology Organization

approach based on the metastasis suppressor RKIP to identify downstream *let-7* targets and relate them to specific metastasis genes in human tumours. Notably, our analyses demonstrate the importance of evaluating genes responsible for tumour progression and metastasis in the context of the signalling cascades that regulate them.

Results

A predicted metastasis signalling pathway connects RKIP and let-7 to bone metastasis genes

To investigate the RKIP/let-7 metastasis pathway without limiting analyses to cell lines, we utilized gene expression data from large data sets of breast cancer patients as a bioinformatics-driven method for generating testable hypotheses about pathway connectivity and for clinically testing pathway predictions. For this, we developed a two-step approach to pathway analysis that utilizes gene set analysis (GSA) (Efron and Tibshirani, 2007) and non-parametric multivariable regression (see Materials and methods). This approach was applied to a gene expression data set of 443 breast cancer patients (BrCa443; see Supplementary Figure S1A) and independently validated using a second data set of 871 breast cancer patients (BrCa871, see Figure 1A).

As the first step in pathway analysis, we confirmed a relationship between RKIP and let-7 in primary human breast cancers. Since let-7 is a microRNA, its expression could not be directly interrogated in the majority of available databases. However, we reasoned that RKIP could be correlated to a gene list comprised of high confidence predicted let-7 targets (Wu *et al*, 2007). Let-7 binds to 3′ UTR complementary sites in target mRNAs to inhibit protein synthesis or promote mRNA degradation so its targets on average should show decreased expression in response to increasing RKIP levels. As predicted, GSA confirmed that a significant proportion of 34 high confidence let-7 target genes do have decreased expression in tumours with higher RKIP levels (Figure 1A, see figure legend for *P*-values).

As the second step in pathway analysis, we derived a candidate list of RKIP-regulated let-7 target genes by determining which predicted let-7 targets are most strongly associated with RKIP. The top importance scores, a measurement of significance based on multivariable analysis (Breiman, 2001), identify let-7 target genes that most accurately predict RKIP expression. Let-7 targets with the highest importance scores (Figure 1A) include HMGA2, IGF2BP2 (IMP2), IGF2BP3 (IMP3), BACH1, GOLT1B, MAP4K4, and N-Ras, genes that also have high GSA scores (Supplementary Figure S1A). Among these, HMGA2, IMP, and Ras have been identified as direct let-7 targets (Johnson *et al*, 2005; Lee and Dutta, 2007; Mayr *et al*, 2007; Boyerinas *et al*, 2008).

RKIP suppresses bone metastasis of 1833 cells, a bone-tropic subline of MDA-MB-231 breast cancer cells, through a let-7-dependent mechanism (Dangi-Garimella *et al*, 2009). Previous studies elucidated a bone metastasis signature (BMS) for 1833 cells comprised of ~100 genes (Kang *et al*, 2003). Five BMS genes were validated experimentally as bone metastasis mediators: Chemokine (C-X-C motif) receptor 4 (CXCR4), the integrin-binding glycoprotein osteopontin (OPN; also SPP1), the matrix metalloproteinase1 (MMP1), connective tissue growth factor (CTGF), and interleukin-11 (IL-11). GSA revealed that a significant proportion of BMS genes exhibit decreased expression with respect to RKIP (Figure 1A), suggesting that the RKIP/let-7 pathway controls bone metastasis by inhibiting BMS mediators. Among the individual BMS genes, MMP1, CTGF, OPN, and CXCR4 had high importance scores. No BMS genes were found to be high confidence let-7 target genes.

RKIP pathway relationships were then analysed and validated in a step-by-step fashion. Among the individual high confidence let-7 target genes correlated to RKIP, we previously identified HMGA2. Therefore, we focused on the leucine zipper transcription factor BACH1 because it is a novel let-7 target, and a gene list of potential BACH1 transcriptional targets is available from the TRANSFAC database. To determine if BACH1 is associated with let-7, a 'meta-gene' for let-7 (LET7-TG) was derived using a weighted average of expression values for let-7 target genes correlated with RKIP (Supplementary Table SI; see Supplementary data). As hypothesized, a significant proportion of BACH1 target genes (BACH1-TG) correlated with LET7-TG, consistent with BACH1 being a target of let-7. Notably, MMP1 is a predicted BACH1 transcriptional target and had one of the highest importance scores relating BACH1-TG to LET7-TG (Figure 1A). Further pathway analysis revealed that expression of BMS genes is associated with a meta-gene for BACH1-TG, and confirmed MMP1 as a potential BACH1-regulated BMS gene (Figure 1A).

To confirm the relationship between RKIP and let-7, we examined two combined primary tumour data sets (181 samples) (Blenkiron *et al*, 2007; Enerly *et al*, 2011) that paired both microRNA and mRNA expression data. These results show a very significant correlation between RKIP and, in particular, let-7e (Supplementary Figure S2A; correlation coefficient = 0.20; $P < 0.006$). Other let-7 family members are close to the 95% confidence level for significance (e.g. $P < 0.08$ for let-7g). The association of RKIP with individual let-7 genes appears less robust than that between the let-7 genes as a set presumably because of the multiplicity of let-7 family members that all contribute to the final output. Since we had previously focused on let-7a and let-7g in our cell lines, we analysed let-7e expression in 1833 cells in response to RKIP expression. These results confirm that RKIP induces let-7e at levels comparable to those of let-7a or g (Supplementary Figure S2B). Analysis of predicted target genes for let-7a, let-7g, and let-7e using PicTar (TargetScan does not distinguish between let-7 family members) revealed 93–94% overlap. Furthermore, BACH1 is a predicted target of all three let-7 family members. In total, these results suggest that multiple let-7 family members contribute to the regulation of metastasis in breast tumour cells.

We also did a GSA using LET7-TG as the dependent variable and eight let-7 genes as a gene set and compared this with random gene sets of eight microRNAs. The let-7a–g genes, when analysed either individually or as a set, show a negative correlation with the let-7 target genes (LET7-TG) as expected (Supplementary Figure S2C; $P < 0.05$). These results suggest that LET7-TG is a good surrogate readout for the activity of the let-7 family.

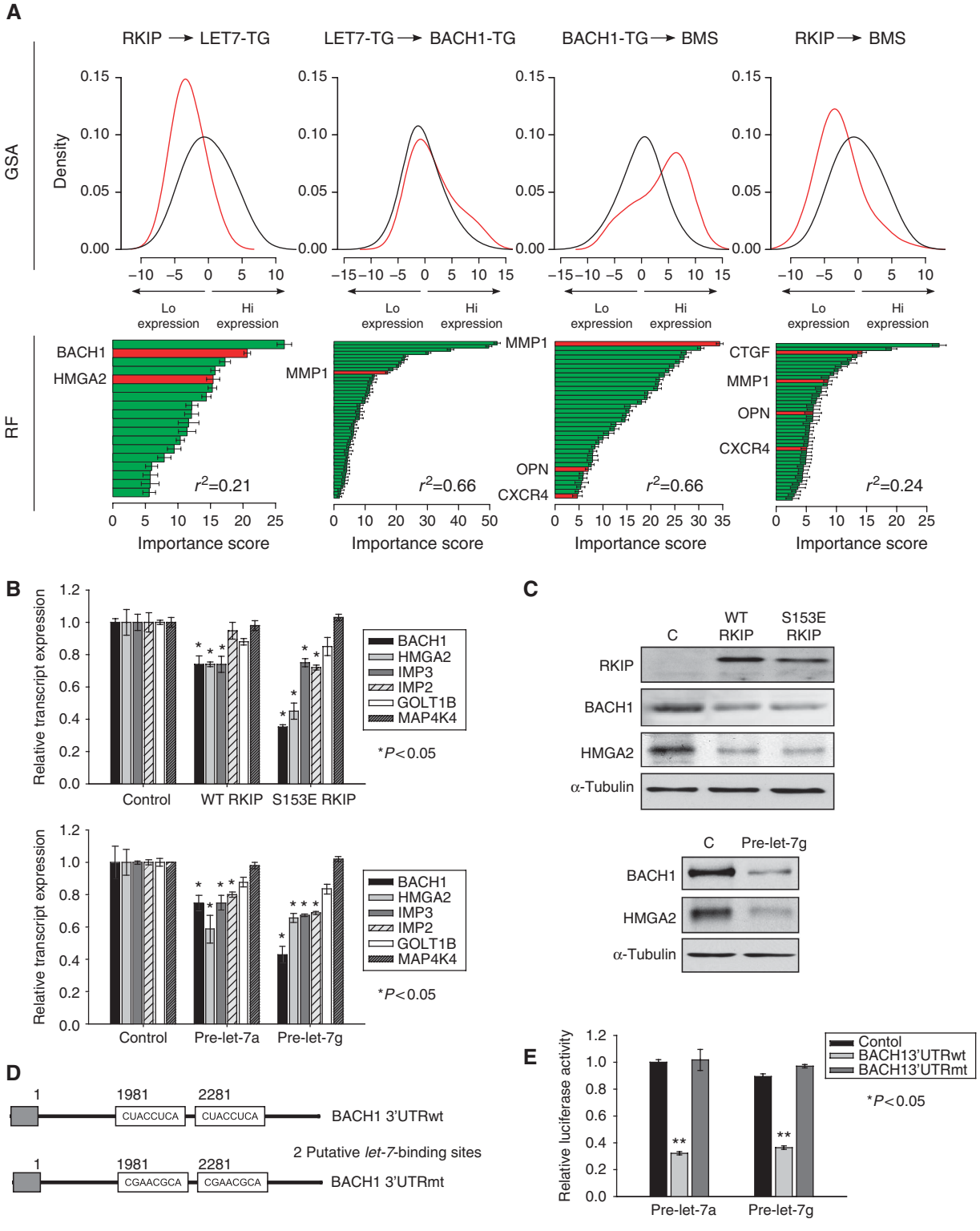
In total, the bioinformatics-driven pathway analysis using clinical expression data suggests that RKIP functions as a bone metastasis suppressor in part by inducing the expression of let-7, which in turn inhibits expression of proteins including HMGA2 and the predicted target BACH1. We identified a set of genes that are potentially transcriptionally activated by BACH1, and these BACH1 target genes correlate

©2011 European Molecular Biology Organization

with BMS genes. Finally, our analyses suggest that BACH1 may be a direct transcriptional regulator of the BMS gene MMP1. A heatmap shows the expression of these genes across primary breast cancers (Supplementary Figure S1B).

