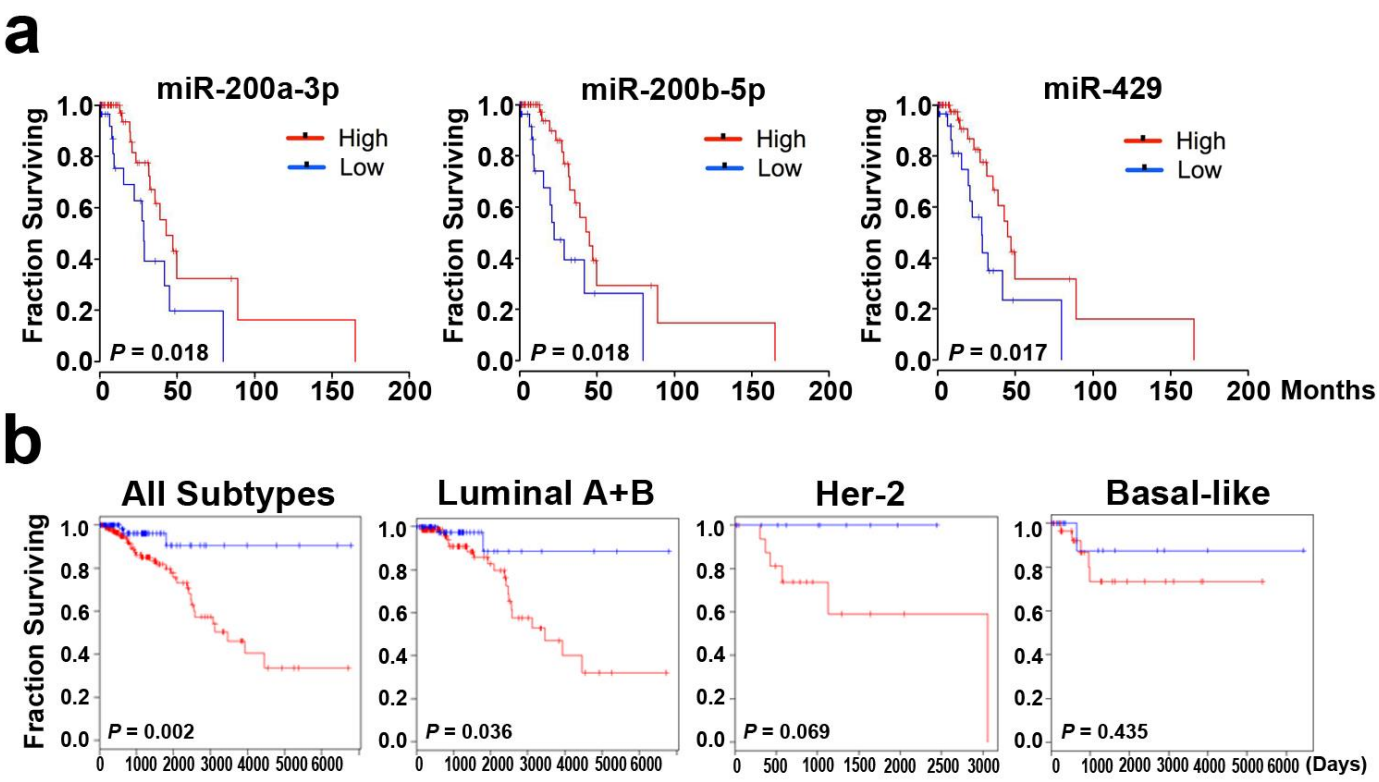
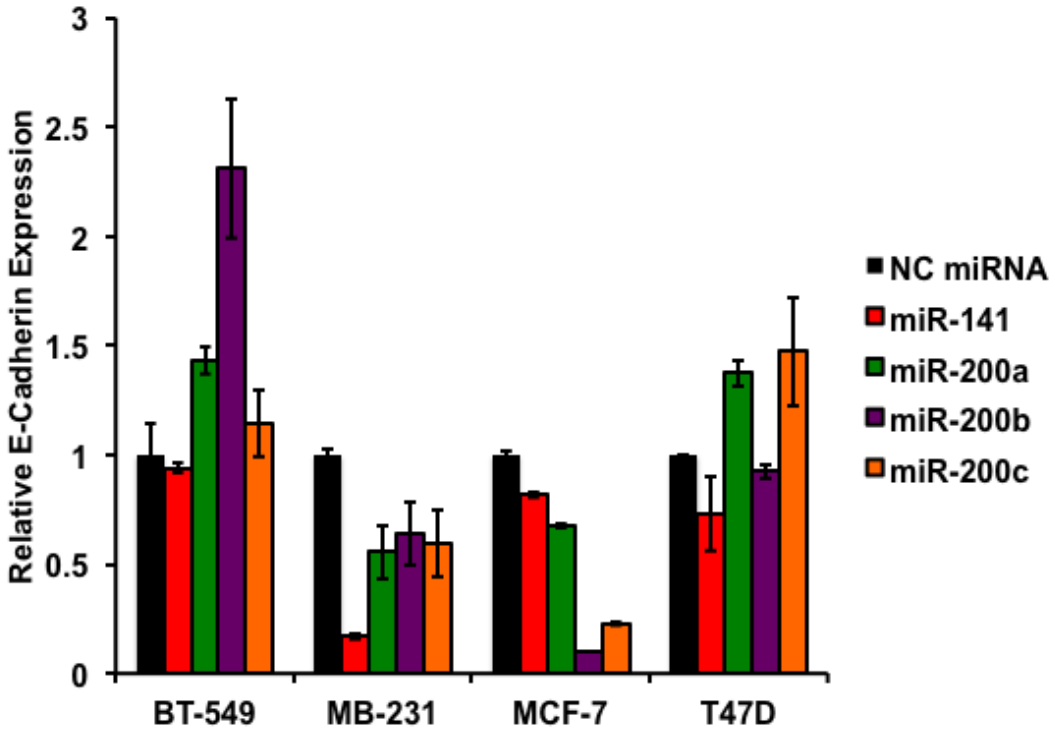
Supplementary Information

Tumor Angiogenesis Regulation by the miR-200 Family

Chad V. Pecot, Rajesha Rupaimoole, Da Yang, Rehan Akbani, Cristina Ivan, Chunhua Lu, Sherry Wu, Hee-Dong Han, Maitri Y. Shah, Cristian Rodriguez-Aguayo, Justin Bottsford-Miller, Yuexin Liu, Sang Bae Kim, Anna Unruh, Vianey Gonzalez-Villasana, Li Huang, Behrouz Zand, Myrthala Moreno-Smith, Lingegowda S. Mangala, Morgan Taylor, Heather J. Dalton, Vasudha Sehgal, Yunfei Wen, Yu Kang, Keith A. Baggerly, Ju-Seog Lee, Prahlad T. Ram, Murali K. Ravoori, Vikas Kundra, Xinna Zhang, Rouba Ali-Fehmi, Ana-Maria Gonzalez-Angulo, Pierre P. Massion, George A. Calin, Gabriel Lopez-Berestein, Wei Zhang, Anil K. Sood



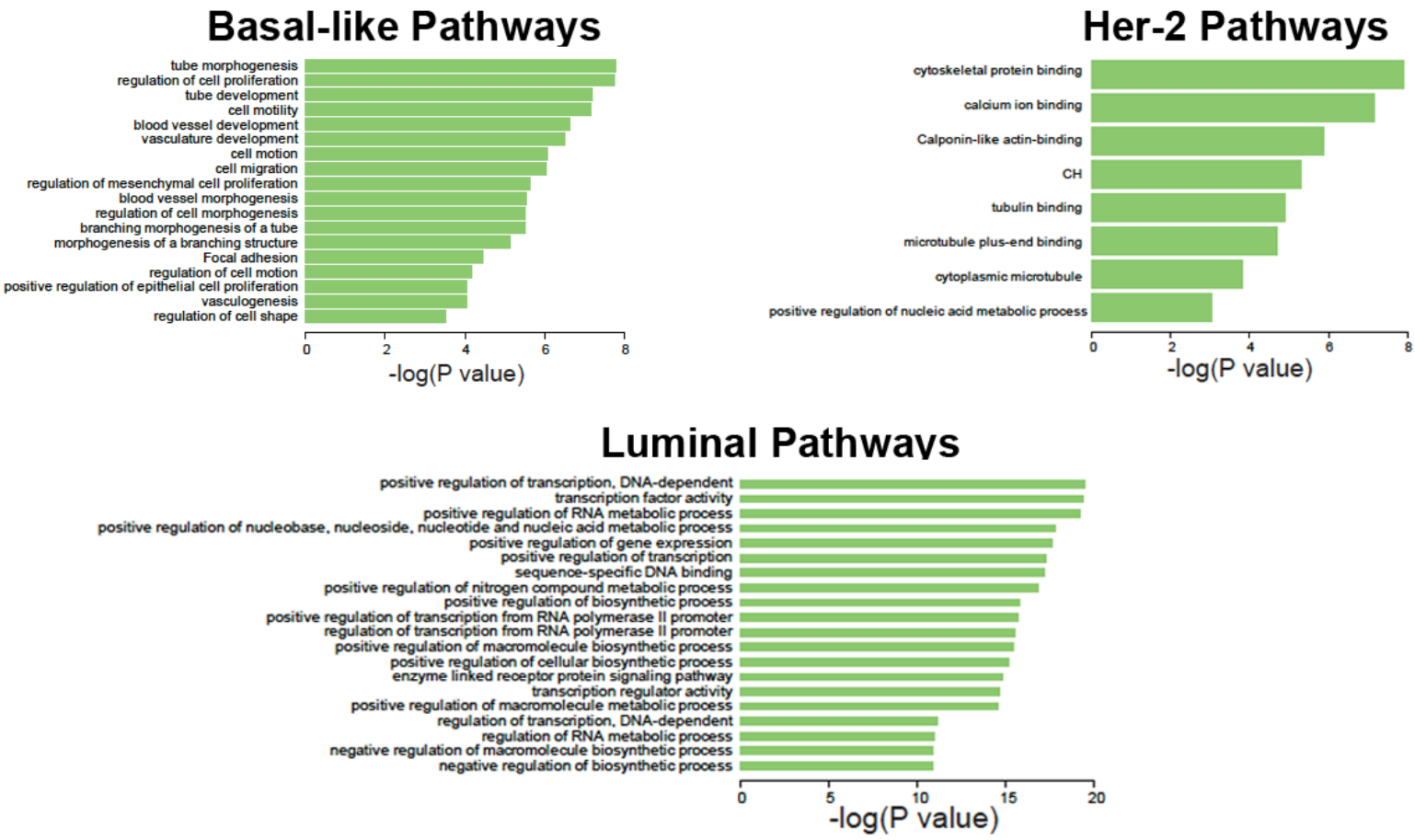
Supplementary Figure S1. Kaplan-Meier curves for overall survival in lung and breast adenocarcinoma patients in The Cancer Genome Atlas (TCGA). Several miR-200 isoforms in lung adenocarcinoma (a) validated as being clinical significant using the threshold established from the training cohort. In breast cancers (b), using the threshold established from the training cohort for miR-141, all subtypes together and the relative contribution by breast cancer subtype are shown. Of note, luminal A and luminal B samples were considered together. *P*-values obtained using log-rank test (Blue: Low expression, Red: High expression).



