An integrated systems genetics screen reveals the transcriptional structure of inherited predisposition to metastatic disease

Farhoud Faraji,1,2,3 Ying Hu,4 Gang Wu,5 Natalie E. Goldberger,1 Renard C. Walker,1 Jinghui Zhang,5 and Kent W. Hunter1,6

1Metastasis Susceptibility Section, Laboratory of Cancer Biology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA; 2Department of Biochemistry and Molecular Biology, School of Medicine, Saint Louis University, Saint Louis, Missouri 63104, USA; 3Howard Hughes Medical Institute, National Institutes of Health Research Scholars Program, Chevy Chase, Maryland 20815, USA; 4Laboratory of Population Genetics, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA; 5Department of Computational Biology, St. Jude Children’s Research Hospital, Memphis, Tennessee 38105, USA

Metastasis is the result of stochastic genomic and epigenetic events leading to gene expression profiles that drive tumor dissemination. Here we exploit the principle that metastatic propensity is modified by the genetic background to generate prognostic gene expression signatures that illuminate regulators of metastasis. We also identify multiple microRNAs whose germline variation is causally linked to tumor progression and metastasis. We employ network analysis of global gene expression profiles in tumors derived from a panel of recombinant inbred mice to identify a network of co-expressed genes centered on Cnot2 that predicts metastasis-free survival. Modulating Cnot2 expression changes tumor cell metastatic potential in vivo, supporting a functional role for Cnot2 in metastasis. Small RNA sequencing of the same tumor set revealed a negative correlation between expression of the Mir216/217 cluster and tumor progression. Expression quantitative trait locus analysis (eQTL) identified cis-eQTLs at the Mir216/217 locus, indicating that differences in expression may be inherited. Ectopic expression of Mir216/217 in tumor cells suppressed metastasis in vivo. Finally, small RNA sequencing and mRNA expression profiling data were integrated to reveal that miR-3470a/b target a high proportion of network transcripts. In vivo analysis of Mir3470a/b demonstrated that both promote metastasis. Moreover, Mir3470b is a likely regulator of the Cnot2 network as its overexpression down-regulated expression of network hub genes and enhanced metastasis in vivo, phenocopying Cnot2 knockdown. The resulting data from this strategy identify Cnot2 as a novel regulator of metastasis and demonstrate the power of our systems-level approach in identifying modifiers of metastasis.

[Supplemental material is available for this article.]
globally interrogate transcriptional determinants of metastasis susceptibility. RI panels are specialized sets of inbred mice generated from two inbred strains produced by 20 or more generations of brother-sister inbreeding (Fig. 1A). Each of the resulting sublines is a distinct inbred composite of two previously established parental inbred mouse lines. RI panels are particularly useful for mapping inherited components for highly variable quantitative phenotypes, such as metastatic dissemination. Variation resulting from stochastic events can be reduced by phenotyping multiple isogenic individuals within each subline of the RI panel. This strategy results in a better estimation of genetic effects and increases power to detect genotype-phenotype associations compared with standard genetic mapping strategies, in which every individual animal is genetically unique.

The AKXD RI panel was derived from the AKR/J and DBA/2J inbred strains (Mucenski et al. 1998). Previous work in our laboratory has demonstrated that, when crossed to highly metastatic male MMTV-PyMT transgenic mice (Guy et al. 1992), the progeny of these two strains exhibit >20-fold difference in metastatic propensity (Lifsted et al. 1998). Crossing each AKXD subline to MMTV-PyMT transgenic mice (denoted henceforth as [AKXDn × PyMT]F1) resulted in progeny with a similar range of variability in metastatic susceptibility across the AKXD panel (Fig. 1A). Since the oncogenic driver (PyMT) and paternal genetic background (FVB/NJ) are identical among all progeny, the phenotypic diversity in metastatic susceptibility is most likely a result of germline polymorphism in the maternal genome. As such, we selected [AKXDn × PyMT]F1 progeny as a genetically defined vehicle for identification of heritable metastasis-associated transcriptional networks.

It has been demonstrated that microRNAs (miRNAs) play a key role in the initiation and progression of multiple tumor types (Tavazoie et al. 2008; Pencheva and Tavazoie 2013). Our group recently showed that heritable differences in miRNA expression can modify gene expression patterns and result in differences in metastatic susceptibility (Faraji et al. 2012; Hu et al. 2012; Winter et al. 2012). Based on this principle, we analyzed gene expression profile data from two to seven tumors from each of the [AKXDn × PyMT]F1 progeny on Affymetrix MOE430 microarrays (Yang et al. 2005). Permutation tests of the gene expression profiles revealed 20 probe sets representing 17 distinct genes passing the genome-wide significance threshold for metastatic progression ($P < 0.05$) (Supplemental Table 1). The gene most strongly correlated with the metastatic phenotype was the C1 protease cathepsin J (Ctsj), suggesting a possible role for protein degradation in metastatic progression. No obvious overrepresentation of any single

**Results**

**Correlation of gene expression and metastatic efficiency**

We have previously demonstrated that germline polymorphisms can modify gene expression patterns and result in differences in heritable metastatic susceptibility (Faraji et al. 2012; Hu et al. 2012; Winter et al. 2012). Based on this principle, we analyzed gene expression profile data from two to seven tumors from each of the [AKXDn × PyMT]F1 progeny on Affymetrix MOE430 microarrays (Yang et al. 2005). Permutation tests of the gene expression profiles revealed 20 probe sets representing 17 distinct genes passing the genome-wide significance threshold for metastatic progression ($P < 0.05$) (Supplemental Table 1). The gene most strongly correlated with the metastatic phenotype was the C1 protease cathepsin J (Ctsj), suggesting a possible role for protein degradation in metastatic progression. No obvious overrepresentation of any single

**Figure 1.** Genome-wide strategy to identify metastasis susceptibility co-expressed networks and their post-transcriptional regulators. (A) The pre-existing AKXD recombinant inbred (RI) panel was constructed by breeding metastasis-prone inbred mice from the AKR/J background to metastasis-resistant DBA/2J mice. F1 progeny from the AKR/J and DBA/2J cross were intercrossed; F2 progeny were bred to homozygosity, generating 24 isogenic AKXD sublines. Crossing each AKXD subline to the MMTV-PyMT model for highly metastatic breast cancer revealed phenotypic diversity in metastatic susceptibility within the AKXD RI panel. (B) RNA purified from AKXD subline primary tumors was subjected to global mRNA profiling by microarray followed by WGCNA and Kaplan-Meier Analysis to identify metastasis-associated co-expressed transcriptional modules. (C) AKXD subline tumor RNA was also subjected to small RNA sequencing followed by miRNA expression-phenotype correlations and eQTL analysis to identify metastasis susceptibility miRNAs. (D) Workflow illustrating analytic integration of global mRNA and miRNA profiling methods to uncover candidate post-transcriptional regulators of co-expressed networks.
functional pathway was observed among the 17 genes, nor were any of the previously identified metastasis susceptibility genes identified in this genetic reference panel. With the exception of Ms4a6d, none of the metastasis-associated genes were linked with regions of the genome previously shown to harbor metastasis susceptibility genes (Hunter et al. 2001; Lancaster et al. 2005; Crawford et al. 2008b; KW Hunter, unpubl.). To test the relevance to human disease, the human homolog to each gene was queried using the Gene expression-based Outcome for Breast cancer Online (GOBO, http://co.bmc.lu.se/goibo/) (Ringner et al. 2011). Of the 17 murine genes, 12 were represented by probes to the human homologs. Of these 12 genes, expression of seven (58%) segregated high- and low-risk groups with respect to distant metastasis-free survival (DMFS) (Supplemental Fig. 1). These data indicate that Csf1, Rngtt, Argc, Umod, Pcpp4, Slc25a5, and Tob1 are mouse candidate metastasis-associated genes predictive of DMFS in humans. Intriguingly, Tob1 and Csf1 are regulators of polyadenylation tail length (Takagaki and Manley 2000; Ezzeddine et al. 2007) and Rngtt is an mRNA 5′-guanylyltransferase capping enzyme (Pilhuta et al. 1998). To our knowledge, this is the first report implicating modulation of mRNA processing and stability in metastasis.