RKIP regulates BACH1, a novel let-7 target

In order to experimentally confirm that BACH1 is a novel RKIP-regulated let-7 target and to investigate other key features of the predicted RKIP metastasis signalling pathway, we analysed 1833 cells that either ectopically expressed wild-type RKIP (wt) or S153E RKIP, a non-phosphomimetic mutant that more potently inhibits Raf-1 (Corbit *et al*, 2003). Increased expression of wt RKIP and S153E RKIP down-regulated BACH1, HMGA2, IMP2, and IMP3 transcripts, as did expression of let-7a and let-7g precursors (pre-let-7) (Figure 1B). BACH1 and HMGA2 were also extensively down-regulated at the protein level (Figure 1C). To confirm that BACH1 mRNA directly binds let-7, the BACH1 3′UTR region



©2011 European Molecular Biology Organization

containing two let-7-binding sites was cloned into a luciferase reporter construct (Figure 1D). Expression of precursor let-7 (a,g) or RKIP (wt, S153E) inhibited reporter activity for the wt BACH1 3′UTR (BACH1 3′UTRwt) but not a mutated BACH1 3′UTR (BACH1 3′UTRmut) (Figure 1E; Supplementary Figure S3A). Similarly, precursor let-7g downregulated BACH1 protein, and a let-7 inhibitor (anti-let-7) upregulated BACH1 protein (Supplementary Figure S3B and C). In total, these results establish that BACH1 is a direct let-7 target, and RKIP regulates BACH1 via let-7 binding.

RKIP and let-7 regulate key BMS genes through HMGA2 and BACH1

Key BMS genes are predicted to be downstream of the RKIP/let-7 pathway. Of the four BMS mediators negatively associated with RKIP based on clinical expression data (Supplementary Figure S1A), at least three (MMP1, CXCR4, and OPN) had decreased transcript and protein expression in wt and S153E RKIP-expressing cells (Figure 2A and B). Similarly, these three genes were downregulated by transfection of let-7a and particularly let-7g (Figure 2A and B).

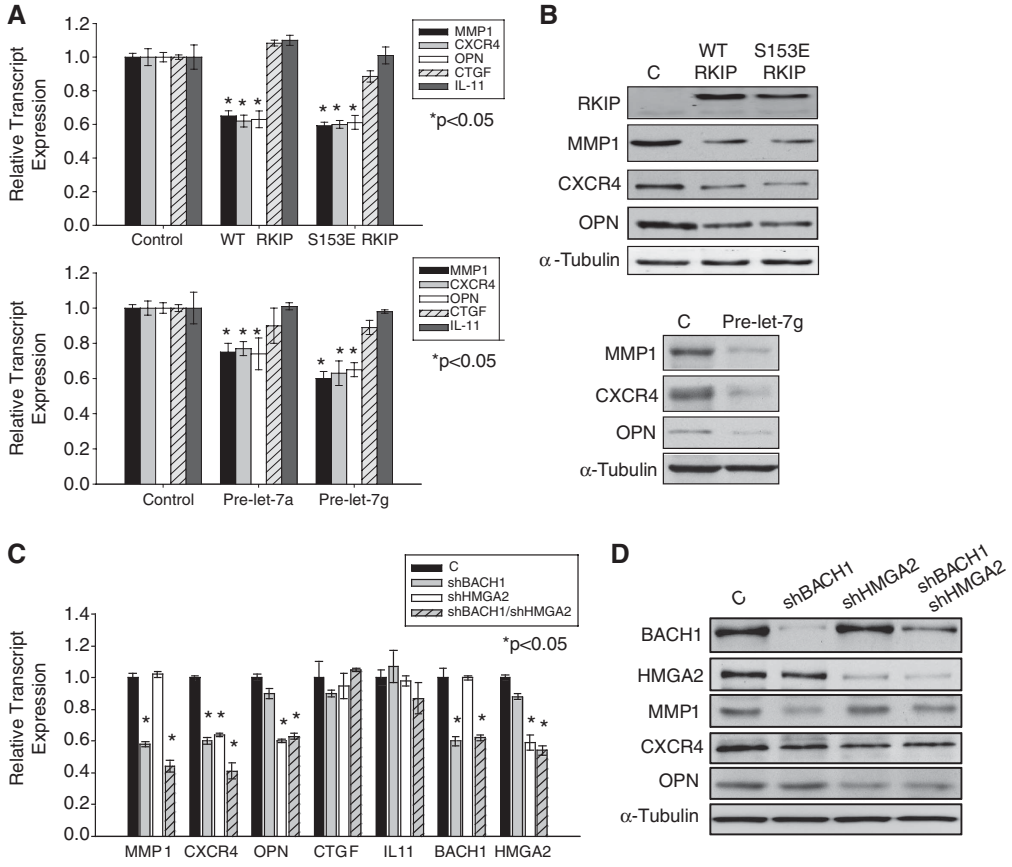
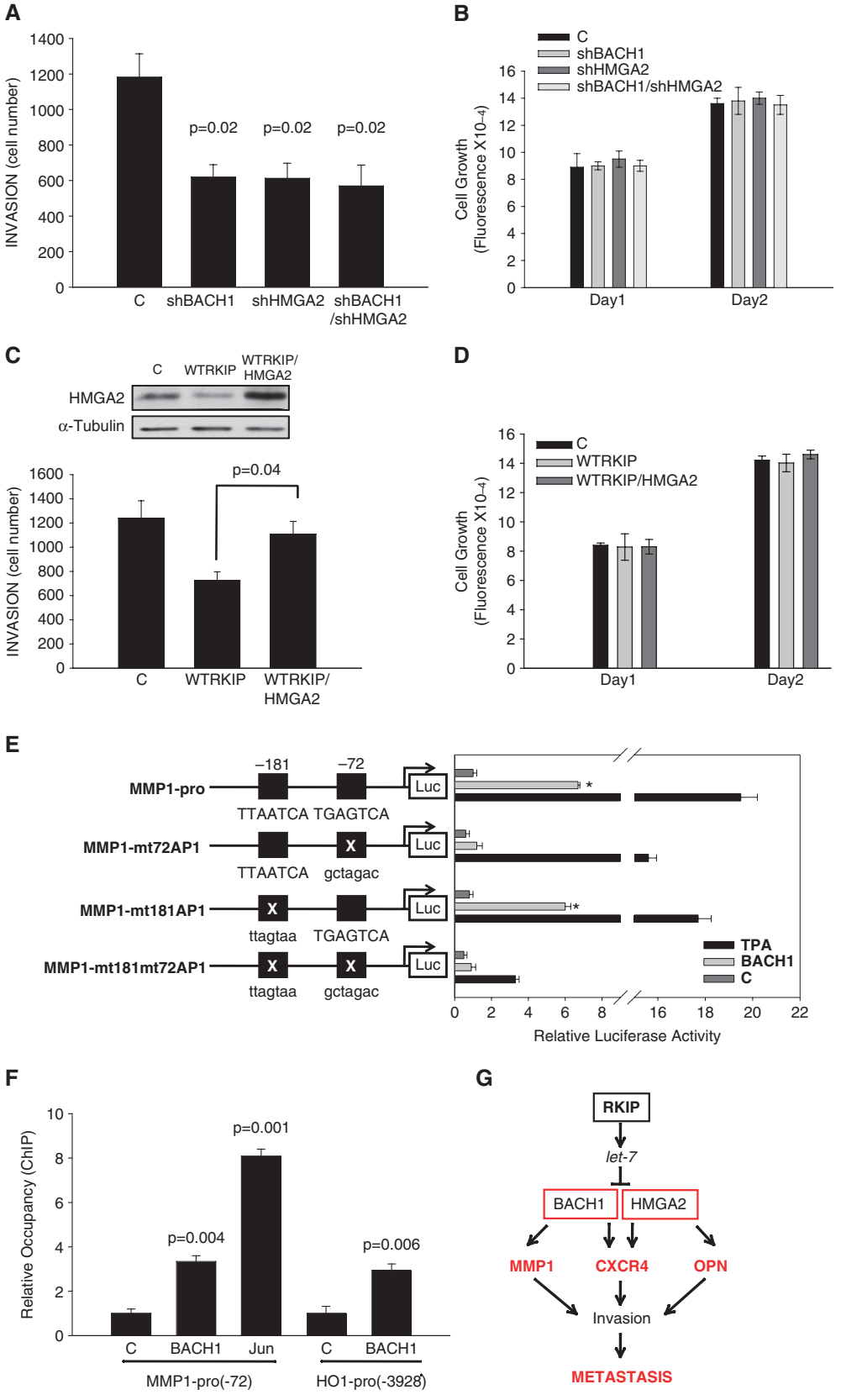


Figure 2 RKIP, let-7, BACH1, and HMGA2 regulate BMS genes. (**A**) RKIP or let-7 downregulate transcripts for bone metastasis gene signatures (BMS). 1833 cells expressing control vector, wt RKIP, S153E RKIP, let-7a, or let-7g were analysed by qRT–PCR (mean ± s.d., $n=3$) for BMS genes MMP1, CXCR4, OPN, IL-11, and CTGF. (**B**) RKIP and let-7 inhibit MMP1, CXCR4, and OPN protein expression. 1833 cells expressing vector, wt or S153E RKIP were immunoblotted for MMP1, CXCR4, OPN, RKIP, or α-tubulin. (**C**) Knockdown of BACH1 or/and HMGA2 in 1833 cells decreases MMP1, CXCR4, and OPN expression. 1833 cells expressing a control vector, shRNA for BACH1 or/and shRNA for HMGA2 were assayed for BMS genes (MMP1, CXCR4, OPN, CTGF, or IL-11) by qRT–PCR (mean ± s.d., $n=3$) or (**D**) were immunoblotted for the indicated proteins.

Figure 1 Identification of an RKIP/let-7 metastasis pathway regulating BACH1, a novel let-7 target. (**A**) RKIP pathway relationships were trained on the BrCa443 data set (Supplementary Figure S1A) and validated using the BrCa871 data set. The top panels show GSA results for the indicated steps in the RKIP pathway. Meta-genes comprised of let-7 or BACH1 target genes were used as readouts for let-7 (LET7-TG) or BACH1 (BACH1-TG), respectively. The red curve indicates the distribution of gene scores for each gene set shown on the right side of the pathway arrow in response to RKIP (first and fourth panels), LET7-TG (second panel), or BACH1-TG (third panel). The black curve is a null distribution. The *P*-values for all results are $P<0.001$. For BACH1-TG, MMP1 was omitted from the GSA to prevent bias. In the bottom panel, a multivariable random forest (RF) model was used to determine how well individual genes in each gene set shown on the right side of the pathway arrow account for the variation in upstream pathway genes shown on the left side of the pathway arrow. Genes are ranked by an importance score, which measures the contribution of each gene in the gene set in accounting for the observed variation (higher scores are better). Shown are barplots of the importance scores along with Monte Carlo standard deviations (for presentation purposes, genes with importance scores below the bottom 10% are not shown). Overall model fit is measured by the indicated pseudo-R-squared (see Supplementary data). (**B**) RKIP or let-7 downregulate transcripts for predicted let-7 targets. 1833 cells transduced with control vector, wt RKIP, or S153E RKIP (upper panel) or transfected with pre-miR let-7a or pre-miR let-7g (lower panel) were analysed by qRT–PCR for predicted let-7 targets BACH1, HMGA2, IMP2, IMP3, MAP4K4, and GOLT1B (mean ± s.d., $n=3$, $^*P<0.05$). (**C**) RKIP or let-7g inhibits BACH1 and HMGA2 expression in 1833 cells. 1833 cells expressing vector, wt or S153E RKIP (upper panel) or let-7g (lower panel) were immunoblotted for BACH1, HMGA2, or α-tubulin antibodies. (**D**) Schematic representation of BACH1 3′UTR with two putative let-7-binding sites. (**E**) BACH1 is a direct target of let-7. Pre-miR let-7a and g inhibit wt BACH1 3′UTR (BACH13′UTRwt) reporter activity but not mutant BACH1 3′UTR (BACH13′UTRmt). Luciferase activity was measured from 1833 cells transfected with pre-miR let-7a and g and normalized by Renilla luciferase activity 48 h after transfection (mean ± s.d., $n=3$, $^{**}P<0.005$).