Supplementary Figure S2. Expression levels of E-cadherin. Relative expression levels of E-Cadherin 72 hours following transfection of basal-like (BT-549 and MB-231) and luminal (MCF-7 and T47D) breast cancer cell lines with different members of the miR-200 family. Data are averages ± SEM, n=3.

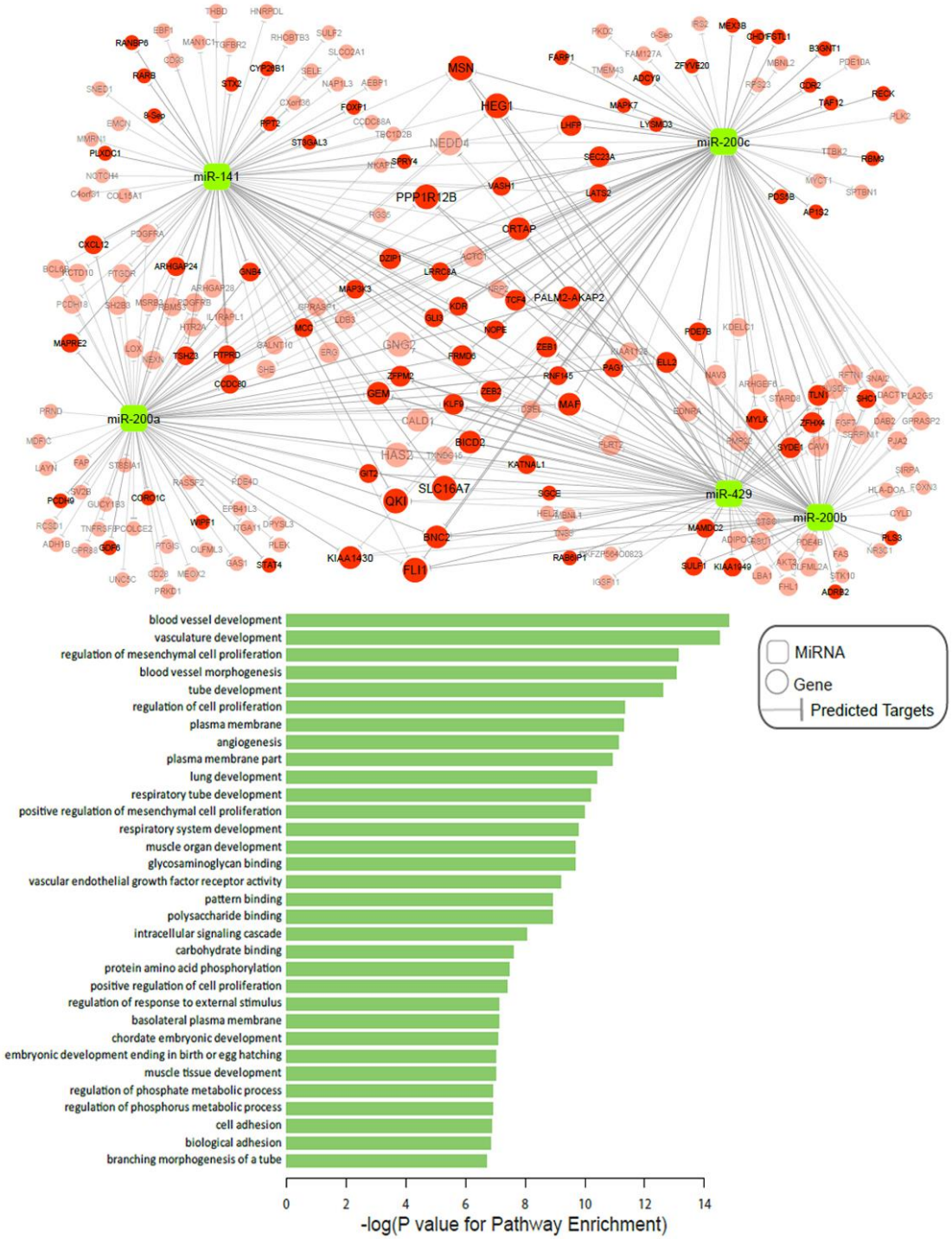
Supplementary Figure 3

a



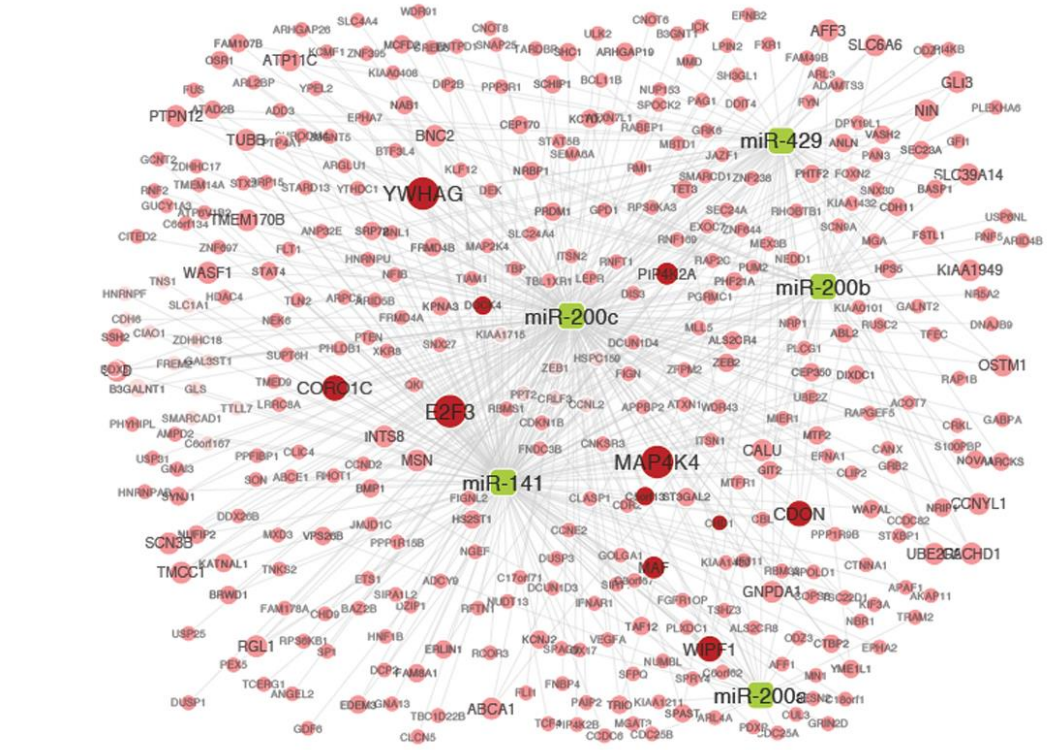
Supplementary Figure 3

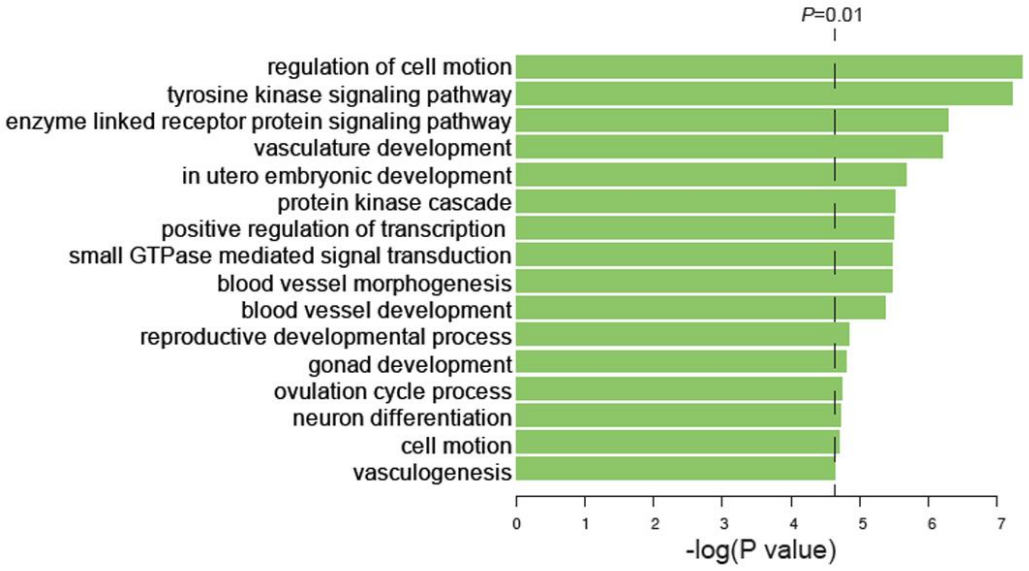
b



Supplementary Figure 3

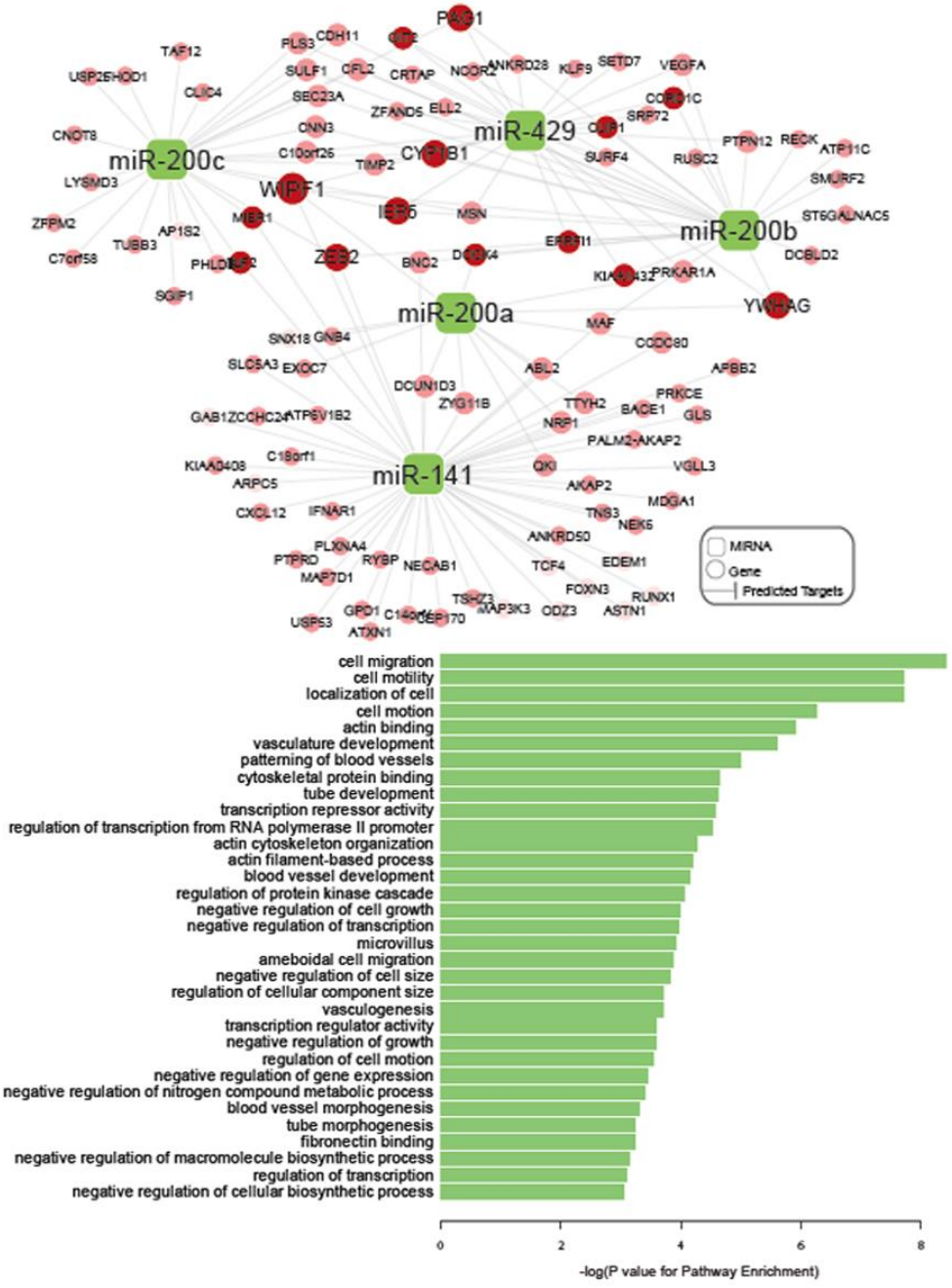
C



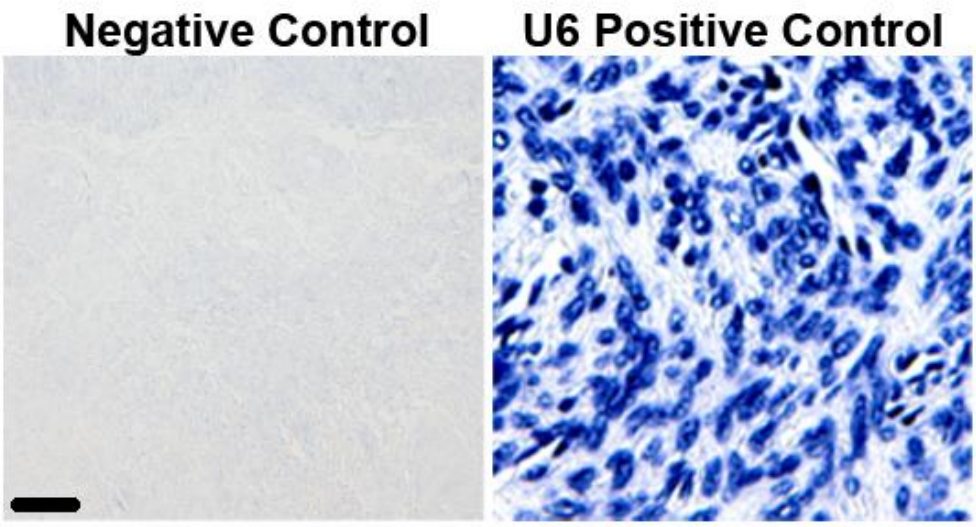


Supplementary Figure 3

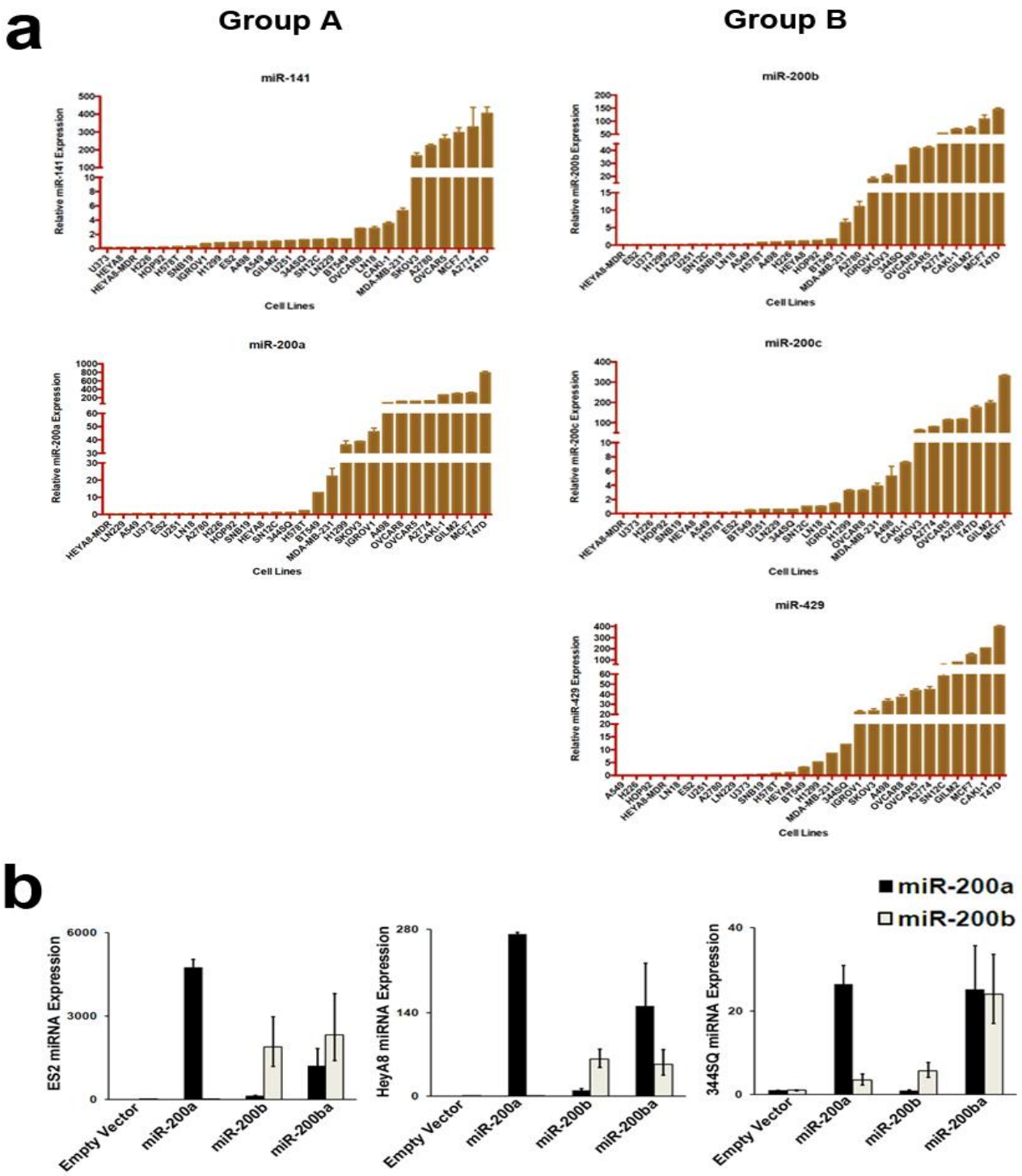
d



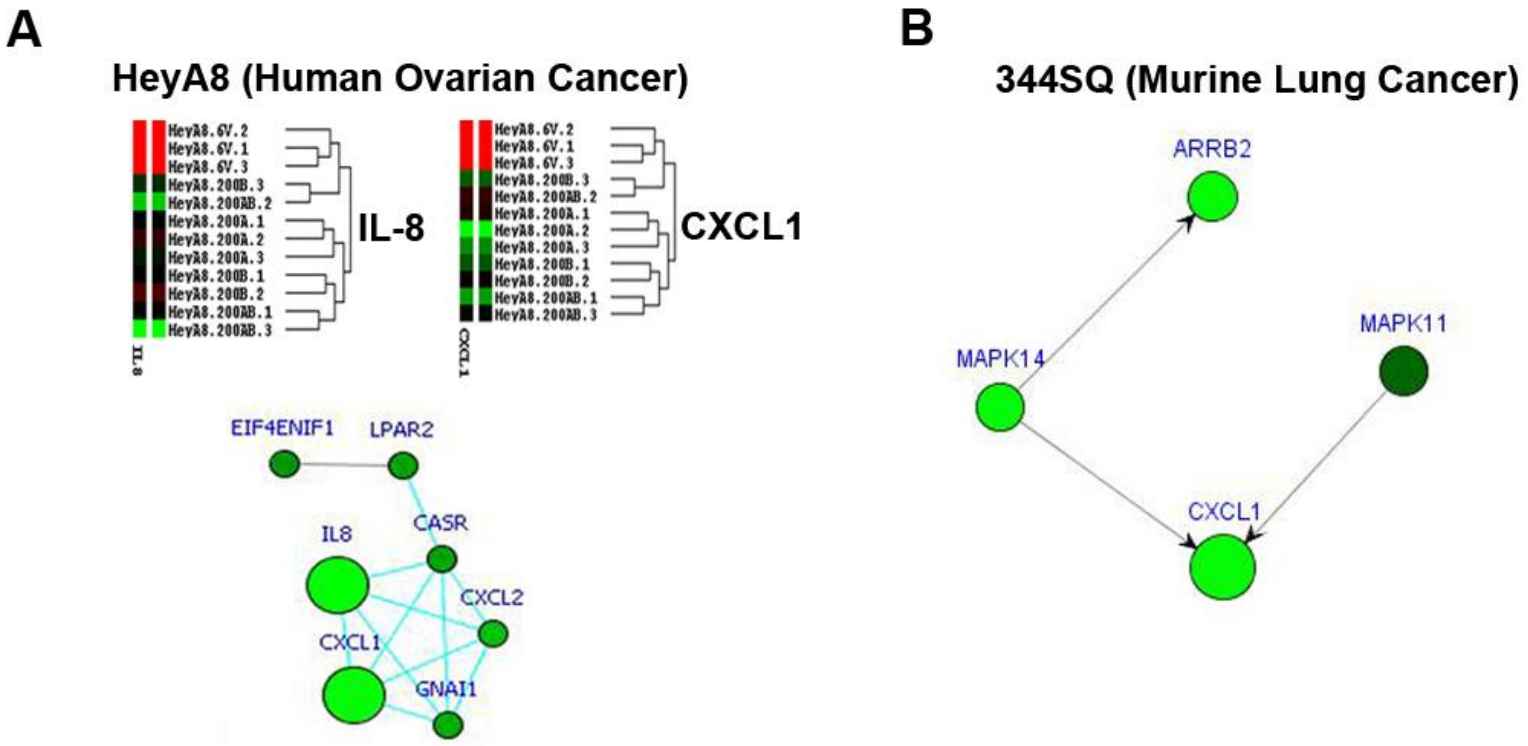
Supplementary Figure S3. An integrated network analysis for the miR-200 family in breast, ovarian, renal and lung adenocarcinomas in The Cancer Genome Atlas. Each breast cancer subtype (**a**) (basal-like, Her-2 and luminal) revealed genes that are significantly (P<0.001) inversely correlated with miR-200 members and also putative targets predicted from TargetScan software. Shown are the most significantly enriched signaling pathways identified. For ovarian cancer (**b**), (top) genes found to be significantly (P<0.01) inversely correlated with miR-200 members and also putative targets predicted from TargetScan software. The size of a gene node indicates the number of predicted miR-200 members; colors indicate whether the microRNA-gene regulation is conserved (dark red shading) or poorly conserved (light red shading). The most significantly enriched signaling pathways identified (bottom) on the basis of the above classified genes. For renal cancer (**c**), (top) the miR-200 family regulatory network in renal cancer shows genes that are significantly (P<0.001) inversely correlated with miR-200 members and also putative targets predicted from TargetScan software. The most significantly enriched signaling pathways identified (bottom) on the basis of the above classified genes where the black dashed line indicates the statistical significance of $P = 0.01$ or less. Note: Log here represents the natural logarithm. For lung cancer (**d**), (top) the miR-200 family regulatory network shows genes that are significantly (P<0.001) inversely correlated with miR-200 members and also putative targets predicted from TargetScan software. The most significantly enriched signaling pathways identified (bottom) on the basis of the above classified genes. *P*-values were calculated using FDR<0.01, Benjamini-Hochberg multiple testing correction.