Analysis of expression data reveals modulation of network expression by previously identified metastasis susceptibility genes

We next undertook gene expression network analyses to move beyond single gene correlations and reveal higher-order expression patterns driving metastatic susceptibility. To define our approach, we proposed that modules of co-expressed genes can indicate co-regulated gene networks comprising functionally coherent molecular pathways (Tavazoie et al. 1999). Network-level analysis of expression data was achieved by examining gene expression network modules using weighted gene co-expression network analysis (WGCNA) based on topological overlap measure (TOM) algorithms (Li and Horvath 2009). Assuming parsimonious network structure, the generated expression networks were visualized with minimum spanning tree algorithms. Thirteen modules containing between 41 and 1085 co-expressed genes were identified (Supplemental Figs. 2–14). For clarity, modules were named for the most highly connected gene within each network.

Encouragingly, we noted the membership of several previously identified and in vivo validated metastasis modifier genes in our networks. The Ndn module was found to contain Arap3 (previously known as Centd3), P16, Ndn, and Csf1r, two of which (Ndn and Arap3) occupied central, highly connected hub positions (Supplemental Fig. 2). These four genes were previously found to modulate metastatic efficiency in vivo (Crawford et al. 2008b). In addition, the Cnot2 module (Supplemental Fig. 3) was found to contain Arid4b and Luc7l while the Pol2g module (Supplemental Fig. 7) contained Brd4 (Crawford et al. 2008a; Winter et al. 2012). The membership of these bona fide metastasis modifier genes within our networks provided increased confidence for our systems-level approach to identifying metastasis-driving transcriptional networks.

Network modules predict outcome in human breast cancer data sets

We next turned to two human breast cancer gene expression data sets to determine if the individual transcription networks were also associated with human metastatic disease. GSE2034 (Wang et al. 2005) and GSE11121 (Schmidt et al. 2008) are gene expression data sets from node-negative breast cancer patients who did not receive adjuvant therapy and for which DMFS data are available. The individual mouse module gene sets were converted to human probe sets using the NetAffx Batch Query tool (Affymetrix.com). Probe sets representing “hub transcripts” (greater than or equal to five connections), which capture the majority of the module gene expression variability (Hu et al. 2012), were used to generate gene signatures to assess the relative impact of each module on DMFS by Kaplan-Meier analysis (Table 1). The individual Cnot2, Khh12, Katna1, Chuc2, Polr2g, and modules significantly discriminated DMFS in both data sets, and are therefore the highest confidence metastasis-associated modules. The Sephs1 and Naa35 modules were significant for GSE2034 and borderline for GSE11121, while the Rad51 and Traf7 modules were significant for GSE2034 but not GSE11121. Conversely, the Fabp4 and Eno3 modules were significant for GSE11121 but not GSE2034. The Ndb module was borderline significant for both, while the Gsk3b module was not significant for either data set (Table 1; Supplemental Figs. 3–28). The lack of concordance for these modules between the two data sets can be explained by statistical fluctuations, differences in the number of samples, or undefined differences in sample composition in the breast cancer data sets. Since human expression data sets are thought to be significantly underpowered to detect stable gene signatures associated with survival (Ein-Dor et al. 2006), we believe that at least some of these modules play a role in metastatic disease and, thus, were included in subsequent computational analyses.

Since each of these statistically derived modules may represent a co-regulated transcription program, we next performed gene set enrichment analysis (GSEA) (Subramanian et al. 2005) to determine if any module acted to maintain specific cellular functions and to identify common underlying functional pathways that might be associated with metastatic progression. Six of the 13 modules were found to be most highly associated with the nucleus. mRNA transport, zinc ion binding, and chromatin modification were also represented by more than one module (Table 1). Together, these results suggest that nuclear biology plays a role in establishment of metastatic predisposition. This idea is supported by previous identification of the metastasis susceptibility genes Brd4, a transcriptional elongation factor and chromatin reader, Bpt1b, a facultative heterochromatin protein, Ndn, a transcription factor, and Arid4b, a member of the Sin3a histone deacetylase complex (Crawford et al. 2007; Crawford et al. 2008a,b; Winter et al. 2012).

Validation of the Cnot2 module co-regulation

The network showing the highest reproducible statistical correlation to DMFS, Cnot2 (Fig. 2A), was selected for biological validation. We hypothesized that the module could be a co-regulated expression unit potentially influenced by the expression of Cnot2, which might serve as a surrogate gene for the network. To test this idea, nine network hub genes (Atg12, Rik1, Kcc1, Angel2, Polrmt, Tmem101, Exoc1, Utp6, Ubx2r2) (highlighted in gray in Fig. 2A) were selected to capture maximum network variance. The expression of these nine hub genes, representing 52% (114 of 217 genes) of the connectivity of the entire module, was assessed by quantitative real time PCR (qRT-PCR). Ectopic expression of Cnot2 in 6DT1 murine mammary tumor cells significantly up-regulated five of nine hub genes (Atg12, Rik1, Kcc1, Angel2, Polrmt) (Fig. 2B), indicating that overexpression of Cnot2 within physiological
Table 1. Summary of co-expressed modules

| Module | Number of genes | Kaplan-Meier analysis | Molecular function | Cellular process | Cell component |
|--------|-----------------|------------------------|--------------------|------------------|---------------|
|        |                 | GSE2034 | GSE11121 | Description | P-value | Description | P-value | Description | P-value |
| Cnot2  | 216             | 0.004   | 0.004   | Methyltransferase activity | 2.66 x 10^-3 | Response to DNA damage stimulus | 1.07 x 10^-3 | Mitochondrion | 2.63 x 10^-6 |
| Klhl12 | 1085            | <0.001  | 0.014   | Zinc ion binding | 1.81 x 10^-7 | mRNA processing | 2.75 x 10^-5 | Nucleus | 2.06 x 10^-6 |
| Katna1 | 220             | 0.042   | 0.004   | Ligase activity | 2.67 x 10^-4 | Cell division | 3.10 x 10^-5 | Nucleus | 1.63 x 10^-8 |
| Chac2  | 218             | <0.001  | 0.05    | Helicase activity | 3.76 x 10^-3 | Chromatin modification | 1.23 x 10^-2 | Golgi apparatus | 2.24 x 10^-2 |
| Polr2g | 376             | 0.05    | 0.008   | SNAP receptor activity | 1.26 x 10^-6 | mRNA transport | 9.71 x 10^-5 | Nucleus | 1.17 x 10^-4 |
| Sephs1 | 146             | <0.001  | 0.054   | Guanyl-nucleotide exchange factor activity | 7.47 x 10^-4 | Intracellular protein transport | 2.41 x 10^-3 | Nucleus | 3.88 x 10^-3 |
| Mak10  | 219             | 0.002   | 0.052   | Cysteine-type peptidase activity | 3.25 x 10^-3 | Protein transport | 4.12 x 10^-7 | Cytoplasm | 2.82 x 10^-5 |
| RadS1  | 87              | 0.016   | 0.178   | ATP binding | 1.53 x 10^-5 | Cell division | 7.53 x 10^-14 | Nucleus | 3.10 x 10^-9 |
| Traf7  | 199             | 0.022   | 0.224   | RNA binding | 1.04 x 10^-3 | Metabolic process | 1.51 x 10^-3 | Nucleolus | 2.71 x 10^-9 |
| Fatp4  | 124             | 0.362   | 0.024   | Aciyltransferase activity | 2.07 x 10^-4 | Gluconeogenesis | 6.44 x 10^-9 | Extracellular region | 1.54 x 10^-9 |
| Emo3   | 41              | 0.118   | 0.04    | — | — | — | — | Mitochondrion | 9.10 x 10^-3 |
| Ndn    | 98              | 0.056   | 0.058   | Receptor activity | 4.48 x 10^-6 | Immune response | 9.06 x 10^-4 | Extracellular region | 8.85 x 10^-8 |
| Gsk3b  | 148             | 0.076   | 0.206   | Zinc ion binding | 4.67 x 10^-5 | Chromatin modification | 3.34 x 10^-5 | Nucleus | 4.90 x 10^-5 |

Minimum of five genes per GO category.
limits (~2–3 fold) resulted in co-upregulation of greater than half of the tested genes within the Cnot2 module and suggesting that at least a subset of genes within the Cnot2 network is part of a regulatory module potentially controlled by Cnot2 expression.