©2011 European Molecular Biology Organization

The effect of the transcription regulators BACH1 and HMGA2 on BMS genes was then determined by shRNA depletion. Examination of mRNA and protein revealed that BACH1 depletion decreased MMP1 and CXCR4, and inhibition of HMGA2 lowered CXCR4 and OPN expression (Figure 2C and D). Knockdown of either gene inhibited invasion without affecting cell proliferation (Figure 3A and B). Similarly, overexpression of HMGA2 in RKIP-expressing cells rescued invasion by RKIP



©2011 European Molecular Biology Organization

without affecting cell proliferation (Figure 3C and D). Although three independent shRNA constructs for BACH1 showed similar inhibition of invasion, we were unable to isolate 1833 cells that overexpress BACH1

BACH1 directly regulates MMP1

Pathway analysis based on clinical expression data reveals that expression of MMP1 is highly predictive of the expression of BACH1 target genes; therefore, we focused on the relationship between the two genes. Analysis of the promoter region of MMP1 reveals at least two potential BACH1 sites that are similar to those of AP1. These sites were mutated singly or together to identify the site of BACH1 binding in 1833 cells (Figure 3E). A luciferase reporter linked to either the wt or mutated MMP1 promoter region confirmed that BACH1 induces MMP1 expression by binding to a specific AP1-like site in the promoter; by contrast, TPA activated both AP1-like sites (Figure 3E). Chromatin immunoprecipitation (ChIP) analysis of BACH1 binding similarly revealed an increase in MMP1 promoter association (Figure 3F). These results confirm a novel signalling cascade from RKIP to MMP1 via let-7 and its target BACH1, highlighting the role of let-7 targets in suppressing key metastasis genes. A scheme summarizing the RKIP metastasis pathway is depicted in Figure 3G.

A metastasis signalling pathway for RKIP regulates experimental metastasis

When introduced into the left cardiac ventricle, 1833 cells give rise to bone metastasis in immunocompromised mice. To determine if genes we identified along the RKIP metastasis signalling pathway affect experimental metastasis, a series of shRNA gene targeting and rescue experiments were performed. Depletion of BACH1, HMGA2, or both resulted in decreased metastasis of 1833 cells to bone, similar to RKIP expression (Figure 4A, left panel). In each case, overexpression of MMP1, OPN, and/or CXCR4 in combination could rescue this effect on metastasis and on invasion, with no effect on proliferation (Supplementary Figure S4A–D). Similarly, overexpression of HMGA2 was able to partially restore bone metastasis in RKIP-expressing cells (Supplementary Figure S4E). As noted above, we were unable to isolate BACH1-overexpressing cells. BACH1 depletion, like RKIP expression, had no effect on cell growth in culture or on primary tumour growth in mice but did inhibit intravasation (Figure 4B–D). HMGA2 was previously shown to promote primary tumour growth, which could account in part for the observed effects on metastasis (Mayr *et al*, 2007). In total, these results indicate that BACH1 selectively regulates invasion, intravasation, and metastasis.

More extensive analysis revealed that combinations of BMS genes are generally more effective than single genes at reversing RKIP inhibition of invasion (Supplementary Figure S5A and C) and bone metastasis (Figure 4A, right panel) with no effect on proliferation (Supplementary Figure S5B). However, even high overexpression of the three BMS genes was insufficient to completely overcome RKIP inhibition of metastasis (Figure 4A), and combinations of two genes poorly rescued HMGA2 or BACH1 loss (Figure 4A). Similarly, in 1833 cells expressing inducible let-7g, the three BMS genes rescued let-7g inhibition of invasion with no change in cell growth (Figure 5A–C) and partially rescued let-7g suppression of bone metastasis (Figure 5D). Since prolonged induction of let-7 expression also inhibits cell proliferation and presumably growth at distant sites (Dangi-Garimella *et al*, 2009), incomplete rescue by BMS genes is not surprising.

To ensure that the RKIP pathway is not limited to a single cell type, we tested the RKIP–BACH1–BMS axis using other invasive breast cancer cell lines. Similar to 1833 cells, ectopic RKIP expression in MDA-MB-436 cells inhibited invasion and bone metastasis but not proliferation (Supplementary Figure S6A–C). RKIP induced both let-7a and let-7g, and let-7g-regulated invasion (Supplementary Figure S6D–F). Finally, RKIP decreased expression of the let-7 target gene BACH1, and the BACH1-regulated BMS genes, MMP1 and CXCR4 (Supplementary Figure S6G). To test the effect of RKIP loss, we used MDA-MB-435 cells. Although the origin of this cell line has been disputed (breast or melanoma), it has many properties similar to those of other aggressive breast cancer cell lines (Chambers, 2009). MDA-MB-435 cells have ~20-fold more RKIP than 1833 cells and are weakly metastatic. RKIP depletion causes a decrease in let-7g and an increase in BACH1 expression (Supplementary Figure S7A–C), consistent with the reported increase in invasion and metastasis (Li *et al*, 2009).

Taken together, these results demonstrate that the RKIP signalling pathway regulates experimental metastasis in multiple breast cancer cell lines. Furthermore, downstream pathway genes in combination are more effective at promoting metastasis than individual genes, downstream mediators at least partially counteract the inhibitory effect of RKIP, and both upstream and downstream RKIP pathway genes contribute to metastasis.

The RKIP pathway metastasis signature

To further explore the significance of the RKIP metastasis pathway, we define a 'signalling pathway signature' termed the RKIP Pathway Metastasis Signature, or RPMS. Unlike

Figure 3 HMGA2 and BACH1, a direct regulator of MMP1, are required for breast cancer invasion. (**A**) Knockdown of BACH1 or/and HMGA2 inhibit invasion of 1833 cells. Cells stably expressing shBACH1 and/or shHMGA2 were assayed for invasion (mean ± s.d., $n=3$). (**B**) Knockdown of BACH1 or/and HMGA2 does not affect cell proliferation of 1833 cells. Cells from (**A**) were assayed for cell proliferation for the indicated times (mean ± s.d., $n=3$). (**C**) HMGA2 overcomes RKIP inhibitory effect on invasion in 1833 cells. Cells stably expressing RKIP, HMGA2, or both (inset) were assayed for invasion (mean ± s.d., $n=3$). (**D**) HMGA2 does not affect cell proliferation of RKIP-expressing 1833 cells. Cells from (**C**) were assayed for cell proliferation (mean ± s.d., $n=3$). (**E**) Schematic representation of MMP1 promoter with the putative BACH1-binding sites. BACH1 induces transcription by the MMP1 promoter. TPA was used as a positive control (50 ng/ml; 12 h). The MMP1 promoter region was fused to a luciferase reporter, 1833 cells were transfected with BACH1, and cells were assayed for luciferase activity as described in Materials and methods ($^*P<0.05$). (**F**) ChIPs were carried out with anti-BACH1 antibody and anti-Jun antibody (a positive control for ChIP assay) using 1833 cells. The promoter for HO1 was used as a positive control for BACH1. ChIP was analysed by qRT–PCR with primers in the MMP1 and HO1 promoters. Results represent the mean ± s.d. for three samples. (**G**) Scheme showing mechanism for RKIP regulation of invasion and metastasis via BACH1 and HMGA2 and its target BMS pathway. BACH1 directly enhances MMP1 transcription, BACH1 and HMGA2 induce CXCR4 expression, and HMGA2 induces OPN expression.