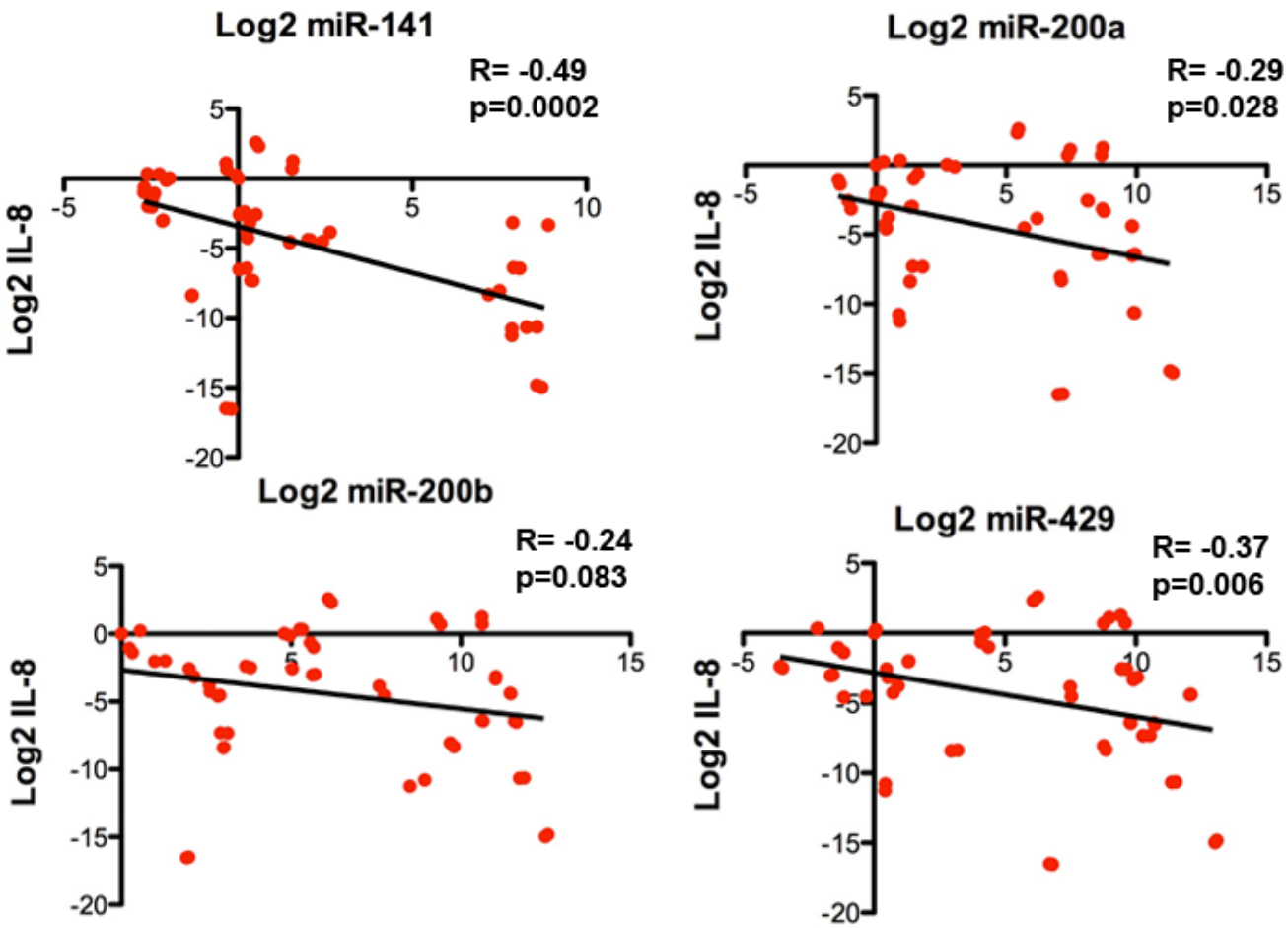
Supplementary Figure S4. *In situ* hybridization. Representative images of ovarian cancer following *in situ hybridization* using a negative (left) or U6 positive (right) control probe. Scale bars 250 μm.



Supplementary Figure S5. Relative expression levels of miR-200 members in a panel of cell lines. A 27-cell line panel (**a**) of ovarian, lung, renal and breast cancers showing relative expression levels of miR-200 members using quantitative PCR. Data are averages ± SEM, n=3. Note: Group A contains miR-141 and miR-200a, Group B contains miR-200b, miR-200c and miR-429. (**b**) Relative expression levels (compared to empty vector) of miR-200a and miR-200b for cell lines stably transfected with lenti-viral clones for empty vector, miR-200a, miR-200b or miR-200ba. Data are averages ± SEM, n=3.

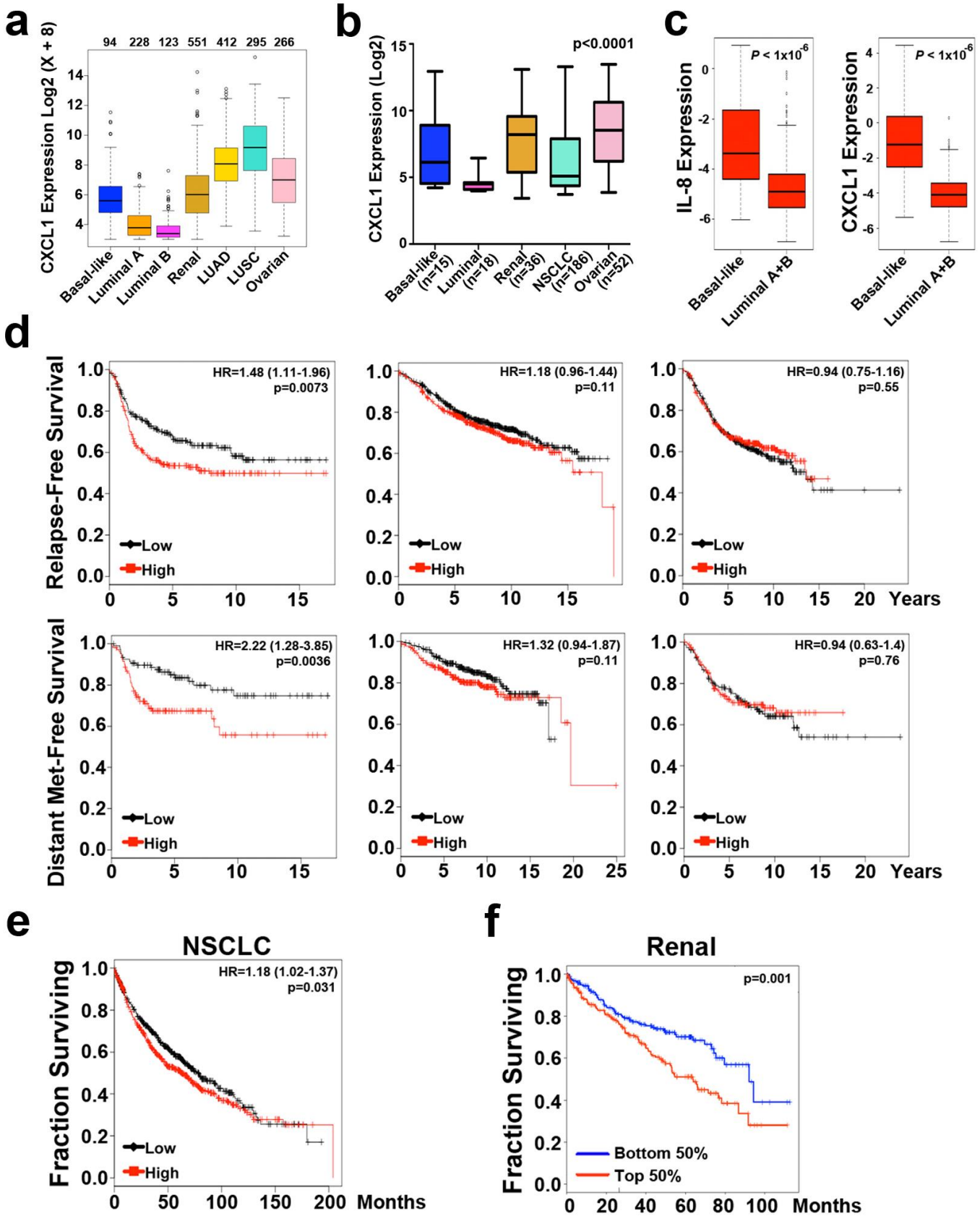


Supplementary Figure S6. Gene expression analysis of HeyA8 clones. Using an Illumina microarray to compare expression levels between (a) HeyA8 clones [Empty Vector (EV), miR-200a, miR-200b, miR-200ba] a net-walk analysis of angiogenesis regulators suggested miR-200 regulates pro-angiogenic cytokines IL-8 and CXCL1. For arrays and net-walk figures, red represents increased expression while green represents decreased expression. For 344SQ (b), a net-walk analysis for miR-200a, miR-200b and miR-200ba clones each demonstrated miR-200 regulation of CXCL1.