**In vivo validation of Cnot2 in metastatic progression**

The Cnot2 network signature stratified breast cancer patients into DMFS high- and low-risk groups (Fig. 2C). To further probe the functional role of variations in Cnot2 expression in tumor progression a search of publicly available human breast cancer microarray data sets using Oncomine (Compendia Bioscience) was performed, revealing that CNOT2 expression was 16-fold lower in 53 invasive breast carcinoma samples compared with six normal breast tissue samples (Fig. 2D; Finak et al. 2008). Also, comparison of 67NR and 4T1 cells, clones with distinct metastatic potential isolated from the same spontaneously formed murine mammary tumor (Aslakson and Miller 1992), showed that Cnot2 along with four module hub genes

---

**Figure 2.** A transcriptional module centered on Cnot2 shows regulation by Cnot2 expression and predicts outcome in human breast cancer cohorts. (A) The Cnot2 module. Circles indicate individual genes. Connections between the genes were generated by minimizing the number of connections necessary to explain the gene expression correlations. (B) The effect of Cnot2 overexpression in 6DT1 cells on network hub transcripts. (*) $P < 0.05$, (**) $P < 0.01$. (C) A gene signature generated from homologous Cnot2 module hub transcripts predicts survival in human breast cancer cohorts. (D) Oncomine data set (Finak et al. 2008) shows CNOT2 is down-regulated in invasive breast carcinoma relative to normal breast tissue (normal breast tissue $n = 6$, invasive breast carcinoma $n = 53$; fold change $= -16.3$; $P = 1.12 \times 10^{-3}$).
were down-regulated in highly metastatic 4T1 cells relative to nonmetastatic 67NR cells (Supplemental Fig. 29). Taken together, these data show a consistent negative correlation between Cnot2 expression and tumor progression in independent mouse breast tumor model systems and human patient data.

To test if variation in Cnot2 expression has a direct, causative role in tumor progression, 6DT1 cells overexpressing Cnot2 were orthotopically implanted into syngeneic immune competent FVB/NJ female mice. Assessment at 30 d post-implantation demonstrated a reduction in primary tumor mass and a pronounced suppression of pulmonary metastases (Fig. 3A). 6DT1 cells were also transduced to stably express short hairpin RNA constructs targeting Cnot2, achieving modest knockdown of Cnot2. Cnot2 knockdown resulted in a significant increase in pulmonary metastasis but had no effect on primary tumor burden in our model (Fig. 3B,C). To increase confidence in the generalizability of our results, the orthotopic implantation assay was repeated on an independent murine mammary tumor cell line, Mvt1. Knockdown of Cnot2 in Mvt1 cells enhanced both pulmonary metastasis and primary tumor mass (Supplemental Fig. 30), consistent with the activity of a progression gene in this cell line. These results validate the causative role of Cnot2 as a negative regulator of tumor progression.

Identification of inherited metastasis susceptibility microRNAs

By definition, heritable metastasis modifier miRNAs display allelic variations in expression that dictate differences in metastatic susceptibility (Goldberger et al. 2013). We proposed that the highest confidence heritable miRNAs would contain cis-expression quantitative trait loci (eQTL), as these are most likely to result from heritable variations that directly and locally influence transcript levels (Doss et al. 2005). In order to uncover such miRNAs we undertook high throughput sequence analysis of small RNAs derived from [AKXDn × PyMT]F1 tumors (Fig. 1C) followed by eQTL analysis (Schadt et al. 2003) in the AKXD RI panel.

Comparison of expression levels of significantly expressed miRNAs to the genetic map resulted in the detection of seven distinct eQTLs for 13 miRNAs exceeding the genome-wide statistical threshold for metastatic progression (Table 2). Interestingly, the human homologs of eight of these 13 miRNAs (miR-671, miR-99a, miR-140, miR-582, miR-186, miR-216b, miR-23a, and miR-544) have previously been implicated in metastatic dissemination (Deng et al. 2011; Haga and Phinney 2012; Jahid et al. 2012; Rutnam and Yang 2012; Kuo et al. 2013; Li et al. 2013a,d; Uchino et al. 2013; Yang et al. 2013a). Further, the highest confidence eQTL observed was for miR-200c*, a poorly characterized member of the Mir200c/141 cluster, which is thought to drive metastatic dissemination by inducing tumor cells to undergo epithelial-to-mesenchymal transition (Gregory et al. 2008).

The sequence data were then queried for miRNAs whose expression directly correlated with metastatic propensity across the AKXD RI panel. miR-216a, miR-216b, and miR-217 showed significant inverse correlations to metastasis (Fig. 4A), indicating that inherited variation in the expression of this set of miRNAs contributes to differences in metastatic susceptibility across the AKXD panel.

Comparison of human and mouse Mir216/217 sequences revealed a high degree of conservation of pre-miRNA and mature miRNAs, with 100% sequence identity for mature miR-216a and miR-217 and 96% identity for mature miR-216b (Supplemental Fig. 32). In addition, the target prediction tool TargetScan (Lewis et al. 2005) showed significant overlap in targets in mice and humans (Supplemental Fig. 33). At least one EST (DA732292 in humans) encompasses all three miRNAs, suggesting this miRNA cluster is expressed from the same primary transcript (Supplemental Fig. 34), which raised the possibility that repression of this transcript may cause a synergistic enhancement of tumor cell metastatic potential. To test this idea, the Broad Institute miRNA target prediction tool was applied to predict and combine all miR-216/217 putative targets. A gene signature was then developed from predicted targets that showed significant variance in two independent mouse tumor data sets (Diversity Outcross and MOLF.

**Figure 3.** Cnot2 expression suppresses tumor progression. (A) Overexpression of Cnot2 in 6DT1 cells showed 22% reduction in primary tumor mass (P = 0.021) and 46% reduction in pulmonary metastases (P = 0.006, n = 20 per group). (B) Short hairpin-mediated knockdown Cnot2 by 51% (sh62) and 42% (sh64) in 6DT1 cells resulted in no significant change in primary tumor mass but enhanced metastatic potential respectively by 183% (P = 0.002) and 87% (P = 0.026, ncontrol = 18 mice, nsh62 = 10 mice, nsh64 = 9 mice). (C) Representative lungs from knockdown experiment sectioned and stained with hematoxylin and eosin (H&E). (*) P < 0.05, (**) P < 0.01.

232 Genome Research www.genome.org
Systems analysis of metastasis predisposition

We searched within all networks generated from expression profiling of [AKXD × PyMT]F1 tumors, focusing on genes that capture the majority of the network variability by restricting analysis to hub transcripts. A gene signature was constructed from the 153 hub transcripts (Supplemental Table 16) and screened for the ability to discriminate patient outcome. This gene signature was found to discriminate patient outcome in both GSE2034 and GSE11121 data sets (Supplemental Fig. 38), suggesting that these 153 hub transcripts were significantly associated with DMFS.

Both mRNA and miRNA genome-wide screens were performed on RNA from the same tumors. The data captured by miRNA expression array and miRNA sequencing could thus be integrated for genome-wide query of potential miRNA–mRNA target pairs by identifying negatively correlated miRNA–gene pairs. Correlations with a P-value ≤ 0.01 after 10,000 permutations were considered significant, producing a total of 16,475 miRNA–gene pairs consisting of 369 miRNAs in negative correlation with 4153 genes (Supplemental Table 17). The resulting data were then subjected to three filters. First, to increase the probability of identifying metastasis-driver microRNAs, all subsequent analysis was restricted to the 153 metastasis-associated hub transcripts. Second, the results were filtered for genes with shared computationally predicted MREs. Third, those miRNAs targeting the filtered set of miRNAs more frequently than predicted by chance were selected as candidate metastasis-driving miRNAs.