©2011 European Molecular Biology Organization

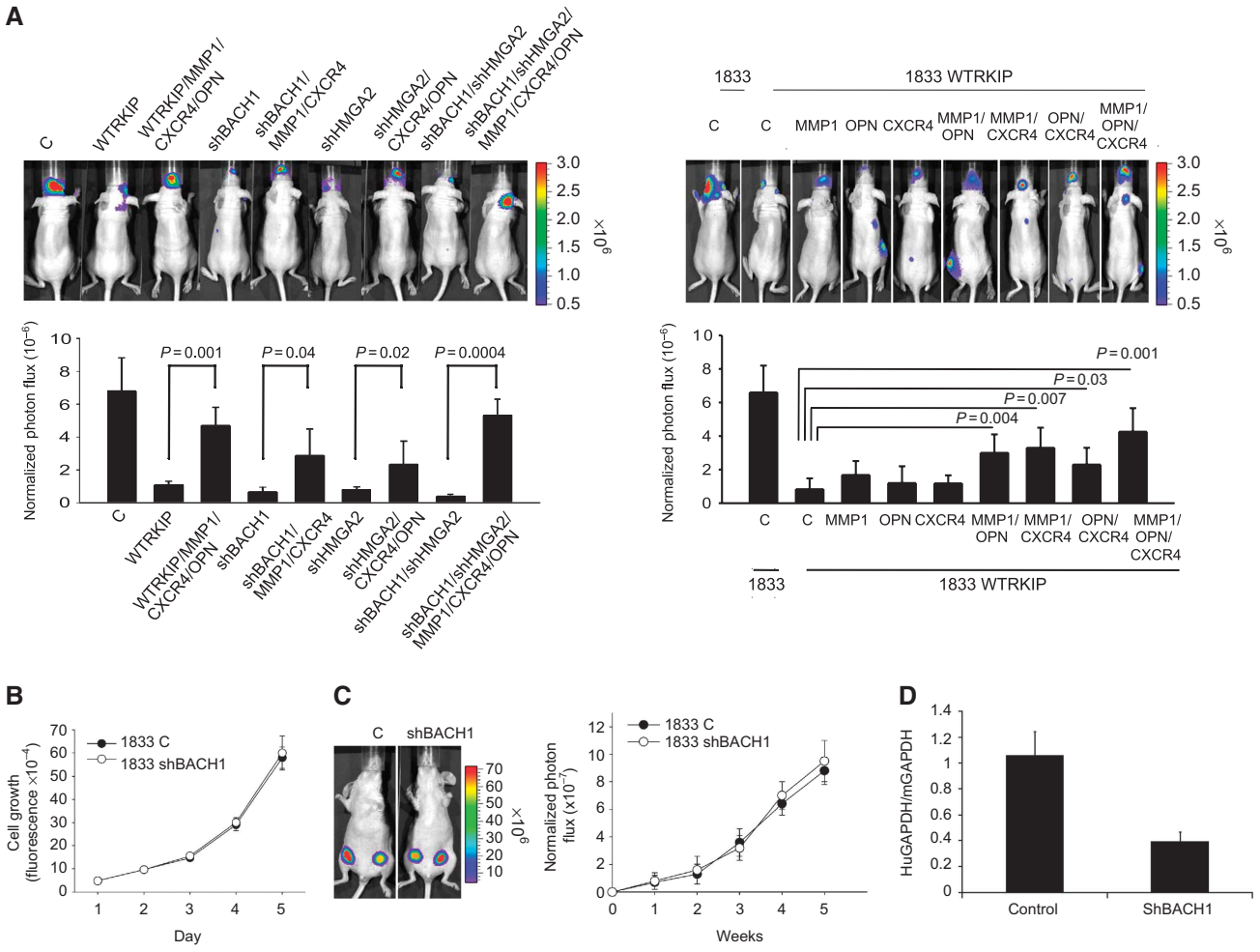


Figure 4 Experimental validation of the RKIP signalling pathway for metastasis. (**A**) BMS genes overcome the inhibitory effect of wt RKIP, shBACH1, shHMGA2, and shBACH1/shHMGA2 on bone metastasis and are more effective in combination. (Left panel) 1833 cells expressing either control vector, wt RKIP, shBACH1, shHMGA2, or shBACH1/shHMGA2 were stably transduced with control vector or target BMS genes, injected into the left cardiac ventricle of mice, and imaged for luciferase activity after 3 weeks (mean ± s.d., $n = 5$–6 for each group). (Right panel) 1833 cells expressing either control vector or wt RKIP were stably transduced with control vector or target BMS genes, injected into the left cardiac ventricle of mice, and imaged for luciferase activity after 3 weeks (mean ± s.d., $n = 5$–6 for each group). (**B**) BACH1 depletion does not alter tumour cell growth. Knockdown of BACH1 does not affect proliferation. (Left panel) Cells were assayed for cell proliferation as described in Materials and methods. Results represent the mean ± s.d. for three independent samples. (**C**) BACH1 depletion does not suppress tumour growth of 1833 cells. 1833 cells expressing luciferase and either control vector (six mice), or shBACH1 (six mice) were injected into the mammary fat pad of mice. After 1 week, tumours were imaged and measured for luciferase activity once a week. Representative images show shBACH1 does not affect tumour growth. Results represent the mean ± s.d. for the animals. (**D**) BACH1 depletion inhibits intravasation of bone-tropic tumour cells (1833). 1833 cells stably expressing control vector (five mice) or shBACH1 (five mice) were injected into the mammary fat pad of mice. After 3 weeks, cells isolated from the blood were analysed for GAPDH transcripts derived from human (tumour) or mouse (control). Results represent the mean ± s.e. for the animals ($P < 0.009$ for shBACH1 relative to control).

more typical gene expression signatures, this signalling pathway signature is based upon an experimentally validated signalling pathway that connects the genes. For five RPMS genes (RKIP, HMGA2, MMP1, CXCR4, and OPN), the signature solely measures their mRNA expression. For let-7 and BACH1, the signature measures meta-genes for let-7 targets (LET7-TG) and BACH1 targets (BACH1-TG). Thus, although the RPMS is based on a pathway of only seven genes, it measures the expression of 117 genes due to the large number of genes used as readouts for let-7 and BACH1 (see Figure 6A and Supplementary Table SI).

The association between RPMS expression and invasion was examined for a panel of 13 common breast cancer cell lines with published gene expression array data (Neve *et al*, 2006). The expression of each RPMS gene or meta-gene in each cell line was related to the median value for all cell lines to determine whether their expression was high or low (Figure 6B). Cells with an RPMS expression pattern that is generally inhibitory for metastasis (blue) are poorly invasive, while a pattern showing low RKIP and high downstream RPMS genes (orange) correlates with high invasiveness. These results experimentally support the ability of the RPMS to identify aggressive breast cancer cell lines.

Critical role of RKIP in the ability of the RPMS to predict clinical metastasis

To determine whether the RPMS is associated with metastatic risk, a univariable Kaplan–Meier analysis of the individual RPMS genes was performed. When considered individually,

 ©2011 European Molecular Biology Organization

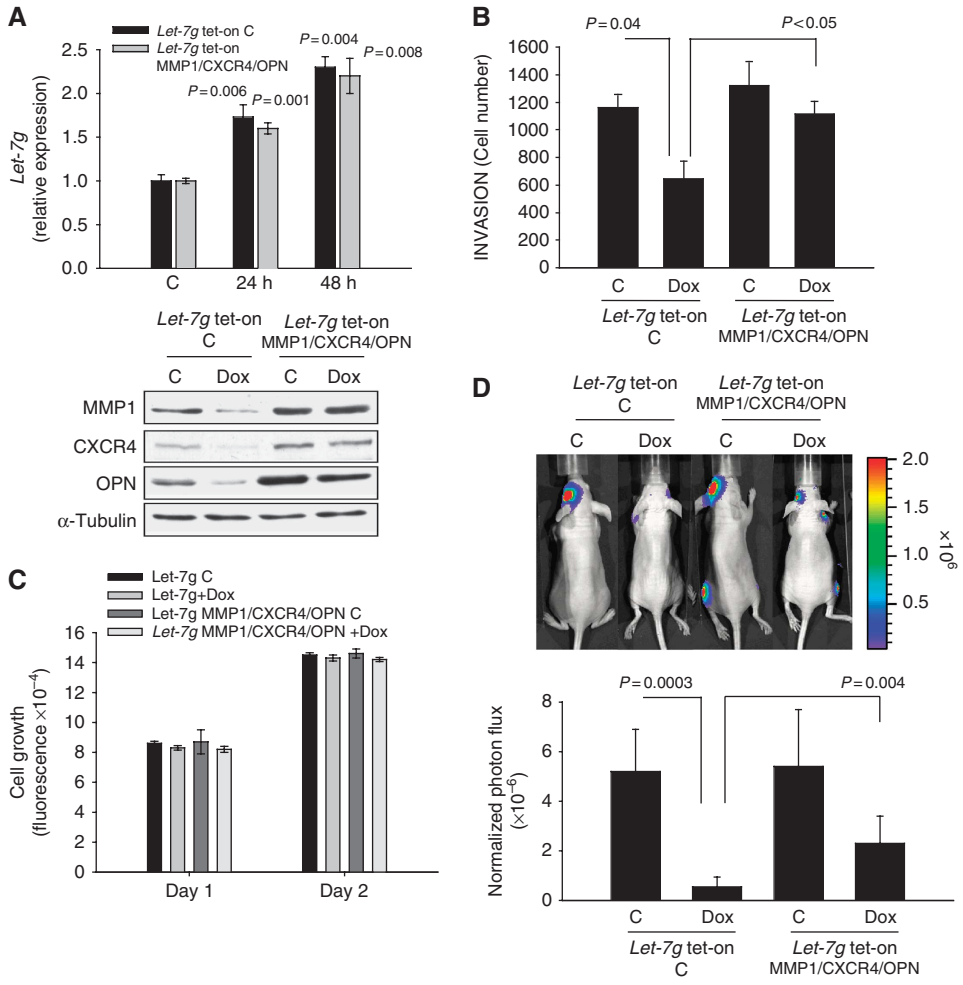


Figure 5 (**A**) BMS genes do not alter *let-7g* induction. (Upper panel) 1833 cells expressing tet-inducible *let-7g* and either control vector or BMS genes (MMP1, CXCR4, and OPN) were treated with 2 μg/ml doxycycline for 24 and 48 h and then assayed for *let-7g* expression by qRT–PCR. Results represent the mean ± s.d. for three samples. (Lower panel) Expression of BMS proteins in 1833 cells expressing tet-inducible *let-7g*. 1833 cells expressing tet-inducible *let-7g* and either control vector or BMS genes (MMP1, CXCR4, and OPN) were treated with 2 μg/ml doxycycline for 48 h. Cell lysates were immunoblotted with antibodies to MMP1, CXCR4, OPN, and α-tubulin. (**B**) BMS genes rescue the inhibitory effect of *let-7* on invasion. 1833 cells from (**A**) were assayed for invasion as described in Materials and methods. Results represent the mean ± s.d. for three independent samples. (**C**) *Let-7* expression for 2 days does not affect cell proliferation. 1833 cells from (**A**) were assayed for cell proliferation as in Materials and methods for the indicated times. Results represent the mean ± s.d. for three samples. (**D**) BMS genes partially restore metastasis to cells expressing *Let-7g*. 1833 cells expressing luciferase, tetracyline-inducible *let-7g* and either control vector or BMS genes (MMP1, CXCR4, and OPN) were grown in the presence of 2 μg/ml doxycycline for 24 h. Cells were injected into the left ventricle of mice, and 2 days later, mice were administered with drinking water containing 4% sucrose only or 2 mg/ml doxycycline and 4% sucrose. Mice were imaged for luciferase activity after 3 weeks. Results represent the mean ± s.d. for the animals.

RPMS genes show modest or no association with metastasis-free survival among the BrCa871 data set (Figure 6C). In contrast, patients with RPMS expression indicating high pathway activity (RKIP expression less than the median and expression of downstream RPMS genes greater than the median) have a markedly greater risk for metastasis compared with those patients with a less active pathway (Figure 6C, bottom right panel).

The importance of RKIP status is further underscored by examining all patients that coexpress downstream RPMS genes above the median. The successive addition of downstream RPMS genes leads to greater separation of the survival curves among patients with low RKIP (Figure 7A, RKIP low). By contrast, downstream RPMS genes do not significantly increase metastatic risk among patients with RKIP expression above the median (Figure 7A, RKIP high). The striking increase in metastatic risk for patients with highly expressed downstream RPMS genes and low but not high RKIP was further confirmed by independent breast cancer patient data sets (Ishwaran H, Boelens MC, Rosner MR, and Minn AJ, in preparation). These findings highlight the potent ability of RKIP to act as a metastasis suppressor.