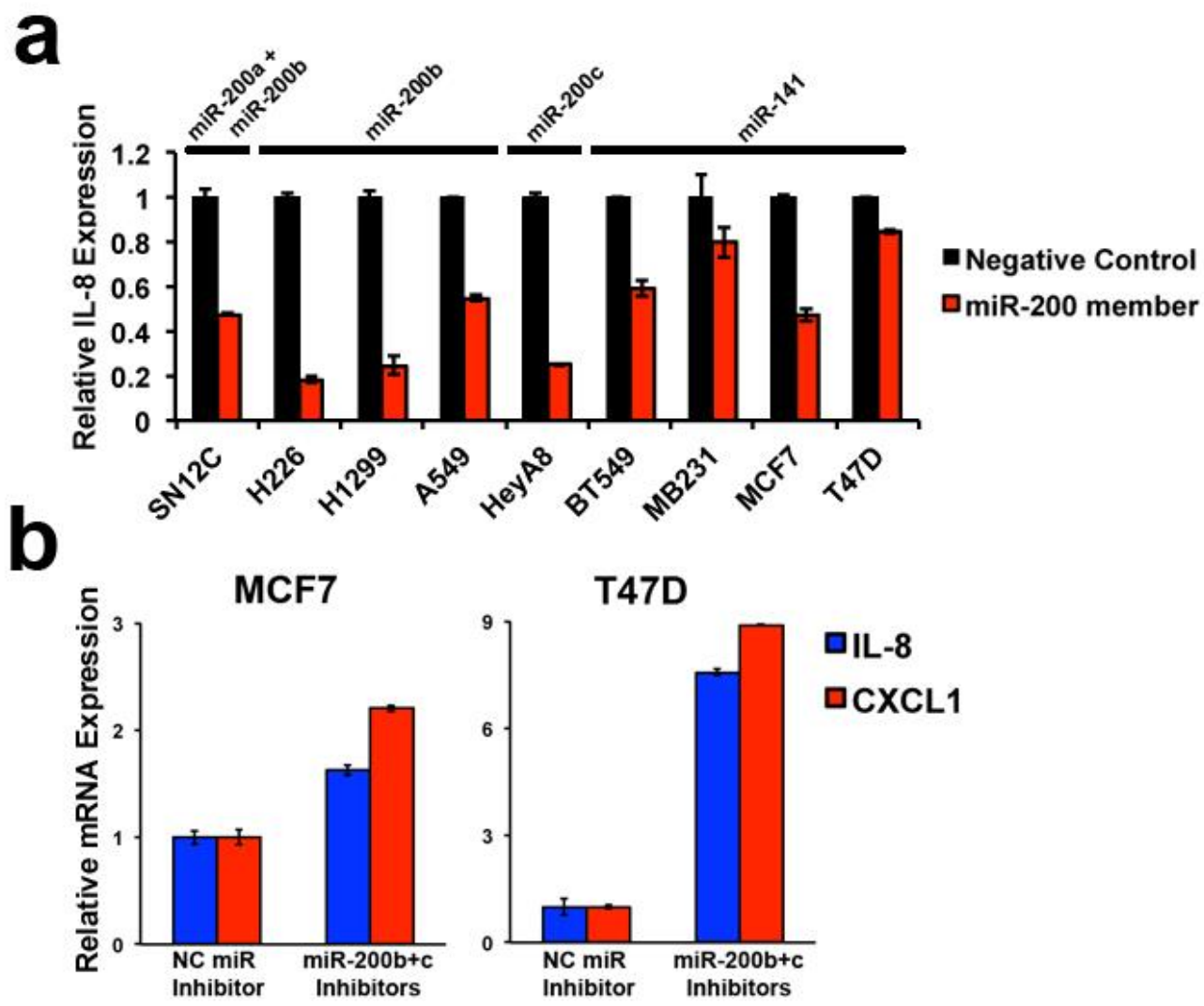


Supplementary Figure S7. Spearman correlation of miR-200 members and IL-8 expression levels of a 27 cell line panel. Quantitative PCR data was obtained for each cell line (in triplicate) for IL-8 and all 5 miR-200 members (shown: miR-141, miR-200a, miR-200b and miR-429). Spearman tests were performed to assess for a negative correlation.

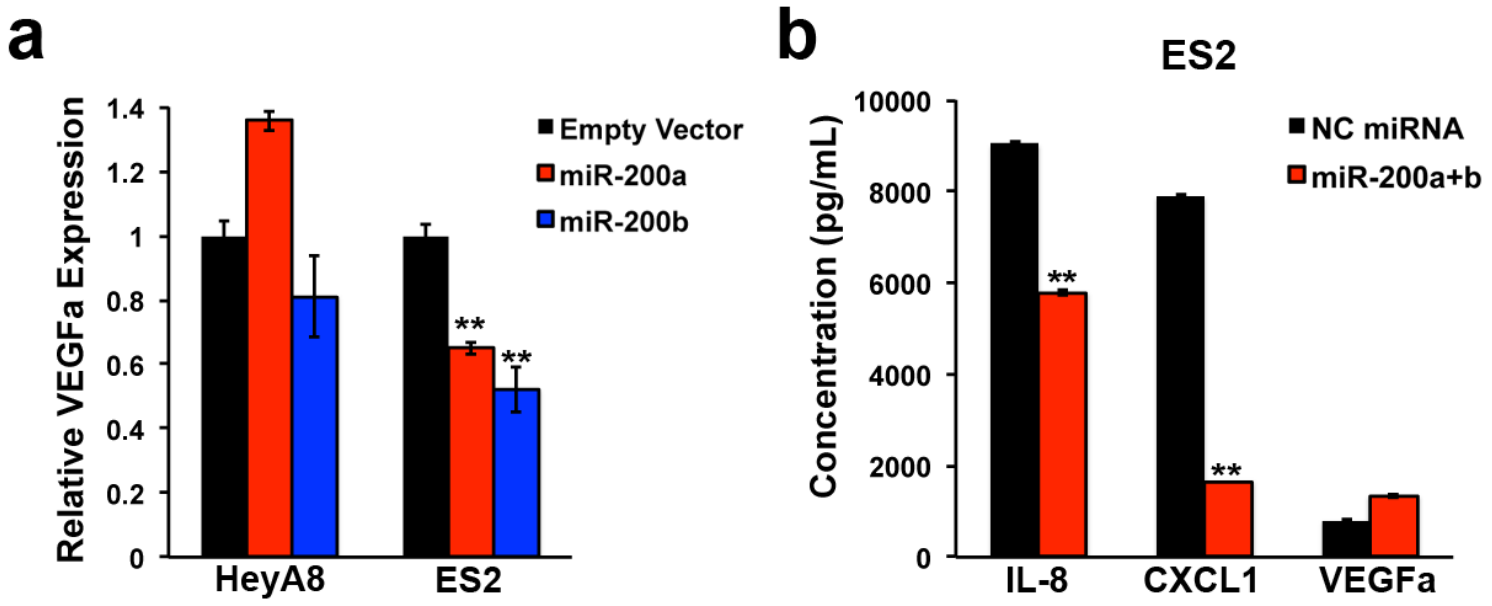
Supplementary Figure S8



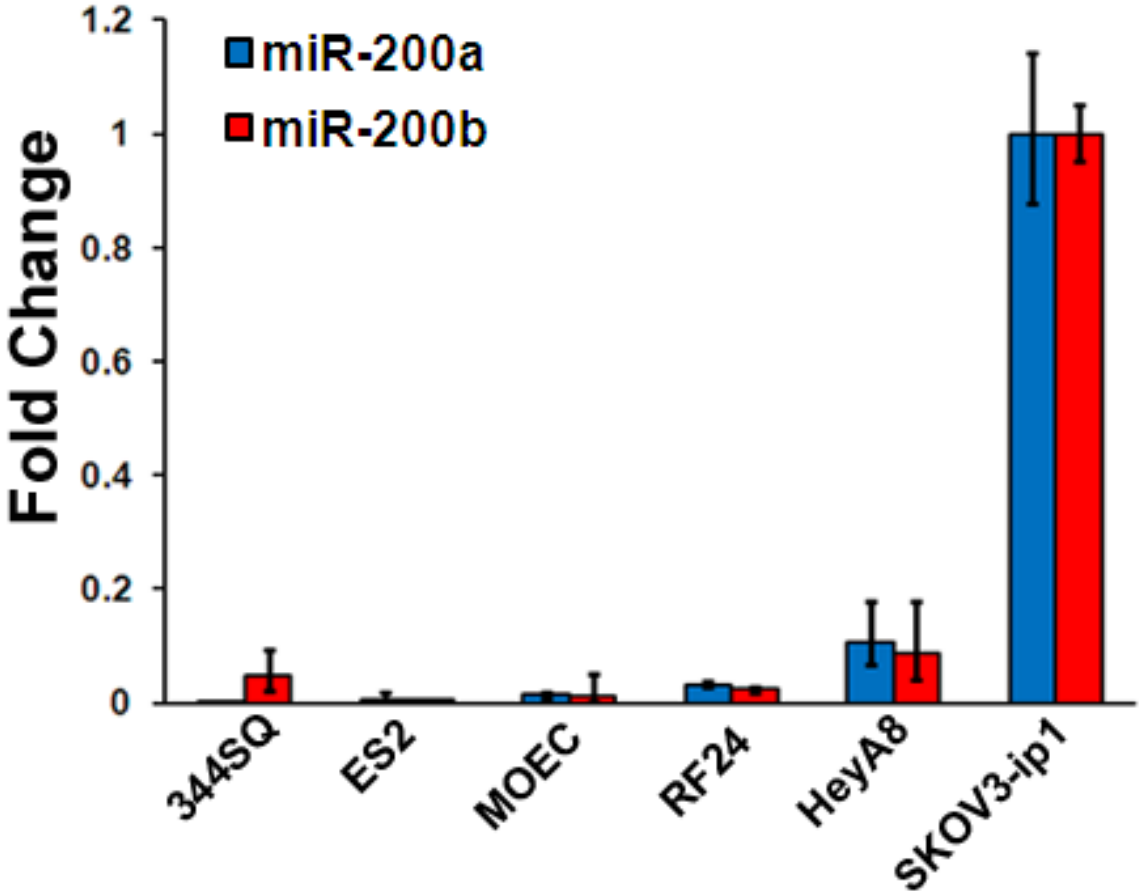
Supplementary Figure S8. Expression levels of CXCL1 in tumor and cell lines from several cancer types. (**a**) Relative expression levels of CXCL1 using RNA-Seq data from the Cancer Genome Atlas (TCGA) datasets. The numbers above each tumor type (top of panel) represents the sample size used to calculate expression levels. (LUAD: lung adenocarcinoma, LUSC: lung squamous carcinoma). Box plot represents first (lower bound) and third (upper bound) quartiles, whiskers represent 1.5 times the interquartile range. (**b**) Relative expression levels of CXCL1 from the cancer cell line encyclopedia (CCLE) dataset partitioned by designated tumor type. ANOVA test was used to test for statistical significance. (NSCLC: non-small cell lung cancer). (**c**) Relative expression levels (mRNA) of IL-8 and CXCL1 cytokines using Agilent microarray data from patients with Basal-like (54 patients) and Luminal A+B (232 patients) breast cancers. Box plot represents first (lower bound) and third (upper bound) quartiles, whiskers represent 1.5 times the interquartile range. (**d**) Kaplan-Meier plots for Relapse-Free Survival and Distant Metastases-Free Survival based on IL-8 expression were generated for patients with basal-like (n=478 for RFS, 215 for DMFS), luminal A (n=1370 for RFS, 707 for DMFS) and luminal B (n=869 for RFS, 321 for DMFS) breast cancers. Above (high, in red) and below (low, in black) the median expression level was used as the threshold in all cohorts. *P*-value calculated using log-rank test. (**e**) Kaplan-Meier plot for overall survival based on CXCL1 expression for patients with non-small cell lung cancer (NSCLC, n=1,404) and (**f**) renal adenocarcinomas (n=469) from the TCGA dataset. Above (high, in red) and below (low, in blue) the median expression level was used as the threshold. *P*-values for Kaplan-meier plots calculated using log-rank test.



Supplementary Figure S9. Regulation of IL-8 by various miR-200 members in several cancer types. Compared with negative control miRNA (black bars), varying degrees of regulation of IL-8 (red bars) is observed following transfecting renal (SN12C), lung (H226, H1299, A549), ovarian (HeyA8) and breast (MB231, BT549, MCF7, T47D) cancer cell lines with various miR-200 members. The specific miR-200 member is indicated above each cell line. Relative expression levels of IL-8 and CXCL1 in luminal breast cancer cell lines (MCF-7 and T47D) 48 hours following transfection with miRNA inhibitors. Data are averages ± SEM, n=3.