The first two filters yielded 38 miRNAs putatively targeting 49 of the 153 highly connected hub genes (Supplemental Table 18). Of these 38 candidate miRNAs 25 have known human homologs, of which 11 have previously been identified to directly act in the metastatic cascade in various tumor types (let-7i, miR-30c, miR-31, miR-127, miR-134, miR-140, miR-148a, miR-196b, miR-210, miR-503, miR-582) (Valastyan et al. 2009; Li et al. 2010; Liang et al. 2011; Ying et al. 2011; Zhou and Wang 2011; Bockhorn et al. 2013; Li et al. 2013c; Liu et al. 2013; Uchino et al. 2013; Yang et al. 2013a,b) while several other miRNAs have been associated with distant metastatic disease (miR-152, miR-298, miR-326, miR-423) (Farazi et al. 2011; Bao et al. 2012; Hui et al. 2013; Valencia et al. 2013). Moreover, as shown above, miR-140 and miR-582 were also found to contain a trans-eQTL (Table 2) implicating them as candidate heritable metastasis susceptibility miRNAs, which is consistent with the notion that these miRNAs target heritable metastasis susceptibility transcriptional networks. In total, 15 (60%) of 25 human homologs of murine miRNAs identified by

---

**Table 2. miRNA eQTLs**

| miRNA   | Chromosome | Position (Mb) | Locus          | eQTL       | Chromosome | Position (Mb) | LOD score | Permutation P-value |
|---------|------------|---------------|----------------|------------|------------|---------------|-----------|---------------------|
| miR-200c* | 6          | 124.7         | rs6248036      |            | 5          | 21.5          | 5.01      | 0                   |
| miR-671-3p | 5          | 24.1          | rs4188505      |            | 16         | 23.9          | 5.25      | 0.002               |
| miR-331-5p | 10         | 93.4          | rs6248036      |            | 5          | 21.5          | 5.6       | 0.004               |
| mir-217   | 11         | 28.7          | rs13480929     |            | 11         | 14           | 6.71      | 0.006               |
| mir-99a   | 16         | 77.6          | rs3657839      |            | 8          | 60.5          | 4.98      | 0.011               |
| mir-216a  | 11         | 28.7          | rs13480929     |            | 11         | 14           | 7.16      | 0.013               |
| mir-140   | 8          | 110.1         | rs6352812      |            | 9          | 44           | 5.45      | 0.016               |
| mir-582-5p | 13         | 110.1         | rs3657839      |            | 8          | 60.5          | 4.33      | 0.018               |
| mir-186   | 3          | 157.2         | rs3657839      |            | 8          | 60.5          | 4.72      | 0.02                |
| mir-216b  | 11         | 28.7          | rs13480929     |            | 11         | 14           | 5.4       | 0.025               |
| mir-25a   | 8          | 86.7          | rs3476243      |            | 1          | 85.8          | 4.34      | 0.032               |
| mir-376c  | 12         | 111           | rs6300613      |            | 8          | 62.4          | 4.53      | 0.041               |
| mir-544   | 12         | 111           | rs3660779      |            | 2          | 29.5          | 14        | 0.05                |

---

[Hu et al. 2013, Supplemental Table 15]. This miR-216/217 target gene signature demonstrated statistically significant stratification of metastatic risk in murine mammary tumors as well as in human breast cancer (Supplemental Figs. 35, 36). Interestingly, the human data demonstrated statistically significant stratification of metastatic risk specifically in estrogen receptor (ER) negative tumors.

To test the direct causative role of miR-216a, miR-216b, and miR-217 in breast cancer, 6DT1 cells were transduced to stably express pre-miRNA sequences (Fig. 4C). As shown in Figure 4D, orthotopic implantation of 6DT1 cells ectopically expressing each miRNA resulted in minimal effect on primary tumor mass although expression of Mir216b and Mir217 trended toward reduction of primary tumor growth. Despite insignificant effect on tumor growth, Mir216b and Mir217 expression in 6DT1 cells resulted in dramatic suppression of pulmonary metastasis (Fig. 4D).

The Mir216/217 cluster predicted target signature stratified metastatic risk in basal-like murine tumors (Hu et al. 2013) as well as ER-negative human tumors. We thus also tested the Mir216a, Mir216b, and Mir217 expression in the 4T1 cell line, an ER-negative murine tumor model (Kaur et al. 2012). Orthotopic implantation of 4T1 cells stably expressing Mir216a, Mir216b, or Mir217 showed statistically significant tumor suppression with a dramatic reduction in tumor mass for 4T1 cells expressing Mir216a and a more moderate tumor suppressive effect for cells expressing Mir216b (Supplemental Fig. 37). In 4T1 cells, all three miRNAs showed statistically significant inhibitory effects on metastasis. Despite observed differential effects on primary tumor growth between 6DT1 and 4T1 cells, expression of the Mir216/217 cluster showed similar suppressive effects on pulmonary metastasis. This directional concordance between genetic associations derived from sequence analysis in AKXD mice with in vivo metastasis assays in two independent mammary tumor cell lines indicates that the Mir216/217 cluster is a heritable suppressor of tumor progression.

Integrated analysis of gene expression profiling and miRNA expression data

To gain insight into post-transcriptional regulation of the co-expressed modules associated with metastasis susceptibility, we undertook an approach that integrated our mRNA and small RNA profiling data. We hypothesized that querying transcripts for the presence of common microRNA recognition elements (MREs) might yield metastasis master regulator miRNAs (Fig. 1D).
Figure 4. *Mic216/217* are inherited metastasis suppressors. (A) Expression-phenotype correlations for *Mic216a*, *Mic216b*, and *Mic217* in [AKXD<sub>3</sub> x PyMT]<sub>1</sub> tumors. Blue, yellow, and red points indicate AKXD sublines with respective weak, intermediate, and high metastatic propensities. (B) miR-eQTL analysis demonstrated peaks reaching genome-wide significance (red line) for *Mic216a* and *Mic217* and reaching the "suggestive" threshold (gray line) for *Mic216b*. Consistent with a cis-eQTL, these peaks co-localize to the *Mic216/217* locus on chromosome 11. Please refer to Supplemental Figure 31 for a detailed, high-resolution format of eQTL plots. (C) Ectopic expression of *Mic216a*, *Mic216b*, and *Mic217* in 6DT1 cells. (D) Orthotopic implantation of 6DT1 cells expressing control miRNA, *Mic216a*, *Mic216b*, or *Mic217*. *Mic216b* and *Mic217* expression reduced pulmonary metastasis by 79% and 80%, respectively (p<sub>*Mic216b*</sub> = 0.011, p<sub>*Mic217*</sub> = 0.022; n = 10 mice per group). (*) P < 0.05.
integrating transcriptional network analysis with small RNA sequencing data have previously been implicated in metastasis.

Unexpectedly, the third filter revealed that 32 (65%) of the 49 target genes were predicted targets of miR-3470a or miR-3470b, suggesting that this miRNA family plays an important role in the establishment of metastatic susceptibility. Fourteen genes were predicted targets for both miR-3470a and miR-3470b and each miRNA targeted two hubs in the Cnot2 network (Supplemental Fig. 39). miR-3470a and miR-3470b were therefore selected for validation as candidate metastasis-driver microRNAs.

In vivo validation of miR-3470a and miR-3470b as mouse-specific promoters of tumor progression

We initially turned to 67NR and 4T1 cells, two cell lines with distinct differences in metastatic potential, to verify the biological role of Mir3470a and Mir3470b in metastasis. Measurement of Mir3470a and Mir3470b expression by miR-Taqman qRT-PCR revealed that these miRNAs were more highly expressed in highly metastatic 4T1 cells than nonmetastatic 67NR cells (Supplemental Fig. 40), indicating that these miRNAs may be pro-metastatic.