RPMS genes cooperatively drive clinical metastasis

To test whether the genes in the RPMS predict metastasis risk in a cooperative rather than an additive way, we utilized random survival forests (RSFs). An ensemble partitioning tree method for survival data that considers the effects of all possible interactions between genes, RSF allows coopera-

©2011 European Molecular Biology Organization

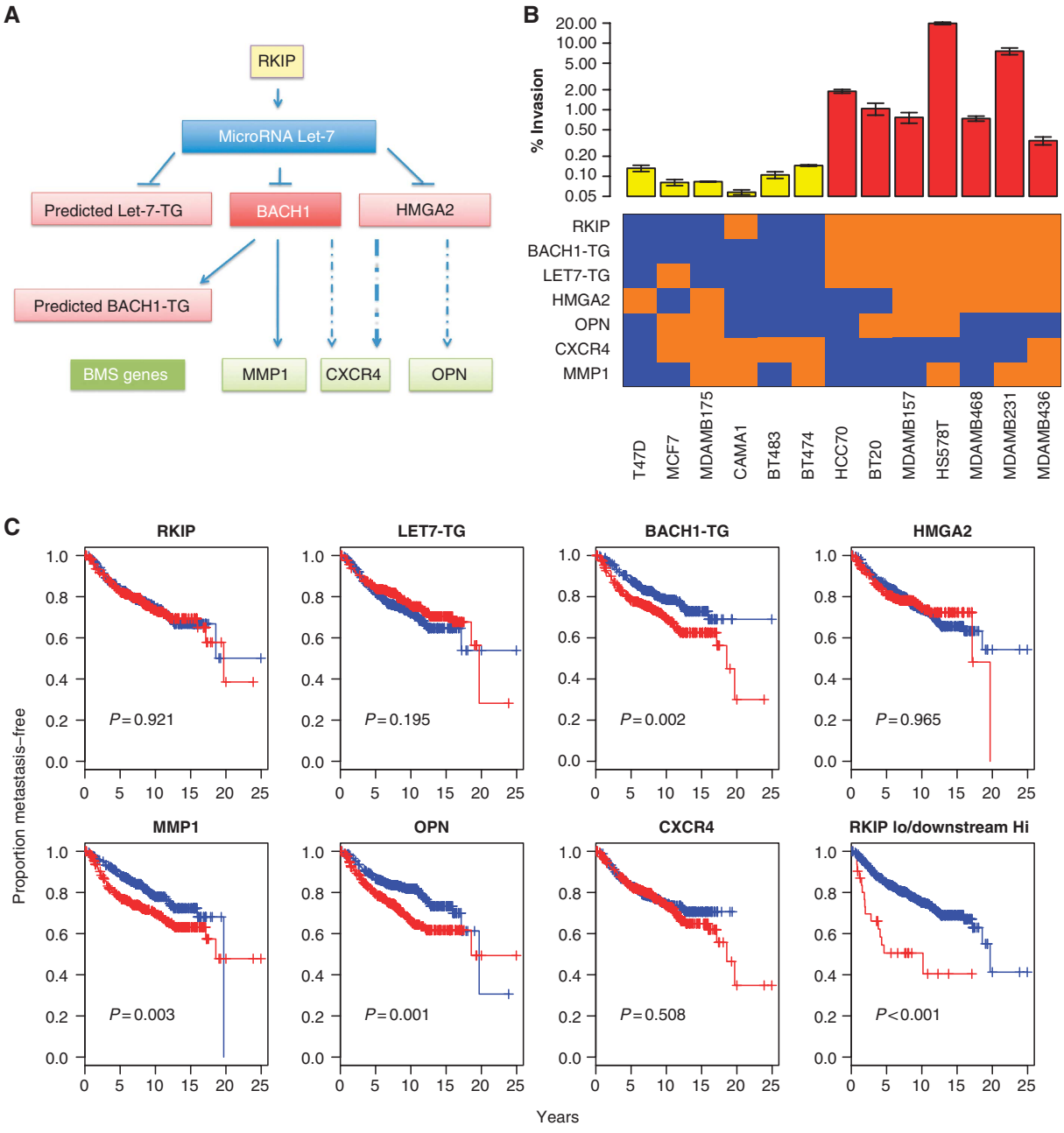


Figure 6 Association of individual RPMS genes with clinical metastasis. (**A**) Schematic showing the genes comprising the RPMS. Only genes shown in black lettering are elements of the RPMS. Both the LET7-TGs and the BACH1-TGs are based on bioinformatic predictions. (**B**) Correlation between invasive breast cancer cell lines and the RPMS. Gene expression microarray data from 13 breast cancer cell lines were analysed for expression of genes in the RPMS. For five RPMS genes (RKIP, HMGA2, MMP1, CXCR4, and OPN), the mRNA expression of these genes is measured relative to the median. For let-7 and BACH1, the meta-gene for let-7 targets (LET7-TG) and the meta-gene for BACH1 targets (BACH1-TG) are measured relative to the median. Expression of RKIP greater than the median value or expression of other RPMS genes less than the median value are expected to inhibit invasion/metastasis and are coloured in blue. Low expression of RKIP or high expression of other RPMS genes are expected to promote invasion/metastasis and are coloured in orange. Hierarchical clustering was used to group cells by RKIP pathway genes. The percentage of invasion (mean ± s.d.) is indicated above the heatmap. (**C**) Metastasis-free survival for patients from the BrCa871 data set is shown stratified by median cut point for the expression values for each of the indicated genes. For each gene, the corresponding probe intensity values from microarray data were used, except for let-7 and BACH1 where the meta-gene was used. Red indicates expression greater than the median and blue is less than the median. For the plot 'RKIP low/downstream high', red indicates patients that have RKIP expression less than median and all other downstream genes have expression greater than median, while blue indicates patients that do not meet this condition. Survival curves are determined by the Kaplan–Meier method and the indicated *P*-values are calculated by the log-rank test.

tivity to be assessed by comparing the sum of the importance score of individual or groups of genes with the joint importance scores of all genes (Minn *et al*, 2007; Ishwaran *et al*, 2008, 2010; Weichselbaum *et al*, 2008). The importance score is greatest when all genes are considered together (coop) compared with the sum of the importance scores for the upstream and downstream genes considered separately (add) (Figure 7B). Although a small cooperative effect is

 ©2011 European Molecular Biology Organization

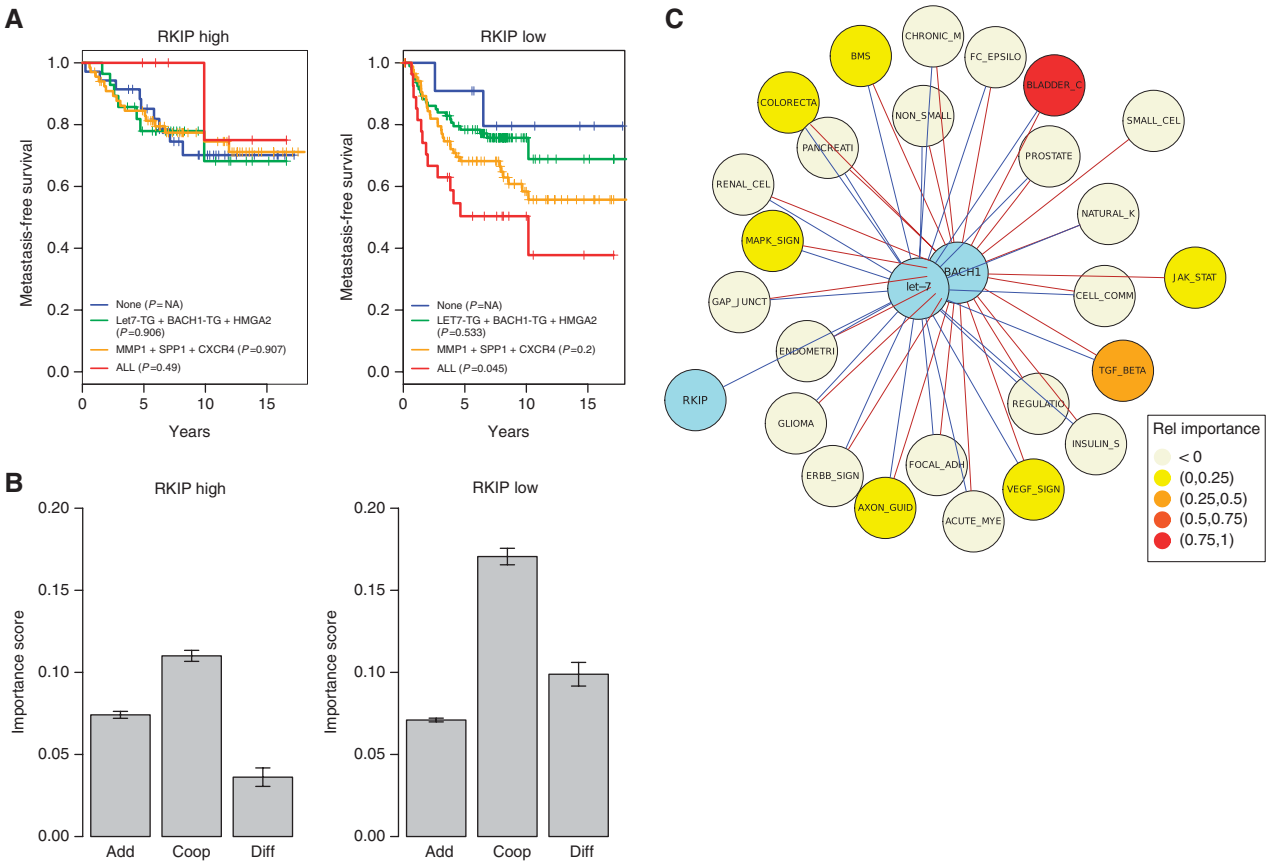


Figure 7 The RPMS forms a network of genes that cooperatively predict and contribute to metastasis. (**A**) Metastasis-free survival for patients stratified by RPMS status is shown. RKIP low (RKIP transcript expression below the median) and RKIP high (RKIP transcript expression above the median) patients in the BrCa871 cohort of patients are analysed separately. Statistical significance for the indicated pathway conditions (legend on bottom) was determined by comparison to patients with low expression of all RKIP pathway genes downstream of RKIP (none) and reported as a *P*-value based on the log-rank test. (**B**) The BrCa871 data set was used to determine cooperativity between the RPMS genes in predicting clinical metastatic risk for tumours that are RKIP high or RKIP low. An importance score, which measures the degree to which genes contribute to overall prediction accuracy (higher scores are better) was calculated for RPMS genes separately or together (Coop). The cooperative importance score was then compared with the sum of the separate importance scores for the downstream genes in the pathway (Add) and the difference (Diff) was used as a measure of cooperation. Error bars are standard deviations from 500 Monte Carlo runs. (**C**) RPMS genes are part of a network that contributes to clinical metastasis. Cancer-related pathways were tested for association with the RKIP pathway by GSA and by overrepresentation among let-7 and BACH1 target genes using a hypergeometric test. Each node represents a pathway that passes both test criteria. Pathways with genes enriched in let-7 target genes are shown using a blue edge and pathways with genes enriched in BACH1 targets are shown with a brown edge. A meta-gene score for each pathway shown was used along with the indicated RPMS genes (light blue) in a predictive model for metastasis using the BrCa871 data set (see Materials and methods). The contribution of each pathway to metastatic risk was measured by a relative importance score. This relative importance score is displayed in quartiles as colour-coded pathway nodes. When considering the entire network, pathways with an importance score less than zero (beige) contribute little to metastatic risk prediction. The prediction error for the entire network is 37.9 ± 0.66%.

observed in RKIP high tumours, the effect is markedly higher in RKIP low tumours (diff).