Supplementary Figure S10. Effects of miR-200 on VEGFa in ovarian cancer cell lines. (**a**) Relative expression levels of VEGFa for HeyA8 and ES2 clones expressing either empty vector, miR-200a or miR-200b. (**b**) ELISA for IL-8, CXCL1 and VEGFa from the conditioned media of ES2 cells 48 hours following miRNA transfection. ** $P<0.01$ Data are averages ± SEM, n=3. *P*-values calculated using student's t-test.



Supplementary Figure S11. MiR-200 express levels of endothelial versus ovarian cancer cell lines. Quantitative PCR was used to compare relative miR-200a and miR-200b expression levels of murine (MOEC) and human (RF24) endothelial cell lines with ovarian carcinoma cell lines of low (HeyA8) and more epithelial (SKOV3-ip1) expression levels. Data are averages ± SEM, n=3.

IL-8

```
    3' u-GUAGCAAUGGUCUGUCACAAU 5'      miR-200a        3' ug-UAGCAAUGGUCUGUCACAAU 5'      miR-200a
         |||| |||||  |||||| |:                             :|| ||:|||| | |||||:
553:5' acCAUC-UUACCUCACAGUGAUG 3'      IL8  3'UTR  950:5' cuuGUCAUUGCCAG-CUGUGUUG 3'      IL8  3'UTR

    3' aGUAGUAAUGGUCCG--UCAUAAU 5'     miR-200b        3' aguaguaaugguccgUCAUAAu 5'       miR-200b
        |||  |||:   |:  |||||||                                 ||||||
183:5' aCAUACUUAUAU-GUAAAGUAUUA 3'     IL8         191:5' cauacuuauauguaaAGUAUUa 3'       IL8

    3' aguaguAAUGGUCCG-UCAUAAu  5'     miR-200b        3' aguaGUAAUGGUCCGUCAUAAU 5'       miR-200b
           |||       ||||||                                 | |   ::   |||||||
206:5'  cauacUUAUAUGUAAAGUAUUau 3'     IL8         191:5' cauaCUUAUAUGUAAAGUAUUA 3'       IL8

    3' aguaguaaugguccgUCAUAAu 5'       miR-200b        3' AGUAGUAAUGGUCCGUCAUAAU 5'       miR-200b
                  ||||||                                  |:  :|     |   |||||||
528:5' uuucuaaguggaaaaAGUAUUa 3'       IL8         528:5' UUUCUAAGUGGAAAAAGUAUUA 3'       IL8

    3' a-GUAGUAAUGGUCCGUCAUAAU 5'      miR-200b        3' aguaguaaugguccgUCAUAAu  5'      miR-200b
         |||| |||||   ||||: |:                                 ||||||
553:5' acCAUC-UUACCUCACAGUGAUG 3'      IL8  3'UTR  543:5' uuucuaaguggaaaaAGUAUUag 3'      IL8
```

```
        3'  AGUAGUAAUGGUCCGUCAUAAU 5'      miR-200b
            |  |||||:|||| | ||:||:
950:5' cUUGUCAUUGCCAG-CUGUGUUG 3'      IL8  3'UTR

950:5'  cUUGUCAUUGCCAGCUCACAAG 3'      IL8  3' UTR Mutation
```

Legend

microrna.org	=	blue
TargetScan	=	purple
microT	=	green
RNA22	=	red
PITA	=	orange

CXCL1

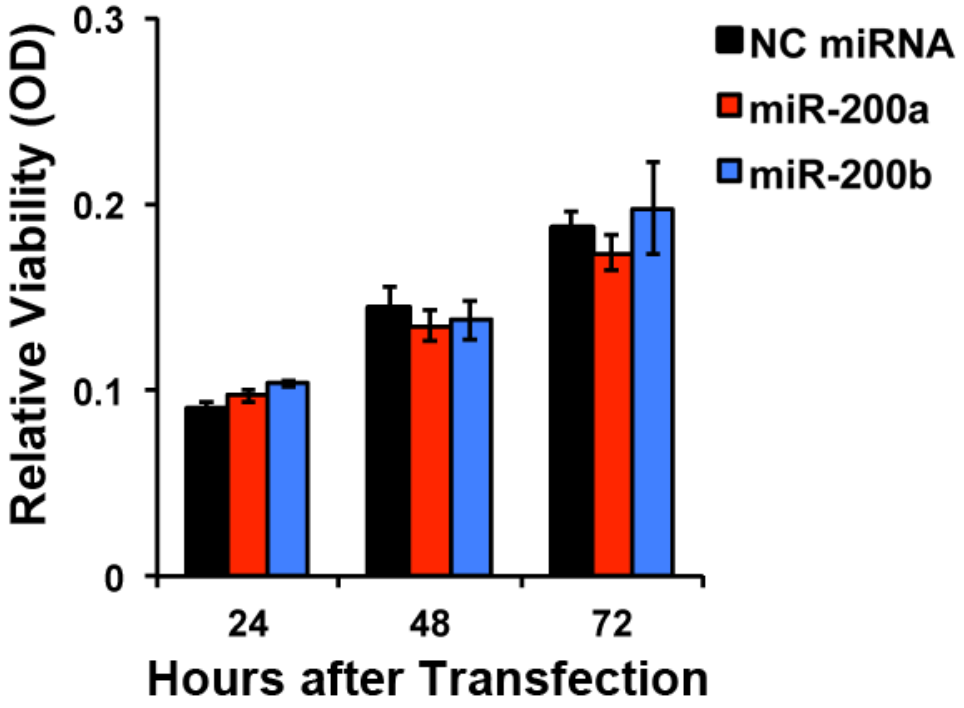
```
    3' ugUAGCAAUGGUCUGUCACAAu 5'       miR-200a        3'  uguagcaauggucuGUCACAAu 5'      miR-200a
       || || :::  ||||||||                                        |||||||
611:5' uaAUGGUAGUUUUACAGUGUUu 3'       CXCL1       625:5' uuaaugguaguuuuaCAGUGUUu 3'      CXCL1

    3' ugUAGCAAUGGUCUGUCACAAu 5'       miR-200a
       || || :::  ||||||||
611:5' uaAUGGUAGUUUUACAGUGUUu 3'       CXCL1
```

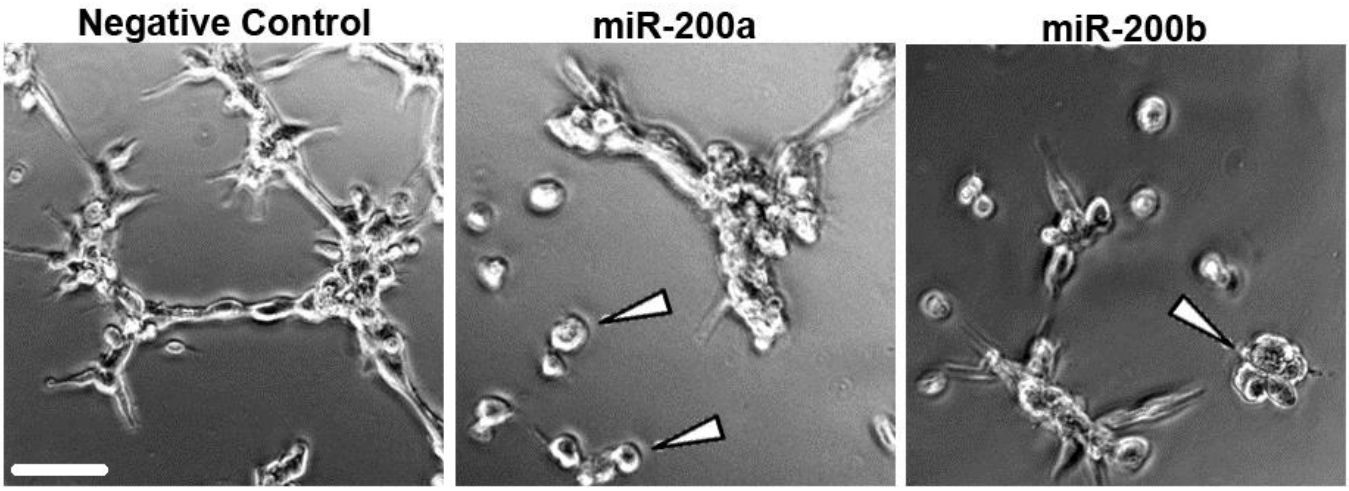
```
        3' AGUAGUAAUGGUCCGUCAUAAu 5'       miR-200b
           | || :| :::   ||||:||
611:5' UAAUGGUAGUUUUACAGUGUUu 3'       CXCL1

611:5'  UAAUGGUAGUUUUACACACAAU 3'      CXCL1   3' UTR Mutation
```

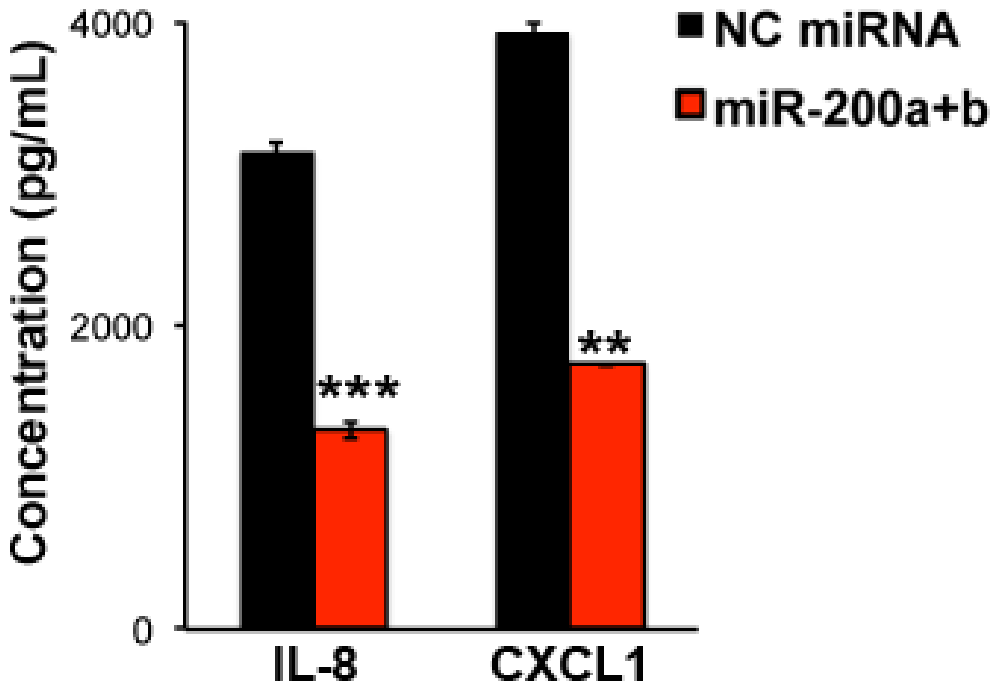
Supplementary Figure S12. Predicted miR-200 binding and site mutagenesis for IL-8 and CXCL1. Predicted binding sites identified using multiple *in silico* predictive tools of the miR-200 family for IL-8 (top) and CXCL1 (bottom). The 3' UTR regions for IL-8 and CXCL1 were mutated (red) at the indicated regions in the miR-200 seed sequence.