As all previous in vivo analyses and Cnot2 network validation were conducted in 6DT1 cells, we next turned this cell line to test direct, causative effects of Mir3470a/b on network structure, gene expression, and tumor cell metastatic potential. We generated 6DT1 cells stably expressing Mir3470a and Mir3470b and confirmed Mir3470a/b expression by qRT-PCR, demonstrating sevenfold overexpression of Mir3470a and 88-fold overexpression of Mir3470b (Fig. 5A). Assessment of hub transcript expression levels upon Mir3470a overexpression demonstrated that three (Chac2, Dnpep, Arfp1) of the nine tested hub transcripts predicted to be miR-3470a/b targets were down-regulated upon miR-3470a overexpression while seven (Ap2a2, Insig2, Chac2, Ddx46, Dnpep, Cnot8, Arfp1) of the nine tested hub transcripts were down-regulated upon miR-3470b overexpression (Supplemental Fig. 41). Since miR-3470a and miR-3470b possess the same seed sequence—the 5’-end miRNA sequence thought to contribute most strongly to miRNA recognition of target transcripts—it is possible that the 10-fold difference in expression between Mir3470a and Mir3470b-

Figure 5. Mir3470a and Mir3470b promote metastatic progression. (A) Stable transduction of Mir3470a and Mir3470b in 6DT1 cells. (B) Orthotopic implantation of 6DT1 cells overexpressing Mir3470a showed 40% increase in tumor mass (P = 0.014) and 230% increase in pulmonary metastasis (P = 0.0069). Overexpression of Mir3470b resulted in 213% increase in pulmonary metastasis (P = 0.088; ncontrol = 9 mice, nMir3470a = 10 mice, nMir3470b = 10 mice). (C) Representative H&E stained lung sections from mice implanted with control or cells overexpressing Mir3470a/b. (*) P < 0.05, (**) P < 0.01.
expressing cells explains the difference in predicted target downstream regulation.

In addition, Mir3470a and Mir3470b overexpression down-regulated multiple network hub genes within the Cnot2 module (Cnot2, 2mem401, Utp6, Exoc1, Polmt, Ube2r2, and Angel2), only some of which are predicted to be miR-3470a/b targets (Supplemental Fig. 42). Intriguingly, Mir3470b overexpression appeared to have a unidirectional effect on previously identified metastasis susceptibility genes, up-regulating the pro-metastatic Brd4 short isoform (Brd4-SF) while down-regulating anti-metastatic genes: Brd4 long isoform (Brd4-LF), Rpl1b, and Cadm1 (Supplemental Fig. 43; Crawford et al. 2007, 2008a; Alsarraj et al. 2011; Faraji et al. 2012). Consistent with the correlative data in 67NR and 4T1 cells, the negative regulatory effect of Mir3470a/b on Cnot2 module transcripts and the changes in expression of previously identified metastasis susceptibility genes were suggestive of a pro-metastatic role for Mir3470a/b.

We next implanted Mir3470a or Mir3470b overexpressing 6DT1 cells into syngeneic mice to test the direct effect of miR-3470a/b on tumor progression and metastasis. Orthotopic implantation of Mir3470a overexpressing 6DT1 cells into mammary fat pads of immune competent FVB/NJ mice resulted in mild but statistically significant enhancement in tumor mass and a dramatic increase in pulmonary metastases while Mir3470b overexpressing 6DT1 cells showed specific promotion of metastatic colonization (Fig. 5B,C). Consistent with gene expression correlation studies in 67NR and 4T1 cells, as well as expression impact on the Cnot2 module, these in vivo results confirm Mir3470a as a promoter of tumor progression, and identify Mir3470b as a metastas-miRNA (Hurst et al. 2009).

Discussion

The systems genetics analysis performed here provides a view of the transcriptional networks and corresponding post-transcriptional regulatory mechanisms driving metastatic breast cancer. Rather than just attempting to identify single genes acting in metastasis, we undertook a higher-order approach to gain insight into the transcriptional profiles potentially driving metastatic susceptibility by emphasizing connectivity and transcript-level interactions (Fig. 6A). Networks of co-expressed genes derived from global expression profiling of [AKXDn × PyMT/Jf] primary tumors were identified agnostic to the metastatic phenotypes observed in the parental strains or RI panel sublines. This approach allowed for unbiased, genome-wide isolation of all biologically relevant co-expressed transcriptional modules. Each module was then simultaneously tested for association with metastasis and human clinical significance by deriving module gene signatures and applying them to independent expression-level patient data sets of node-negative nonadjuvant treated breast cancer with associated clinical endpoints. Our strategy, based on mRNA and miRNA expression in primary tumors, identified both metastasis-driving genes and miRNAs, further supporting a paradigm in which transcription programs within the primary tumor are deter-

Figure 6. Integration of systems genetics analysis. (A) Circos integration of global transcript and miRNA profiling analyses. From periphery inward: clockwise proximal to distal chromosomal positioning (blue), (1) mRNAs showing statistically significant eQTL peaks (red peaks), (2) mRNAs with expression significantly associated with metastatic propensity (blue peaks), (3) hub transcripts encompassed within eQTL intervals (black peaks), (4) miRNAs encompassed within miR-eQTL intervals (green peaks), (5) miRNAs with metastasis-associated expression (pink peaks), (7) miRNAs predicted to target highly connected network hubs (blue line links miRNA and its predicted target mRNA, red point indicates miRNA position). Genes labeled in purple indicate module central nodes. Genes labeled in green indicate hub transcripts with predicted miR-3470b MREs that are down-regulated upon Mir3470b overexpression. miRNAs shown to have a causal role in metastasis are indicated in blue. Peak lengths are in –log10 P-value. Please refer to Supplemental Figure 44 for a detailed high-resolution version. (B) Proposed model for murine metastatic progression. Global genome demethylation associated with neoplastic transformation results in the DNA demethylation and transcriptional activation of short interspersed elements (SINEs) including Mir3470b. Mir3470b expression down-regulates metastasis suppressive genes and the metastatic progression inhibitory Cnot2 network while concurrently up-regulating the metastasis promoting Brd4-SF, driving metastatic progression.
ominants of metastatic progression (van de Vijver et al. 2002; Ramaswamy et al. 2003). In addition to describing a novel approach to the identification of disease-specific metastasis susceptibility transcription modules, the networks themselves provide a resource for selecting and prioritizing genes and miRNAs for analysis, based on position, connectivity, and number of potential target genes.

The Ndh module, for example, incorporated four previously identified metastasis susceptibility genes (Ndh, Pti16, Arap3, Cs1fr). The Cnot2 module also contained at least four previously identified regulators of tumor progression: the metastasis susceptibility gene, Arid4b, its protein interaction partner, Sin3a, Luc7l, and the anti-proliferative cell cycle component Btg2 (Crawford et al. 2008b; Takahashi et al. 2011; Winter et al. 2012; Das et al. 2013). Our laboratory validated the direct role of both Arid4b and Luc7l using genetic and in vivo modeling of metastasis (Crawford et al. 2008b; Winter et al. 2012). Since expression modules were identified by unsupervised clustering of array data, the incidental membership of bona fide metastasis modifier genes within these networks provide another measure of confidence in our method for identifying potential transcription networks that may drive metastatic susceptibility. Based on the reproducibility of the prognostic power of the Cnot2 module we chose to investigate Cnot2 as both a potential metastasis modifier gene and a factor contributing to the establishment of the transcriptional network module.