The RPMS defines a metastatic gene network

Because the RPMS predicts metastasis and the RKIP-regulated genes are likely members of a more extensive signalling network, we determined the potential relationship of the RPMS to known cancer-related pathways. For this analysis, KEGG cancer-related gene sets along with the BMS were tested for a significant association with the RKIP pathway and for connections to let-7 and/or BACH1 target genes. Cancer gene sets determined to be significant are shown as nodes (Figure 6A). Notably, breast cancer proliferation genes and common breast cancer prognostic signatures were also analysed but were not significantly associated with the RKIP/let-7/BACH1 pathway (Ishwaran H, Boelens MC, Rosner MR, and Minn AJ, in preparation). Edges between nodes indicate that at least one or more genes are shared between the pathway and the BACH1 and/or let-7 target gene sets. Of these genes, N-Ras, SHC3, TGFβR1, and AKT3 stood out as most connected. The relative contribution of each pathway node to clinical metastatic risk was also measured by a relative importance score denoted by differences in colour. The top metastatic nodes include the bladder cancer pathway, TGFβ pathway, MAPK signalling, VEGF signalling, JAK-STAT signalling, and the BMS. Pathways with a relative importance score of less than zero have the lowest contribution to predicting metastatic risk when the entire network is considered. The results of these analyses demonstrate that the RKIP/let-7/BACH1 axis interacts with multiple pathways through key genes that have previously been implicated in metastasis.

©2011 European Molecular Biology Organization

Discussion

By integrating experimental systems and clinical expression data with molecular and statistical modelling, we leveraged the ability of each approach to predict and validate a unique RKIP signalling pathway that regulates metastasis. We demonstrate that RKIP inhibits let-7 target genes BACH1 and HMGA2 that in turn promote breast cancer metastasis genes MMP1, OPN, and CXCR4. We utilized this RKIP signalling pathway as the basis for generating a differentially weighted metastasis gene signature termed the RPMS (Figure 7C). This approach for constructing pathway-specific gene signatures downstream of a key tumour suppressor is a general one that should have applicability to other genes and tumour types.

The RPMS is expressed in part or in full by a subpopulation of primary human breast cancers expressing RKIP at levels lower than the median. Individually, each RPMS gene is negligibly or weakly associated with metastasis; however, when considered together the ability to identify high-risk patients dramatically improves. The RPMS predicts invasive potential in breast cancer cell lines and predicts metastasis across breast cancer patient cohorts. Taken together, these results suggest that the RPMS identifies a cancer patient subpopulation with an intracellular environment that enables induction of metastasis by the mechanistically linked RKIP signalling cascade.

Further clinical validation and translation of the RPMS is addressed in a related study (Ishwaran H, Boelens MC, Rosner MR, and Minn AJ, in preparation). For example, basal triple-negative breast cancers (TNBCs), which comprise 15–20% of invasive breast cancer tumours overall and 60% among African American populations, have the shortest survival and highest rates of relapse after treatment (Schneider *et al*, 2008). In contrast to proliferation-based signatures, the RPMS also predicts metastasis for basal/TNBCs (Ishwaran H, Boelens MC, Rosner MR, and Minn AJ, in preparation). Identification of the RPMS as a prognostic metastasis pathway signature for difficult to predict subtypes such as TNBC reveals how it complements the weaknesses of existing prognostic gene signatures.

In contrast to most gene signatures, an overall mechanistic connection between individual RPMS component genes was experimentally and clinically validated. We therefore refer to the RPMS as a 'signalling pathway signature' to distinguish it from signatures that are not constructed based on pathway topology. Specifically, we show that targets of let-7 required for experimental metastasis include epigenetic regulators such as HMGA2, a chromatin remodelling factor, and BACH1. BACH1 is a basic region-leucine zipper (bZip) transcription factor that binds to MAREs (Maf recognition elements), acts as a repressor of genes that regulate the oxidative stress response (Kyo *et al*, 2004), and suppresses p53-mediated cellular senescence (Dohi *et al*, 2008). Our work reveals that BACH1 also directly regulates MMP1, one of four key pro-angiogenic genes shown to potentiate breast metastasis and mediate invasion (Gupta *et al*, 2007) and HMGA2 regulates OPN, an instigator of distant tumour growth (McAllister *et al*, 2008). Interestingly, MMP1 and OPN contribute to metastasis most strongly when RKIP is depleted, and network analysis shows that the BMS, which is regulated in part by RKIP, has one of the strongest associations with clinical metastatic risk. Furthermore, the RKIP/let-7 to BACH1/MMP1 axis is particularly predictive for breast cancer metastasis.

The RPMS likely identifies a subpopulation of high-risk human breast tumours by revealing a cellular signalling environment that is favourable to metastatic progression. Similar to other tumour suppressors, RKIP loss alone is not sufficient to promote invasion and metastasis unless RKIP depletion occurs in certain cellular signalling contexts (Figure 7A–C). High invasion does not require high expression of all three BMS genes in breast cancer cell lines, suggesting that other metastasis genes might also play a role in the RPMS pathway. Finally, it is likely that the let-7 pathway is only one mechanism by which RKIP regulates RPMS genes, and there are other pathways controlled by the RKIP axis. Detailed investigation of genes that comprise the RKIP network should yield further insight into the mechanism by which RKIP suppresses metastatic progression. Our findings imply that targeting BACH1, MMP1, OPN, and possibly other BMS genes would be highly effective for RPMS-positive tumours, and illustrate the potential therapeutic utility of signalling pathway signatures.

Materials and methods

Cell culture and generation of cell lines

Cell lines (MDA-MB-231 1833, MDA-MB-436, MDA-MB-435) expressing RKIP or inducible let-7g, or with stable knockdown of RPMS genes, were generated and maintained as previously described (Dangi-Garimella *et al*, 2009). MMP1, CXCR4, or/and OPN expressing cell lines were generated by transducing target cells with MMP1, CXCR4, or OPN subcloned into a pLenti4 lentiviral vector (Invitrogen Corp., Carlsbad, CA). The HMGA2 expressing cell line was generated by transfection with pH3HX-HMGA2. Cells were selected in 400 μg/ml G418. Methods for transient transfection of pre-miR let-7g and anti-miR let-7g have been described (Dangi-Garimella *et al*, 2009). BACH1 3′UTR luciferase reporter constructs were cloned into psiCheck2 plasmid (Promega, Madison, WI). MMP1-pro is a luciferase reporter construct containing the human MMP1 276 bp upstream of the transcriptional start site in pGL3basic plasmid backbone (Promega).

Immunoblotting

Cells were lysed in RIPA buffer (50 mM Tris–Cl, pH 8.0, 150 mM NaCl, 0.1% sodium dodecyl sulphate, 1% Nonidet P-40) supplemented with a cocktail of protease inhibitors (Calbiochem, San Diego, CA). The protein concentration of each lysate was determined using the BCA assay (Bio-Rad, Hercules, CA) and resolved and probed as described (Dangi-Garimella *et al*, 2009). Antibodies specific for BACH1 (sc-14700), MMP1 (sc-12348), OPN (sc-20788), and α-tubulin (sc-8035) were purchased from Santa Cruz Biotechnology (Santa Cruz, CA). Antibody for CXCR4 (551852) purchased from BD Biosciences (Bedford, MA) and the antibody for HMGA2 (59170AP) was purchased from BioCheck Inc. (Foster City, CA). Anti-RKIP antisera used was generated by immunizing rabbits with purified GST-RKIP (α-RKIP).

In vitro cell assays

Invasion and cell proliferation assays were performed as previously described (Dangi-Garimella *et al*, 2009). Luciferase reporter assays were performed using a Dual-Luciferase® Reporter Assay (Promega) as described by the manufacturer. Transfections were performed in 24-well plates. At 48 h post-transfection, medium was removed, and the cells were washed twice with cold PBS. A measure of 100 μl of Passive Lysis Buffer (Promega) was added to each well. A total of 10 μl of lysate was analysed by a GloMax microplate luminometer (Promega). Renilla luciferase activity was used as an internal control.

Animal studies

All animal work was done in accordance with a protocol approved by the Institutional Animal Care and Use Committee. Bone

 ©2011 European Molecular Biology Organization

metastasis assays have been described (Dangi-Garimella *et al*, 2009). For the tetracycline induction of let-7g, 1833 let-7g tet-inducible cells were plated in the presence of 2 μg/ml doxycycline. Twenty-four hours later, cells (10^5) were injected into the left ventricle of mice. Two days later, mice were administered drinking water containing 4% sucrose only or 2 mg/ml doxycycline and 4% sucrose. Mice were imaged for luciferase activity after 3 weeks. Bone metastases were localized in either the skull or femur/tibia regions of the mice dependent upon cell type and age of animal. The intravasation assays were conducted as previously described (Dangi-Garimella *et al*, 2009).

RNA isolation and RT–PCR

Total RNA was isolated from cells with RNeasy Mini Kit according to the manufacturer's instructions (Qiagen, Valencia, CA). Reverse transcription and RT–PCR was performed as before (Dangi-Garimella *et al*, 2009).

Quantitative ChIP analysis

ChIP was carried out as described previously (Wu *et al*, 2005) with 2 μg of antibodies against either BACH1 (sc-14700, Santa Cruz Biotechnology), Jun (Santa Cruz Biotechnology), or with preimmune serum. The precipitated chromatin (2 μl) was analysed by quantitative real-time PCR with the indicated primer pairs. Forward and reverse primers for real-time PCR (5′–3′) ChIP analysis were as follows: MMP1(BACH1), CTTGTTTGAAGTTAATCGTGACAC and AGCCTCTTGCTGCTCCAATATC; HO-1(BACH1), CAGTGCCTCCTCA GCTTCTC and CTCGGTGGATTGCAACATTA, and rRNA as an internal control, ATTAGTCAGCGGAGGAAAAGAAAC and TCGCCGTTACTGA GGGAATC.

Statistical analysis of experimental results

Samples were analysed by the two-sample Student's *t*-test assuming unequal variances (two-tailed). Excel software was used for statistical analysis. *P*-values were calculated for samples from three independent experiments unless otherwise indicated.

Software

All analyses were performed using the R language and environment for statistical computing version series 2.8 through 2.9 (R core development team, 2009). The base packages from Bioconductor version 2.4 (Gentleman *et al*, 2004) were used and other R packages used are described. Unless otherwise noted, default parameters were used.

Microarray data processing and normalization

The microarray expression data used in this study are available through the GEO or other public repository. The BrCa443 data set is comprised of three cohorts: GSE5327, GSE2034, and GSE2603. The BrCa871 data set is comprised of five cohorts: GSE1456, GSE2990, GSE3494, GSE7390, and GSE11121. Duplicate samples (126 out of 997) were identified by sample identifier and removed. The BrCa295 data set is comprised of a single cohort (http://www.rii.com/publications/2002/vantveer.html). For the BrCa443 and BrCa295 data set, raw data were not available for all the samples at the time this work was initiated; therefore, processed data were used unless otherwise noted. For the BrCa871 data set, Affymetrix CEL files for all samples were processed at once using the RMA method as implemented in the 'affy' R package version 1.20.2 (Gautier *et al*, 2004). Only genes common to all platforms were kept, and if multiple probes existed for the same gene, the probe with the highest variance was used. Final normalization of all three data sets was performed by *z*-score transformation (Cheadle *et al*, 2003). The GEO accession numbers for the two microRNA expression data sets used are GSE19783 and GSE7842.