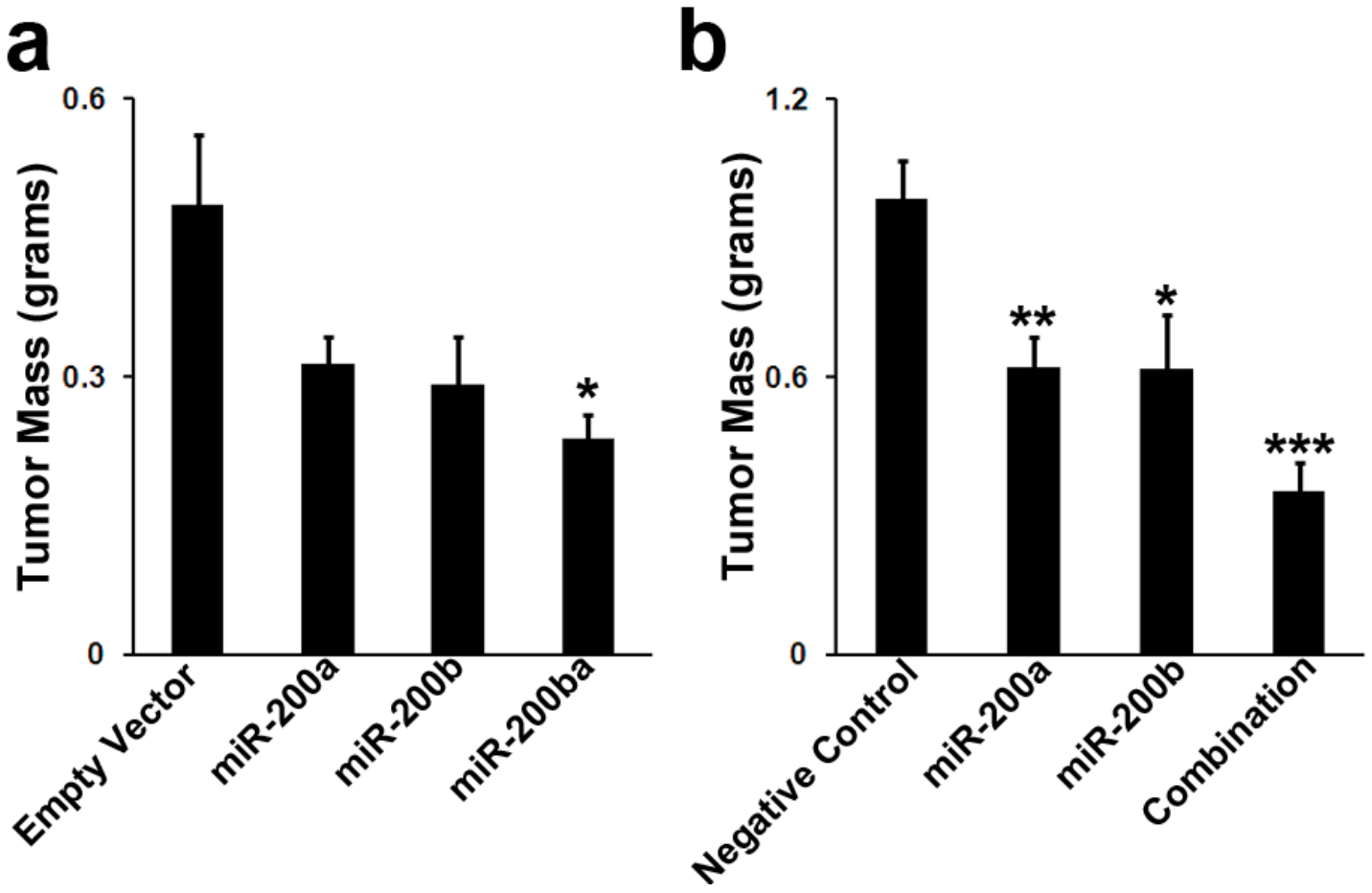
Supplementary Figure S13. Viability of RF-24 endothelial cells following miRNA transfection. After RF-24 endothelial cells were transfected with negative control miRNA, miR-200a or miR-200b, MTT assays were performed to assess cell viability at 24, 48 and 72 hour time-points. Data are averages ± SEM, n=3.



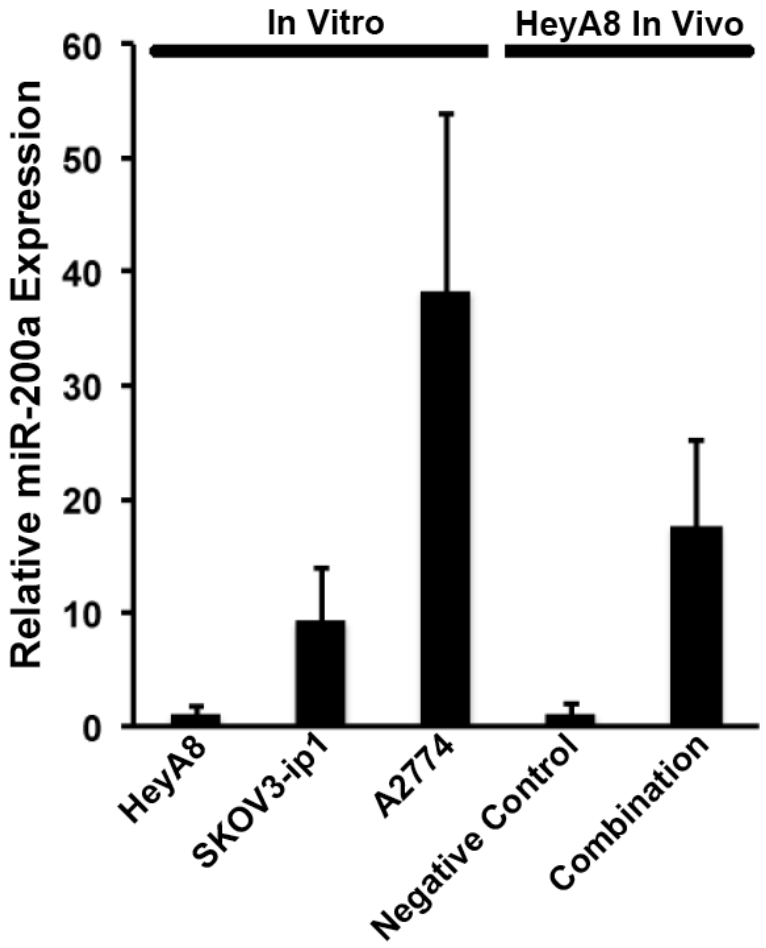
Supplementary Figure S14. Morphologic changes of endothelial cells following direct miR-200 transfection. Bright-field microscopy of tube formation at 6 hours shows significant blunting of pseudopodia and more rounded three-dimensional structures (arrows) in the miR-200a and miR-200b transfected endothelial cells as compared to those following negative control miRNA transfection. Scale bar 500 μm.



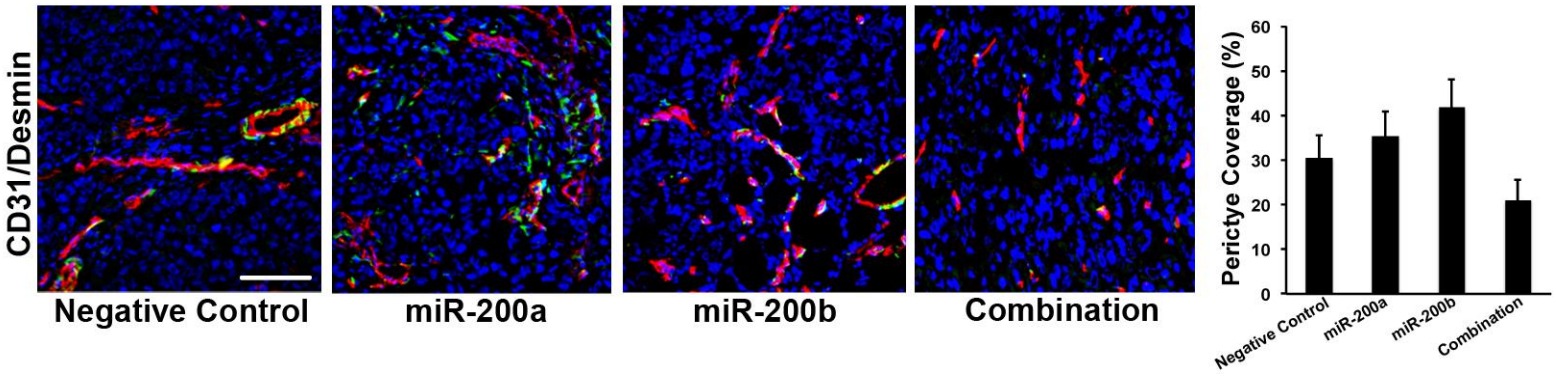
Supplementary Figure S15. ELISA for IL-8 and CXCL1 from the conditioned media of HeyA8 cells 24 hours following miRNA transfection. This conditioned media was also used for the matrigel plug assay (Fig. 3o). Data are averages ± SEM, n=3. *P*-values calculated using student's t-test. ** $P<0.01$, *** $P<0.001$



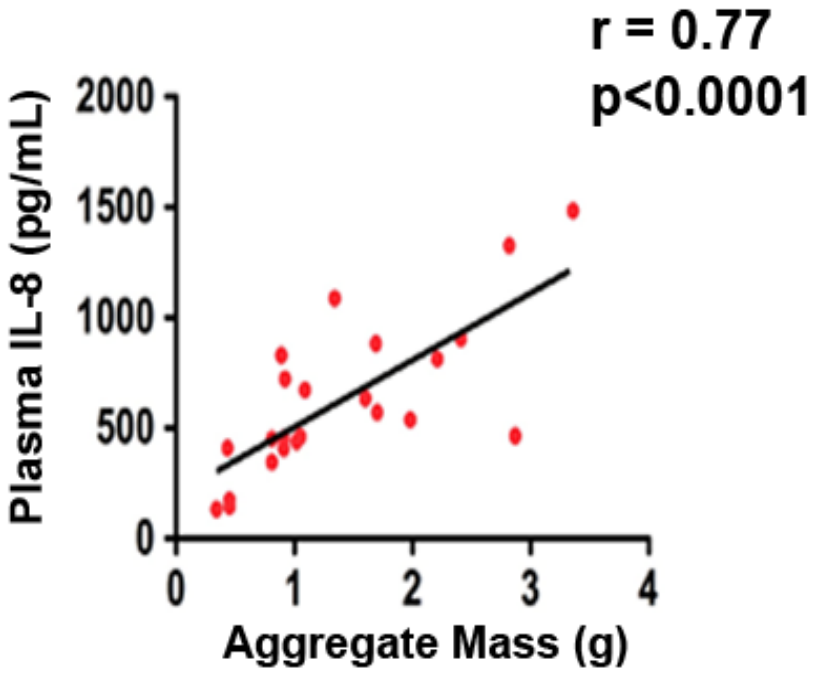
Supplementary Figure S16. Average tumor masses after injection of either (a) 344SQ clones, or (b) 344SQ wild-type cells. For 344SQ clones, $1x10^6$ cells per mouse (10 mice per group) were injected subcutaneously over the left flank and allowed to grow for 14 days, at which point mice were moribund. For 344SQ wild-type cells, $1x10^5$ cells per mouse (10 mice per group) were injected subcutaneously over the left flank. Once tumors became palpable (approximately 7 days), twice weekly delivery of miRNA-DOPC began and continued for a total of 4 treatments. Data are averages ± SEM, n=10. *P*-values calculated using student's t-test. * $P<0.05$, ** $P<0.01$, *** $P<0.001$