Cnot2 is a core component of the CCR4–NOT complex, a multifunctional protein complex that is highly conserved in Eukarya. It has been reported to have transcriptional regulatory, deadenylase, ubiquitin ligase, and chromatin modifying activities (Collart and Panasenko 2012). It is additionally thought to play an essential role in miRNA-mediated transcript repression by destabilizing miRNA-bound target transcripts (Behm-Ansmant et al. 2006). CCR4–NOT complex subunit 2, Cnot2, was originally identified as a negative regulator of transcription in yeast (Collart and Struhl 1994). Although Cnot2 has not been reported to have any enzymatic activity, it has been shown to be necessary for the CCR4–NOT deadenylase activity in mammalian cells (Ito et al. 2011). Moreover, studies indicate the contribution of Cnot2 to cellular viability (Ito et al. 2011) as well as differentiation state (Zheng et al. 2012), both thought to play critical cellular functions in the metastatic cascade (Luzzi et al. 1998; Kourosh-Mehr et al. 2008). In yeast Cnot2 was observed to repress specific gene transcription in a histone deacetylase-dependent manner (Jayne et al. 2006) as well as to promote resumption of transcriptional elongation likely by binding directly to RNA polymerase II (Kruk et al. 2011). Intriguingly, Cnot2 has also been reported to physically interact with the metastasis-associated proteins BRD4 (which incidentally is a member of the Pol2q module) and TPX2 (Crawford et al. 2008a; Lau et al. 2009; Hu et al. 2012), raising the possibility that these proteins act in complex to modify tumor cell metastatic potential. All of these activities and associations are under current investigation as potential mechanisms that contribute to the regulation of the Cnot2 module and Cnot2-mediated metastasis suppression as observed in the present study. Nonetheless, the identification of Cnot2 as an inhibitor of tumor progression represents a significant finding as it links an entirely novel protein complex as a determinant of metastasis.

Further, we show by two distinct methods that miRNAs have functional regulatory roles in networks of heritable metastasis susceptibility: (1) Metastasis-suppressing miRNAs possess cis-eQTLs, indicative of heritable miRNA expression levels, and (2) miRNAs can target heritably derived metastasis-driving transcriptional networks. Two independent analyses identified several novel heritable metastatic susceptibility miRNA genes (Mir216a, Mir216b, Mir217, Mir3470a, Mir3470b). The discrepant effects on tumor progression of the Mir216/217 cluster between 6DT1 and 4T1 cells is under investigation but is likely attributable to differences in expressed genes between these cell lines. 6DT1 cells were derived from a MMTV–MYC transgenic tumor (Pei et al. 2004) with a gene expression profile resembling luminal-like human breast tumors (Lim et al. 2010) while 4T1 cells are derived from a spontaneous tumor in BALB/c mice with an expression profile resembling basal-like human breast tumors (Kaur et al. 2012). As such, 6DT1 and 4T1 cells contain major transcriptomic differences (JE Green, unpubl.) and therefore likely express different subsets of miR-216/217 target transcripts, which could account for differences observed in tumor progression phenotypes. Our results are consistent with the notion that these miRNAs regulate numerous genes that promote tumor progression and which may be expressed differentially in different metastatic tumors. Notably, mir-216b and mir-217 have previously been implicated as inhibitors of tumor growth and invasion in nasopharyngeal, renal clear cell, and pancreatic ductal carcinoma (Deng et al. 2011; Ali et al. 2012; Li et al. 2013b). Further, each member of the Mir216/217 cluster was demonstrated to possess a significant cis-acting eQTL, indicating that differences in Mir216/217 expression across the AKXD RI panel are inherited and not acquired somatically during neoplastic transformation or tumor progression.

In contrast mirR-3470a/b were selected to validate our combined small RNA sequencing and miRNA profiling approach. We surmised that the relative overrepresentation of mirR-3470a/b targets in our transcription networks would render these miRNAs the most powerful in testing the validity of our miRNA–mRNA integrated network approach. Although mirR-3470a/b were identified by target prediction analysis by combining all metastasis-associated network transcripts, their role in metastasis may be in part mediated by direct or indirect targeting of previously identified metastasis susceptibility genes (Brd4, Rpl1b, Cadm1), genes within the Cnot2 network, genes within all networks constructed here, or a combination of the above. Despite no conserved human homolog to these murine miRNAs, their effect on transcriptional networks conserved among mice and humans validates our approach of integrating network analysis and miR-seq data and provides another line of evidence for the potential functional role of our statistical correlation-based expression networks in tumor progression.

A closer look at the genomic context of Mir3470a/b reveals that both are degenerated B1-type short interspersed nuclear elements (SINEs), which are thought to be predominantly silenced in adult cell types largely by DNA methylation (Su et al. 2012). This is consistent with the initial report identifying mir3470a and mir3470b showing that its expression is restricted to the gonads (Ahn et al. 2010). There is also accumulating evidence supporting observations of global DNA hypomethylation in cancer (Feinberg and Vogelstein 1983), with the most comprehensive genome-wide, nucleotide-level data available for breast cancer (Hon et al. 2012), and that the resulting hypomethylation leads to expression of repetitive elements (Florl et al. 1999). Given these observations, we propose that murine tumorigenesis leads to hypomethylation and expression of repetitive elements including Mir3470a/b, which drive tumor progression and metastasis (Fig. 6b). It remains unclear whether hypomethylation is a somatic or heritable event, but the strength of the effect by which Mir3470a/b can drive tumor progression may depend on heritable expression levels of metastasis susceptibility networks described in the current study and/or previously described susceptibility genes.
The networks described here represent statistically associated sets of expressed genes and therefore are likely partial representations of the true transcriptional structure underlying metastatic disease. Like the prognostic gene signatures, the exact structure and membership of the metastasis-associated transcriptional network will likely change somewhat as additional samples and analyses are performed (Ein-Dor et al. 2006). Encouragingly, significant overlap of seven of the AKXD modules has been observed in TOM analysis of an independent mapping cross (P-values 0.001–6 × 10^{-12}) (K Hunter, Y Hu, J Zhang, unpubl.), indicating that at least some portion of the network will be stable. These results also suggest that a similar analysis performed on a much larger sample set, derived for example from the use of the new Collaborative Cross mouse genetic mapping resource (Churchill et al. 2004), will improve this higher-order analysis of the susceptibility mechanisms associated with metastatic breast cancer and identify novel and more robust targets for clinical intervention. Although significant work remains to understand the details uncovered in this study, the observations presented here not only support our strategy of identifying higher-order transcriptional structure but also highlight that such approaches are essential for acquiring a broad-scale understanding of tumor progression.

Methods

Gene expression and small RNA sequencing
Affymetrix MOE430 gene expression analysis was performed as previously described (Yang et al. 2005). Total RNA was isolated from two to four tumors from each of the [AKXD × PyMT]F1 outcrosses; due to availability constraints of AKXD RI panel animals, only 18 of the 24 sublines were crossed to MMTV-PyMT mice. Where possible the same tumors used for the MOE430 expression analysis were selected for small RNA sequencing. Total RNA was extracted from the mammary tissues using the mirVana miRNA Isolation Kit (Ambion). Each RNA sample was applied to the Nanodrop 1000 Spectrophotometer (Thermo Scientific) to calculate sample concentration and purity. All samples displayed a 260/280 ratio >2.0 with a small portion having a 260/230 ratio <2.0. Samples were further analyzed for integrity using the Agilent 2100 Bioanalyzer. Equal amounts of total RNA were pooled for each of the genotypes within the AKXD RI × PyMT panel then size selected for small RNA species before sequencing. The NCI Next Generation Sequencing Core performed enrichment of small RNA and sequencing on Illumina GA IIx.

Network analysis
We used an R package “weighted correlation network analysis (WGCNA)” (Langfelder and Horvath 2008) to find clusters (modules) of highly correlated genes on AKXD mRNA expression array data sets: GSE30864, GSE30865, GSE30866, GSE31223. The network for each module was generated with the minimum spanning tree with a dissimilarity matrix from WGCNA. Hub transcripts were determined by links (degree of freedom) ≥5. This study utilized the high-performance computational capabilities of the Bio-wulf Linux cluster at the National Institutes of Health, Bethesda, MD (http://biowulf.nih.gov).

eQTL analysis
eQTL analysis was performed with R package qtl. Expression levels of all significantly expressed miRNAs were compared with the genetic map using an R/QTL package (Broman et al. 2003) and genome-wide significance levels determined by a bootstrap permutation test (1000 permutations) using both the permutation tail probability test and the one-sided P-value test. The permutation tests were based on resampling with samples drawn without replacement from the pooled sample set. The observed value T(obs) of the test statistic was calculated and the observations/samples were pooled. Next the set of the statistic values were calculated by resampling each sample. For example, for 10,000 permutations, 9999 statistic values were calculated. Finally, the one-sided P-value of the test was then calculated as the proportion of sampled permutations where the difference in means was greater than or equal to T(obs). The genome-wide statistical threshold was set at 0.05.