Construction of candidate let-7 target gene list

To obtain a high-confidence set of *let-7* target genes, we used the intersection of two leading target prediction programs: TargetScanS (Lewis *et al*, 2005) and PicTar (Krek *et al*, 2005). For TargetScanS, we only used the genes whose target sites contain both 7 mer seed match and the anchored adenosine nucleotide (t1A + m8); for PicTar, we only used the five-species predicted targets. In addition, we included a few target genes with experimental evidence, which were compiled from the TarBase database (Sethupathy *et al*, 2006).

GSA and gene set reduction

To test whether gene sets were enriched in response to genes of interest, we utilized GSA as implemented in the 'GSA' R package version 1.03 (Efron and Tibshirani, 2007, 2010). The 'maxmean' test statistic was used to test enrichment with *P*-values and false-discovery rates based on 500–1000 permutations. For restandardization, a method that combines randomization and permutation to correct permutation values of the test statistic and to take into account the overall distribution of individual test statistics, the entire data set was used rather than only the genes in the gene sets tested. This method of restandardization was chosen because we used GSA for hypothesis testing rather than for exploratory analysis; therefore, only a small number of gene sets were analysed. Gene sets analysed include the bone metastasis gene signature (Kang *et al*, 2003), the high-confidence predicted let-7 target genes described above, and predicted BACH1 target genes from TRANSFAC version 7.4. When RKIP was used as a response variable in GSA, we conservatively chose to average all probe sets for RKIP prior to *z*-score transformation if multiple probe sets for RKIP existed on the platform. Prior to GSA and the calculation of gene scores, genes in the gene set that overlap with the response variable or were used to calculate the response variable (in the case of meta-gene values) were removed to prevent bias.

The BrCa443 data set was used as an initial training set to test the expected relationships between the genes in the RKIP metastasis signalling pathway. Also, because not all genes in the gene sets are expected to be significant, the BrCa443 was also used as a training set to reduce the size of the gene sets. For gene set reduction, only the genes with gene scores having the same sign as the maxmean statistic were kept. This resulted in a reduction of the *let-7* target gene set down to 22 genes and a reduction of the BACH1 target gene set down to 90 genes (Supplementary Table SI). The BrCa871 data set was used to validate the GSA results using both the full gene sets and the reduced gene sets.

Random forest regression for pathway analysis

We assume that if a gene set is significantly enriched in response to a gene of interest, individual genes in the gene set may be mechanistically related to the response gene. To select individual genes for functional validation, genes in a gene set most strongly associated with the response gene were prioritized. This selection was based on individual gene scores returned by GSA, which represent a distribution of univariable test statistics. However, GSA is meant to test the significance of a gene set and not individual genes. Thus, to complement GSA results, we used random forest regression analysis (RF-R), a non-parametric ensemble partitioning tree method that can control for the effects of all genes in the gene set and can account for the effects of interactions between genes (Breiman, 2001; Pang *et al*, 2006). For each gene, an importance score is determined, which is a measure of the ability of variables to account for variance (higher scores are better). Therefore, RF-R was used to confirm that each gene in the RKIP pathway that we selected for biological testing was strongly associated with upstream pathway genes based on the clinical expression data. This approach was also combined with GSA as a means to validate RKIP pathway relationships using the BrCa871 data set. Performance of RF-R was measured using a pseudo-R-squared (one minus the mean square errors divided by the variance), and importance scores were measured by the mean decrease in test-mean squared error. In general, RF-R results were based on 500 trees and 100 Monte Carlo replications. RF-R was implemented using the 'random forest' R package version 4.5-30 (Liaw and Wiener, 2002).

The RKIP pathway metastasis signature

For RKIP and the three metastasis mediators MMP1, CXCR4, and OPN, the RPMS solely measures the mRNA expression of these genes. For the microRNA *let-7* and the transcription factor BACH1, the signature measures a meta-gene for *let-7* targets (LET7-TG) and the meta-gene for BACH1 targets (BACH1-TG) as an indicator of the functional activity of these regulatory genes. Ideally, we would also measure a meta-gene for HMGA2 targets; however, a list of predicted targets for this transcription factor is not available, hence, we use the mRNA expression value for the gene.

To calculate a meta-gene for *let-7* targets (LET7-TG), we used a weighted average of the 22 genes in the reduced *let-7* target gene set derived from the BrCa443 data set. Specifically, the expression values for *let-7* target genes with gene scores having the same sign

©2011 European Molecular Biology Organization

as the maxmean statistic were averaged and weighted by the individual gene scores (see Supplementary Table SI for a list of the gene scores). In this way, a meta-gene value for the predicted *let-7* target genes that are enriched in response to RKIP is derived for each sample. This meta-gene could then be used as a variable for further GSA or other analyses. A similar approach was used to calculate a meta-gene value for the predicted BACH1 target genes. Here, the meta-gene for BACH1-TG is a weighted average of the 90 genes in the reduced BACH1 target gene set enriched in response to LET7-TG. Of note, the genes and weights used for meta-gene calculation for all patients in all data sets were determined from the BrCa443 training set.

Univariable survival analysis stratified by individual RPMS genes

All patients with metastasis-free survival information from the BrCa871 data set (695 patients) were included in survival analyses. Univariable survival analysis was performed by the Kaplan–Meier method and log-rank test using the 'survival' R package version 2.35-4 (Therneau and Lumley, 2009). Patients were stratified using a median cutoff for the expression values of the gene or meta-gene of interest.

Determining cooperativity for RPMS genes

All patients with metastasis-free survival information from the BrCa871 data set were included in survival analyses (695 total patients, see Supplementary data). Similar to our previous work (Minn *et al*, 2007; Weichselbaum *et al*, 2008), multivariable survival analysis was performed using RSF, an ensemble partitioning tree method for survival data, using the 'randomSurvivalForest' R package version 3.5.1 (Ishwaran and Kogalur, 2008; Ishwaran *et al*, 2008). In general, 1000 trees were evaluated using the log-rank splitting rule. To improve computational speed, a random version of the log-rank splitting rule was optionally used by invoking random split points rather than deterministic splitting. All genes and meta-genes in the RPMS were used as continuous variables. Any missing data were imputed. Joint importance scores for RPMS genes as a function of RKIP status were determined by using the full RSF model, restricting the data by RKIP status, and calculating the joint importance score using random daughter assignment. If the sum of the importance scores for the RPMS genes considered separately is less than the joint importance score of all genes considered together, this difference was used as a measure of between-group cooperation. Results were based on 500 Monte Carlo replications.

RPMS pathway network analysis

The association of cancer-related gene sets to the RKIP metastasis pathway was tested by using a set of 65 KEGG cancer-related gene sets (see Supplementary data). To assemble the KEGG cancer-related gene sets, a KEGG gene set collection was obtained (http://www.broadinstitute.org/gsea/msigdb/collections.jsp), and we manually removed any KEGG gene sets that was specific for a non-cancer disease (e.g., 'cholera infection') or for a particular pathway/biological process judged to be not closely cancer-associated (e.g., 'caffeine metabolism'). We also included the proliferation meta-gene and various and disparate other common breast cancer prognostic gene signatures such as the MammaPrint 70 gene signature (van 't Veer *et al*, 2002), the wound signature (Chang *et al*, 2005), and an invasiveness gene signature (Liu *et al*, 2007). The BMS gene set was included to determine how it compares to the others. Using the BrCa443 data set, GSA was used to calculate which of the gene sets were associated with RKIP expression using an FDR cutoff of 0.05. As a second test criterion, the significance of the number of genes in each significant gene set that overlapped with either LET7-TG or BACH1-TG was determined by a hypergeometric test. Gene sets with an uncorrected *P*-value >0.05 were removed. Thus, only gene sets that pass both test criteria were kept. These gene sets were used as nodes in the network in addition to RKIP, LET7-TG, and BACH1-TG. Undirected edges were used to represent nodes with shared genes. In total, the network includes 28 nodes connected by 50 undirected edges.

In order to determine the clinical significance of the nodes in the RPMS network, we used the BrCa871 data set. For each patient, a value for each node in the network was assigned by calculating a meta-gene value similarly to LET7-TG and BACH1-TG. That is, the expression value of the genes with gene scores having the same sign as the GSA maxmean statistic from the BrCa443 data set were averaged and weighted by the individual gene scores. These meta-gene values were combined with the expression values for RKIP, LET7-TG, and BACH1-TG and used in an RSF model for metastasis-free survival. The relative importance scores of the nodes were used as a measure of how much each pathway in the network contributes to metastatic risk.

Sweave documentation and custom R packages

To ensure reproducibility, all procedures used for microarray data processing and normalization, and the code to reproduce all key findings in the manuscript are documented using Sweave (see Supplementary data). Functions, data, and meta-data needed to reproduce key results are provided as custom R packages.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (http://www.embojournal.org).

Acknowledgements

We thank S Gomes for excellent technical assistance and Dan Nicolae for helpful discussions. We also thank Robert Tibshirani for assistance with GSA. This work was supported by NIH Grants GM087630 and CA112310 (to MRR) and GM30998 (to WHL), the University of Chicago Breast Cancer SPORE (to MRR, AJM), and the Charlotte Geyer award (to MRR). MCB is supported by the Dutch Cancer Society. AJM is supported by a Department of Defense Era of Hope Scholar Award, NIH Grant 3P30CA016520-34S5, and the WW Smith Charitable Trust.

Author contributions: AJM and MRR were responsible for the experimental plan and wrote the manuscript; JY, CAF, W-LK, MCB, and NC carried out the experiments; JY and EME made the cell lines; HL and W-HL generated the list of potential let-7 target genes; and AJM and HI performed the statistical analyses.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Akaishi J, Onda M, Asaka S, Okamoto J, Miyamoto S, Nagahama M, Ito K, Kawanami O, Shimizu K (2006) Growth-suppressive function of phosphatidylethanolamine-binding protein in anaplastic thyroid cancer. *Anticancer Res* **26:** 4437–4442

Al-Mulla F, Hagan S, Behbehani AI, Bitar MS, George SS, Going JJ, Garcia JJ, Scott L, Fyfe N, Murray GI, Kolch W (2006) Raf kinase inhibitor protein expression in a survival analysis of colorectal cancer patients. *J Clin Oncol* **24:** 5672–5679

Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D, Joshi MB, Harpole D, Lancaster JM, Berchuck A, Olson Jr JA, Marks JR, Dressman HK, West M, Nevins JR (2006) Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* **439:** 353–357

Blenkiron C, Goldstein LD, Thorne NP, Spiteri I, Chin SF, Dunning MJ, Barbosa-Morais NL, Teschendorff AE, Green AR, Ellis IO, Tavare S, Caldas C, Miska EA (2007) MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype. *Genome Biol* **8:** R214

Bodenstine TM, Welch DR (2008) Metastasis suppressors and the tumor microenvironment. *Cancer Microenviron* **1:** 1–11

Boyerinas B, Park SM, Shomron N, Hedegaard MM, Vinther J, Andersen JS, Feig C, Xu J, Burge CB, Peter ME (2008) Identification of let-7-regulated oncofetal genes. *Cancer Res* **68:** 2587–2591

Breiman L (2001) Random forests. *Machine Learning* **45:** 5–32

 ©2011 European Molecular Biology Organization

Chambers AF (2009) MDA-MB-435 and M14 cell lines: identical but not M14 melanoma? *Cancer Res* **69:** 5292–5293