MiRNA–phenotype correlation analysis was conducted using the R package glm.

Correlation of microRNA abundance with gene expression
Correlation analysis between the expression of microRNA and targeted genes was performed for all pairs of genes and miRNAs, then filtered by the genes predicted as the putative targets of a particular microRNA. Targets were predicted from three of four databases: PITA (Witkos et al. 2011), miRBase (Griffiths-Jones et al. 2006), TargetScan, and miRanda (Miranda et al. 2006). The association of the expression between miRNAs and their target genes was analyzed by R package glm. The significant associations were determined by a bootstrap permutation test (10,000 times and P-value < 0.01). For genes whose expression significantly correlated with microRNA expression, a gene set enrichment analysis (GSEA) was performed by a hypergeometric test in R.

Animal studies
Female FVB/NJ or Balb/cJ mice from Jackson Laboratories were injected at 6-8 wk of age. Two days prior to orthotopic injections, cells were placed in nonselective media. On the day of injection, 1 × 10^5 cells were injected orthotopically into the fourth mammary fat pad of age-matched, syngeneic, immune competent virgin females. After 30 d the mice were euthanized by intraperitoneal injection of 1 ml Tribromoethanol with subsequent cervical dislocation. Primary tumors were resected, weighed, and snap-frozen in liquid nitrogen. Lungs were resected, surface metastases were counted; lungs were inflated with 10% nitrate-buffered formalin and sent for sectioning and staining. All procedures were performed under the Animal Safety Proposal (LCBG-004) and approved by the NCI-Bethesda Animal Care and Use Committee.

Data access
The AKXD small RNA sequencing data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE50179.

Acknowledgments
We thank Jude Alsarraj, Thomas R. Geiger, Jeffrey E. Green, Ngoc-Han Ha, and James J. Morrow for critical review of this manuscript and stimulating discussion. We thank Dominic Esposito of the Protein Expression Laboratory, SAIC-Frederick, for constructing the myc-Cnot2 expression vector. This research was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research and the Howard Hughes Medical Institute Research Scholars Program.

Author contributions: All experiments were conceived by F.F. and K.W.H. Experiments were conducted by F.F. and N.E.G. with
technical assistance from R.C.W. Computational analysis was performed by Y.H., G.W., and J.Z. The manuscript was written by E.F. and K.W.H.

References

Ahn HW, Morin RD, Zhao H, Harris RA, Coarfa C, Chen ZJ, Milosavljevic A, Warra MA, Rajkovic A. 2010. MicroRNA transcriptome in the newborn mouse ovaries determined by massive parallel sequencing. Mol Hum Reprod 16: 463–471.

Ali S, Banerjee S, Logina F, Bao B, Philip PA, Korc M, Sarkar FH. 2012. Inactivation of Irk4/Arf leads to deregulated expression of miRNAs in K-Ras transgenic mouse model of pancreatic cancer. J Cell Physiol 227: 3373–3380.

Alsarraj J, Walker RC, Webster JD, Geiger TR, Crawford NP, Simpson RM, Ozato K, Hunter KW. 2011. Depletion of the proline-rich region of the murine metastasis susceptibility gene Brd4 promotes epithelial-to-mesenchymal transition- and stem cell-like conversion. Cancer Res 71: 3121–3131.

Aslakson CJ, Miller FR. 1992. Selective events in the metastatic process defined by analysis of the sequential dissemination of subpopulations of a mouse mammary tumor. Cancer Res 52: 1399–1405.

Bad H, Harati S, Mehra S, Kauhal D, Moroz P, Dash S. 2012. Increased expression of Pglycoprotein and doxorubicin chemoresistance of metastatic breast cancer is regulated by miR-298. Am J Pathol 180: 2490–2503.

Behm-Ansmant I, Rehwinkel J, Doerks T, Stark A, Bork P, Izaurelde E. 2006. miRNA degradation by miRNAs and GW182 requires both CCR4 NOT deadenylase and DCP1/DCP2 decapping complexes. Genes Dev 20: 1885–1898.

Beckmann J, Yee K, Chang YF, Prat A, Hua D, Nwachukwu C, Dalton R, Huang S, Swanson KE, Penou CM, et al. 2013. MicroRNA-30c targets cytoskeleton genes involved in breast cancer cell invasion. Breast Cancer Res Treat 137: 373–382.

Broman KW, Wu H, Sen S, Churchill GA. 2003. R/qtl: QTL mapping in experimental crosses. Bioinformatics 19: 889–890.

Churchill GA, Arvey DC, Alleyaye H, Angel JM, Attie AD, Beatty J, Beavis WD, Belknap JK, Bennett B, Berrettini W, et al. 2004. The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat Genet 36: 1133–1137.

Collart MA, Panasenko OO. 2012. The Ccr4–Not complex. Gene 492: 42–53.

Collart MA, Struhl K. 1994. NOT(CDC39, NOT2)(CDC36, NOT3, and NOT4) encode a global-negative regulator of transcription that differentially affects TATA-element utilization. Genes Dev 8: 525–537.

Crawford NP, Ziegas A, Peil DJ, Hess J, Anton-Culver H, Hunter KW. 2006. Germline polymorphisms in SIPA1 are associated with metastasis and other indicators of poor prognosis in breast cancer. Breast Cancer Res 8: R16.

Crawford NP, Qian X, Ziegas A, Papageorge AG, Boersma BJ, Walker RC, Lukes L, Rowe WL, Zhang J, Ambs S, et al. 2007. Rpl1b, a new candidate susceptibility gene for breast cancer progression and metastasis. PLoS Genet 3: e214.

Doss S, Schadt EE, Drake TA, Lusis AJ. 2005. Integrated cross-species transcriptional network analysis of metastatic susceptibility. Proc Natl Acad Sci USA 102: 1333–1338.

Ein-Dor L, Zuk O, Domany E. 2006. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. Cancer Res 66: 2383–2392.

Feltz JE, Sawa Y, Koyama K, Park JY, Kimura T, Aoki S, et al. 2012. miR21 and miR17-92 cluster targets CCR4-NOT complex directly promotes transcription elongation. Cytoskeleton 70: 447–458.

Finak G, Berto N, Pepin E, Sadekova S, Souleimanova M, Zhao H, Chen H, Omerruqo G, Morozov P, Brown M, Hafner M, Reyal F, van Koonhoven M, et al. 2011. MicroRNA sequence and expression analysis in breast tumors by deep sequencing. Cancer Res 71: 4443–4453.

Feinberg AP, Vogelstein B. 1983. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature 301: 89–92.

Finak G, Berto N, Pepin E, Sadekova S, Souleimanova M, Zhao H, Chen H, Omerruqo G, Morozov P, Brown M, Hafner M, Reyal F, van Koonhoven M, et al. 2011. MicroRNA sequence and expression analysis in breast tumors by deep sequencing. Cancer Res 71: 4443–4453.

Gagnon J, Duggan D, Tarnawski A, Schilling B. 2002. A novel mouse model for metastatic disease. Mol Cell Biol 22: 357–369.

Ghia CL, Phinney DG. 2012. MicroRNAs in the imprinted DLK1-DIO3 region repress the epithelial-to-mesenchymal transition by targeting the TWIST1 protein signaling network. J Biol Chem 287: 42695–42707.

Goldberger N, Walker RC, Green JE, Field LM, Hunter KW. 2013. Deleted in breast cancer susceptibility. Cancer Res 73: 2671–2681.

Gregory PA, Bert AG, Paterson EL, Barry SC, Tsykin A, Farshid G, Vadas MA, Mielke J, Goldberger N, Walker RC, Kim CH, Winter S, Hunter KW. 2013. Inherited metastasis susceptibility gene Arid4b. Cancer Res 73: 2546–2560.

Hui AB, Lin A, Xu W, Waldron L, Perez-Ordonez B, Weirenb I, Shi W, Bruce J, Huang SH, O’Sullivan B, et al. 2013. Potentially prognostic miRNAs in HPV-associated oropharyngeal carcinoma. Clin Cancer Res 19: 2154–2162.

Hunter KW, Brown KW, Voyer TL, Lukes L, Cozza D, Debbies MT, Rouse J, Welch DR. 2001. Predisposition to efficient mammary tumor metastatic progression is linked to the breast cancer metastasis suppressor gene Bcm1. Cancer Res 61: 6866–6872.

Hurst DR, Edmonds MD, Welch DR. 2009. Metastasir: The field of metastasis-regulatory microRNA is spreading. Cancer Res 69: 7495–7498.

Ito K, Inoue T, Yokoyama K, Morita M, Suzuki T, Yamamoto T. 2011. CNOT2 deletion depletes and inhibits the CCR4-NOT deadenylase complex and induces apoptotic cell death. Genes Cells 16: 597–607.

Jahidi S, Sun J, Edwards RA, Dizon D, Panarelli NC, Milsom JW, Sikandar SS, Gusmus ZH, Lipkin SM. 2012. 19-miR-23a promotes the transition from invasive adenocarcinoma to invasive ductal carcinoma in breast cancer. BMC Cancer 12: 490.

Jaye S, Swartjes CG, van Schak HM, Timmers HJ. 2006. Involvement of the DLK1-DIO3 complex in breast cancer is regulated by miR-298. J Biol Chem 281: 72287–72297.

Kaur P, Nagaraj GA, Zheng H, Gizauch D, Galuakande M, Krishnan S, Asse A. 2012. A mouse model for triple-negative breast cancer tumor-initiating cells (TRCN-TCIs) exhibits similar aggressive phenotype to the human disease. BMC Cancer 12: 120.

Kourois-Mehr H, Bechis SK, Siorach EM, Littlarge LE, Ebgald E, Ewald AJ, Pai J, Hou IC, Web Z, Vo C. 2008. GATA-3 links tumor differentiation and dissemination in a luminal breast cancer model. Cancer Cell 13: 141–152.

Kruk JA, Datta A, Fu J, Gilmour DS, Reese JC. 2011. The multifunctional Ccr4-Not complex directly promotes transcription elongation. Genes Dev 25: 581–593.

Kuo YZ, Tai YH, Lo HI, Chen YL, Cheng HC, Fang WY, Lin SH, Yang CL, Tsai ST, Wu LW. 2013. miR-99a exerts anti-metastasis through inhibiting myotubularin-related protein 3 expression in oral cancer. Oral Dis 19: 1111/odi.12133.

Lancaster M, Rouse J, Hunter KW. 2005. Modifiers of mammary tumor progression and metastasis on mouse chromosomes 7, 9, and 17. Mammm Genome 16: 120–126.

Langfelder P, Horvath S. 2008. WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics 9: 559.
Lau NC, Kolkmann A, van Schaik FM, Mulder KW, Pijnappel WV, Heck AJ, Timmers HT. 2009. Human Ccr4-Not complexes contain variable deadenylase subunits. *Biochem J* **422**: 443–453.

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 A, Horvath S. 2009. Network module detection: Affinity search technique with the multi-node topological overlap map. *BMC Res Notes* **2**: 142.

Li Y, Zhang M, Chen H, Dong Z, Ganapathy V, Thangaraju M, Huang S. 2010. Ratio of miR-196 to HOXC8 as a serum microRNA correlates with breast cancer cell migration and metastasis. *Cancer Res* **70**: 7894–7904.

Li Y, Yin C, Zhang B, Sun Y, Shi L, Liu N, Liang S, Liu S, Liu Y, Zhang J, et al. 2013a. PTTG1 promotes migration and invasion of human non-small cell lung cancer cells and is modulated by miR-186. *Carcinogenesis* **34**: 2145–2155.

Li H, Zhao J, Zhang JW, Huang QY, Huang JZ, Chi LS, Tang HJ, Liu GQ, Zhu DJ, Ma WM. 2013b. MicroRNA-217, down-regulated in clear cell renal cell carcinoma and associated with lower survival, suppresses cell proliferation and migration. *Neoplasia* **60**: 511–515.

Li J, Song Y, Wang Y, Luo J, Yu W. 2013c. MicroRNA-148a suppresses epithelial-to-mesenchymal transition by targeting ROCK1 in non-small cell lung cancer cells. *Mol Cell Biochem* **380**: 277–282.

Li X, Liu X, Xia W, Zhou P, Gao P, Jiang S, Lobie PE, Zhu T. 2013d. c-MYC-regulated miR-23a/24-2/27a cluster promotes mammary carcinoma cell invasion and hepatic metastasis by targeting Sprouty2. *J Biol Chem* **288**: 18121–18129.

Liang S, He H, Zhao X, Miao Y, Gu Y, Guo C, Xue Z, Dou W, Hu F, Wu K, et al. 2011. MicroRNA let-7f inhibits tumor invasion and metastasis by targeting MYH9 in human gastric cancer. *PLoS ONE* **6**: e18409.

Lifshitz T, Le Voyer T, Williams M, Muller W, Klein-Szanto A, Buettow K, Hunter KF. 2010. Identification of in vivo mouse strains harboring genetic modifiers of mammary tumor age of onset and metastatic progression. *Int J Cancer* **77**: 640–644.

Linn E, Wu D, Pal B, Bouras T, Asselin-Labat ML, Vaillant F, Yagita H, Lindeman GJ, Smyth GK, Visvader JE. 2010. Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways. *Breast Cancer Res* **12**: R21.

Liu CJ, Shen WG, Peng SY, Cheng HW, Kao SY, Lin SC, Chang KW. 2013. miR-134 induces oncogenicity and metastasis in head and neck carcinoma through targeting WWOX gene. *Int J Cancer* **134**: 811–821.

Lukes L, Crawford NF, Walker R, Hunter KW. 2009. The origins of breast cancer prognostic gene expression profiles. *Cancer Res* **69**: 310–318.

Luzzi KJ, MacDonald IC, Schmidt EE, Kerkvliet N, Morris VL, Chambers AF, Groom AC. 1998. Multistep nature of metastatic inefficiency: Dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. *Am J Pathol* **153**: 865–873.

Miranda KC, Grynpas MD, Visvader JE. 2010. RECP regulates microRNA binding sites and their corresponding heteroduplexes. *Cell Rep* **1**: 1205–1217.

Monsurro ML, Taylor BA, Jenkins NA, Copeland NG. 1986. AKD recombinant inbred strains: Models for studying the molecular genetic basis of murine lymphomas. *Mol Cell Biol* **6**: 4236–4243.

Park YG, Zeng X, Luscher T, Eisen H, Ross KN, Lander ES, Golub TR, Rabinowitz JS. 2002. An integrated molecular description of human glioblastoma multiforme. *Cell* **114**: 2143–2156.

Pei XF, Noble MS, Davoli MA, Rosfjord E, Tilli MT, Furth PA, Russell R, Johnson MD, Dickson RB. 2004. Explant-cell culture of primary mammary gland tumors from MMTV-c-Myc transgenic mice. *Biol Anim Cells* **30**: 347–353.

Pillutla RC, Yue Z, Maldonado E, Shatkin AJ. 1998. Recombinant human Sipa1 gene Arid4b promotes mammary carcinoma cell growth and associated with lower survival, suppresses cell proliferation and migration. *Neoplasia* **60**: 511–515.

Sipa1