Chang HY, Nuyten DS, Sneddon JB, Hastie T, Tibshirani R, Sorlie T, Dai H, He YD, van 't Veer LJ, Bartelink H, van de Rijn M, Brown PO, van de Vijver MJ (2005) Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc Natl Acad Sci USA* **102:** 3738–3743

Chatterjee D, Sabo E, Tavares R, Resnick MB (2008) Inverse association between Raf kinase inhibitory protein and signal transducers and activators of transcription 3 expression in gastric adenocarcinoma patients: implications for clinical outcome. *Clin Cancer Res* **14:** 2994–3001

Cheadle C, Cho-Chung YS, Becker KG, Vawter MP (2003) Application of z-score transformation to Affymetrix data. *Appl Bioinformatics* **2:** 209–217

Corbit KC, Trakul N, Eves EM, Diaz B, Marshall M, Rosner MR (2003) Activation of Raf-1 signaling by protein kinase C through a mechanism involving Raf kinase inhibitory protein. *J Biol Chem* **278:** 13061–13068

Dangi-Garimella S, Yun J, Eves EM, Newman M, Erkeland SJ, Hammond SM, Minn AJ, Rosner MR (2009) Raf kinase inhibitory protein suppresses a metastasis signalling cascade involving LIN28 and let-7. *EMBO J* **28:** 347–358

Dohi Y, Ikura T, Hoshikawa Y, Katoh Y, Ota K, Nakanome A, Muto A, Omura S, Ohta T, Ito A, Yoshida M, Noda T, Igarashi K (2008) Bach1 inhibits oxidative stress-induced cellular senescence by impeding p53 function on chromatin. *Nat Struct Mol Biol* **15:** 1246–1254

Efron B, Tibshirani R (2007) On testing the significance of sets of genes. *Ann Appl Stat* **1:** 107–129

Efron B, Tibshirani R (2010) *GSA: Gene Set Analysis*. R package version 1.03

Enerly E, Steinfeld I, Kleivi K, Leivonen SK, Aure MR, Russnes HG, Ronneberg JA, Johnsen H, Navon R, Rodland E, Makela R, Naume B, Perala M, Kallioniemi O, Kristensen VN, Yakhini Z, Borresen-Dale AL (2011) miRNA-mRNA integrated analysis reveals roles for miRNAs in primary breast tumors. *PLoS One* **6:** e16915

Eves EM, Shapiro P, Naik K, Klein UR, Trakul N, Rosner MR (2006) Raf kinase inhibitory protein regulates aurora B kinase and the spindle checkpoint. *Mol Cell* **23:** 561–574

Fu Z, Kitagawa Y, Shen R, Shah R, Mehra R, Rhodes D, Keller PJ, Mizokami A, Dunn R, Chinnaiyan AM, Yao Z, Keller ET (2006) Metastasis suppressor gene Raf kinase inhibitor protein (RKIP) is a novel prognostic marker in prostate cancer. *Prostate* **66:** 248–256

Fu Z, Smith PC, Zhang L, Rubin MA, Dunn RL, Yao Z, Keller ET (2003) Effects of raf kinase inhibitor protein expression on suppression of prostate cancer metastasis. *J Natl Cancer Inst* **95:** 878–889

Gautier L, Cope L, Bolstad BM, Irizarry RA (2004) affy–analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20:** 307–315

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JYH, Zhang J (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5:** R80.

Granovsky AE, Rosner MR (2008) Raf kinase inhibitory protein: a signal transduction modulator and metastasis suppressor. *Cell Res* **18:** 452–457

Gupta GP, Nguyen DX, Chiang AC, Bos PD, Kim JY, Nadal C, Gomis RR, Manova-Todorova K, Massague J (2007) Mediators of vascular remodelling co-opted for sequential steps in lung metastasis. *Nature* **446:** 765–770

Hagan S, Al-Mulla F, Mallon E, Oien K, Ferrier R, Gusterson B, Garcia JJ, Kolch W (2005) Reduction of Raf-1 kinase inhibitor protein expression correlates with breast cancer metastasis. *Clin Cancer Res* **11:** 7392–7397

Ishwaran H, Kogalur UB (2008) *randomSurvivalForest: Ishwaran and Kogalur's Random Survival Forest*. R Package version 351

Ishwaran H, Kogalur UB, Gorodeski EZ, Minn AJ, Lauer MS (2010) HIgh-dimensional variable selection for survival data. *J Am Stat Assoc* **105:** 205–217

Ishwaran H, Kogular UB, Blackstone EH, Lauer MS (2008) Random survival forests. *Ann Appl Stat* **2:** 841–860

Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, Cheng A, Labourier E, Reinert KL, Brown D, Slack FJ (2005) RAS is regulated by the let-7 microRNA family. *Cell* **120:** 635–647

Kang Y, Siegel PM, Shu W, Drobnjak M, Kakonen SM, Cordon-Cardo C, Guise TA, Massague J (2003) A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* **3:** 537–549

Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M, Rajewsky N (2005) Combinatorial microRNA target predictions. *Nat Genet* **37:** 495–500

Kyo M, Yamamoto T, Motohashi H, Kamiya T, Kuroita T, Tanaka T, Engel JD, Kawakami B, Yamamoto M (2004) Evaluation of MafG interaction with Maf recognition element arrays by surface plasmon resonance imaging technique. *Genes Cells* **9:** 153–164

Lee YS, Dutta A (2007) The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. *Genes Dev* **21:** 1025–1030

Lewis BP, Burge CB, Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120:** 15–20

Li HZ, Gao Y, Zhao XL, Liu YX, Sun BC, Yang J, Yao Z (2009) Effects of raf kinase inhibitor protein expression on metastasis and progression of human breast cancer. *Mol Cancer Res* **7:** 832–840

Liaw A, Wiener M (2002) Classification and regression by randomForest. *R News* **2:** 18–22

Liu R, Wang X, Chen GY, Dalerba P, Gurney A, Hoey T, Sherlock G, Lewicki J, Shedden K, Clarke MF (2007) The prognostic role of a gene signature from tumorigenic breast-cancer cells. *N Engl J Med* **356:** 217–226

Lorenz K, Lohse MJ, Quitterer U (2003) Protein kinase C switches the Raf kinase inhibitor from Raf-1 to GRK-2. *Nature* **426:** 574–579

Massague J (2007) Sorting out breast-cancer gene signatures. *N Engl J Med* **356:** 294–297

Mayr C, Hemann MT, Bartel DP (2007) Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation. *Science* **315:** 1576–1579

McAllister SS, Gifford AM, Greiner AL, Kelleher SP, Saelzler MP, Ince TA, Reinhardt F, Harris LN, Hylander BL, Repasky EA, Weinberg RA (2008) Systemic endocrine instigation of indolent tumor growth requires osteopontin. *Cell* **133:** 994–1005

Minn AJ, Gupta GP, Padua D, Bos P, Nguyen DX, Nuyten D, Kreike B, Zhang Y, Wang Y, Ishwaran H, Foekens JA, van de Vijver M, Massague J (2007) Lung metastasis genes couple breast tumor size and metastatic spread. *Proc Natl Acad Sci USA* **104:** 6740–6745

Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, Giri DD, Viale A, Olshen AB, Gerald WL, Massague J (2005) Genes that mediate breast cancer metastasis to lung. *Nature* **436:** 518–524

Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L, Bayani N, Coppe JP, Tong F, Speed T, Spellman PT, DeVries S, Lapuk A, Wang NJ, Kuo WL, Stilwell JL, Pinkel D, Albertson DG, Waldman FM *et al* (2006) A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10:** 515–527

Nguyen DX, Chiang AC, Zhang XH, Kim JY, Kris MG, Ladanyi M, Gerald WL, Massague J (2009) WNT/TCF signaling through LEF1 and HOXB9 mediates lung adenocarcinoma metastasis. *Cell* **138:** 51–62

Pang H, Lin A, Holford M, Enerson BE, Lu B, Lawton MP, Floyd E, Zhao H (2006) Pathway analysis using random forests classification and regression. *Bioinformatics* **22:** 2028–2036

R core development team (2009) R: A language and environment for statistical and computing. *R Foundation for Statistical Computing*

Schneider BP, Winer EP, Foulkes WD, Garber J, Perou CM, Richardson A, Sledge GW, Carey LA (2008) Triple-negative breast cancer: risk factors to potential targets. *Clin Cancer Res* **14:** 8010–8018

Schuierer MM, Bataille F, Hagan S, Kolch W, Bosserhoff AK (2004) Reduction in Raf kinase inhibitor protein expression is associated with increased Ras-extracellular signal-regulated kinase signaling in melanoma cell lines. *Cancer Res* **64:** 5186–5192

Sethupathy P, Corda B, Hatzigeorgiou AG (2006) TarBase: a comprehensive database of experimentally supported animal microRNA targets. *RNA* **12:** 192–197

Therneau T, Lumley T (2009) Survival: survival analysis, including penalised likelihood. R Package version 235-4

©2011 European Molecular Biology Organization

Trakul N, Menard RE, Schade GR, Qian Z, Rosner MR (2005) Raf kinase inhibitory protein regulates Raf-1 but not B-Raf kinase activation. *J Biol Chem* **280:** 24931–24940

van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415:** 530–536

Weichselbaum RR, Ishwaran H, Yoon T, Nuyten DS, Baker SW, Khodarev N, Su AW, Shaikh AY, Roach P, Kreike B, Roizman B, Bergh J, Pawitan Y, van de Vijver MJ, Minn AJ (2008) An interferon-related gene signature for DNA damage resistance is a predictive marker for chemotherapy and radiation for breast cancer. *Proc Natl Acad Sci USA* **105:** 18490–18495

Wu J, Iwata F, Grass JA, Osborne CS, Elnitski L, Fraser P, Ohneda O, Yamamoto M, Bresnick EH (2005) Molecular determinants of NOTCH4 transcription in vascular endothelium. *Mol Cell Biol* **25:** 1458–1474

Wu W, Sun M, Zou GM, Chen J (2007) MicroRNA and cancer: current status and prospective. *Int J Cancer* **120:** 953–960

Yeung K, Seitz T, Li S, Janosch P, McFerran B, Kaiser C, Fee F, Katsanakis KD, Rose DW, Mischak H, Sedivy JM, Kolch W (1999) Suppression of Raf-1 kinase activity and MAP kinase signalling by RKIP. *Nature* **401:** 173–177

Yeung KC, Rose DW, Dhillon AS, Yaros D, Gustafsson M, Chatterjee D, McFerran B, Wyche J, Kolch W, Sedivy JM (2001) Raf kinase inhibitor protein interacts with NF-kappaB-inducing kinase and TAK1 and inhibits NF-kappaB activation. *Mol Cell Biol* **21:** 7207–7217

Zeng L, Imamoto A, Rosner MR (2008) Raf kinase inhibitory protein (RKIP): a physiological regulator and future therapeutic target. *Expert Opin Ther Targets* **12:** 1275–1287

 ©2011 European Molecular Biology Organization