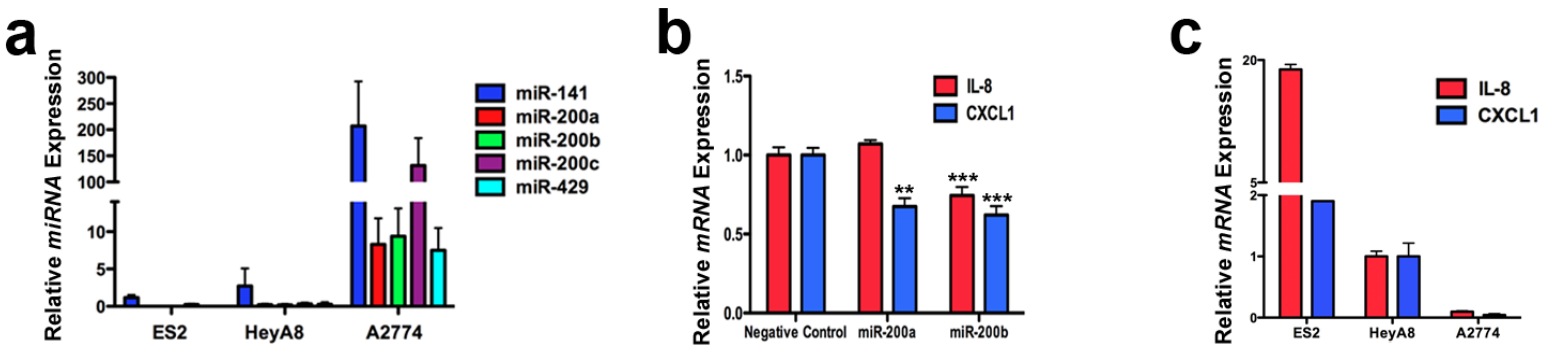
Supplementary Figure S17. Real-time PCR for miR-200a expression was performed on three ovarian cancer cell lines (HeyA8, SKOV3-ip1 and A2774) with varying levels of miR-200 member expression. For the HeyA8 *in vivo* experiment, five tumors from each group (Negative Control miRNA; Combination: miR-200a and miR-200b) were profiled for miR-200a expression. Data are averages ± SEM, n=5.



Supplementary Figure S18. Indirect effects of miR-200 targeting (IL-8 and CXCL1) on pericyte coverage. Dual desmin and CD31 staining and percentage of pericyte coverage for HeyA8 ovarian tumors following miRNA delivery, n=5 tumors per group. Data are averages ± SEM., 100 μm.



Supplementary Figure S19. Correlation of plasma IL-8 with disease burden. Correlation of plasma IL-8 levels, obtained using ELISA, with aggregate tumor burden from a metastatic ovarian cancer model (HeyA8). *P*-value was obtained using Spearman's test.



Supplementary Figure S20. Expression profiles of ovarian cell lines for relative miR-200, IL-8 and CXCL1 expression. (**a**) Relative expression of miR-200 members for ES2, HeyA8 and A2774 cell lines. Data are averages ± SEM, n=3. (**b**) Relative expression of angiogenesis (IL-8 and CXCL1) target genes following transfection of A2774. *P*-values calculated using student's t-test. (**c**) Relative expression of endogenous IL-8 and CXCL1 for ES2, HeyA8 and A2774. Data are averages ± SEM, n=3. ** $P<0.01$, *** $P<0.001$

Supplementary Table S1. Table showing normalized hazard ratios of the miR-200 family isoforms of high-grade serous ovarian, renal, lung and breast adenocarcinomas from The Cancer Genome Atlas (TCGA). The most statistically significant miR-200 members are in bold. Note: progression-free survival (PFS) was only available for ovarian cancer.

Cancer	miRNA	Normalized HR (OS)	p-value (OS)	Normalized HR (PFS)	p-value (PFS)
Ovarian (n=475)	miR-141-3p	1.163	0.253	1.307	0.2
	miR-200a-3p	**-1.796**	**0.059**	**-1.979**	**0.038**
	miR-200b-3p	**-2.042**	**0.032**	**-2.468**	**0.009**
	miR-200c-3p	-1.34	0.175	-0.254	0.799
	miR-429	**-1.248**	**0.192**	**-1.781**	**0.058**
Lung (n=508)	miR-141-3p	-1.350	0.178		
	miR-200a-3p	**-3.317**	**0.001**		
	miR-200b-3p	**-3.426**	**0.001**		
	miR-200c-3p	-0.513	0.609		
	miR-429	**-2.887**	**0.004**		
Renal (n=502)	miR-141-3p	0.727	0.475		
	miR-200a-3p	-1.158	0.26		
	miR-200a-5p	**-2.559**	**0.015**		
	miR-200b-3p	-1.262	0.216		
	miR-200b-5p	**-2.181**	**0.037**		
	miR-200c-3p	1.103	0.287		
	miR-429	-1.016	0.318		
Breast (n=929)	**miR-141-3p**	**1.779**	**0.078**		
	miR-200a-3p	1.378	0.164		
	miR-200b-3p	0.872	0.378		
	miR-200c-3p	0.398	0.691		
	miR-429	**1.835**	**0.065**		

Supplementary Table S2. Comparison of miR-200 levels with EMT markers from patient samples within The Cancer Genome Atlas (TCGA). For breast, ovarian and renal cancers within The Cancer Genome Atlas, the best threshold (as used in survival analyses) was used to divide patients into low and high expression subsets. Demonstrated are p-values (using t-tests) of the comparison between expression levels of common EMT genes between these two subsets [p-values less than 0.05 are highlighted in either red (inversely associated with miR-200 member) or green (directly associated with miR-200 member)]. *P*-values calculated using student's t-test.

Best Thresholds

Tumor	miRNA	Low	High	Total	CDH1	CDH2	SNAI1	SNAI2	TWIST1	VIM	ZEB1	ZEB2
Breast Cancer	hsa.mir.141	108	218	326	0.501128	0.014837	0.012235	<0.000001	0.00002	<0.000001	<0.000001	<0.000001
Ovarian Cancer	hsa.mir.200c	192	374	566	0.002457	0.001923	0.000056	<0.000001	0.000016	<0.000001	0.000016	0.000037
Renal Cancer	hsa.mir.200a	114	291	405	0.00001	0.103044	0.009781	0.000043	0.086431	0.000722	0.190923	0.043078

Supplementary Table S3. Comparison of miR-200 levels with EMT markers from patient samples using above/below median for threshold. For breast, ovarian and renal cancers within The Cancer Genome Atlas, the median threshold for each miR-200 family member was used to divide patients into low and high expression subsets. Demonstrated are p-values of the comparison between expression levels of common EMT genes between these two subsets [p-values less than 0.05 are highlighted in either red (inversely associated with miR-200 member) or green (directly associated with miR-200 member)]. Note: Several samples in renal cancer (miR-141 and miR-429) had values exactly equal to the median and these were arbitrarily placed into the low expression group. *P*-values calculated using student's t-test.

Median Thresholds

Tumor	miRNA	Low	High	Total	CDH1	CDH2	SNAI1	SNAI2	TWIST1	VIM	ZEB1	ZEB2
Breast Cancer	hsa.mir.141	163	163	326	0.648698	0.001705	0.014492	<0.000001	0.000423	<0.000001	0.000054	0.000042
Breast Cancer	hsa.mir-200a	163	163	326	0.422872	0.878685	0.096953	0.051586	0.082838	0.034976	0.005239	0.012118
Breast Cancer	hsa.mir.200b	163	163	326	0.924718	0.489629	0.047334	0.003904	0.117831	0.000847	0.003298	0.002122
Breast Cancer	hsa.mir.200c	163	163	326	0.590151	0.144036	0.001629	0.000002	0.000056	0.000002	0.000153	<0.000001
Breast Cancer	hsa.mir.429	163	163	326	0.408798	0.683148	0.280562	0.004639	0.003865	0.000083	0.000129	0.005459
Ovarian Cancer	hsa.mir.141	283	283	566	0.000227	0.056015	0.00046	0.000224	0.000138	0.000023	0.005922	0.008873
Ovarian Cancer	hsa.mir-200a	283	283	566	0.045331	0.296545	0.002493	<0.000001	0.000079	0.00013	<0.000001	<0.000001
Ovarian Cancer	hsa.mir.200b	283	283	566	0.256613	0.004317	0.038058	0.000001	0.000217	0.000028	<0.000001	<0.000001
Ovarian Cancer	hsa.mir.200c	283	283	566	0.0049	0.000889	0.002497	<0.000001	0.000047	<0.000001	0.000032	0.000141
Ovarian Cancer	hsa.mir.429	283	283	566	0.216419	0.089412	0.156239	0.00045	0.001063	0.00035	0.000033	0.000002
Renal Cancer	hsa.mir.141	225	180	405	0.000756	0.668462	0.955865	0.928528	0.640669	0.037398	0.149439	0.198099
Renal Cancer	hsa.mir-200a	203	202	405	<0.000001	0.416983	0.002911	0.000008	0.06867	0.000028	0.013035	0.019107
Renal Cancer	hsa.mir.200b	203	202	405	<0.000001	0.406914	0.00172	0.000008	0.118284	0.000017	0.006503	0.000397
Renal Cancer	hsa.mir.200c	203	202	405	0.001181	0.217636	0.4991	0.678251	0.544464	0.025172	0.439627	0.589114
Renal Cancer	hsa.mir.429	207	198	405	0.00015	0.105885	0.214459	0.000085	0.158576	0.002449	0.009327	0.002561

Supplementary Table S4. Comparison of miR-200 levels with EMT markers from patient samples using above/below median for threshold. For basal-like and luminal breast cancers within The Cancer Genome Atlas, the median threshold for each miR-200 family member was used to divide patients into low and high expression subsets. Demonstrated are p-values of the comparison between expression levels of common EMT genes between these two subsets [p-values less than 0.05 are highlighted in either red (inversely associated with miR-200 member) or green (directly associated with miR-200 member)]. *P*-values calculated using student's t-test.

Median Thresholds

Tumor	miRNA	Low	High	Total	CDH1	CDH2	SNAI1	SNAI2	TWIST1	VIM	ZEB1	ZEB2
Basal Subtype	hsa.mir.141	27	27	54	0.981253	0.87017	0.673341	0.016835	0.190716	0.020371	0.033809	0.108549
Basal Subtype	hsa.mir-200a	27	27	54	0.975471	0.06229	0.272827	0.200086	0.470222	0.858709	0.186588	0.06007
Basal Subtype	hsa.mir.200b	27	27	54	0.630485	0.020172	0.128065	0.429868	0.90576	0.68016	0.17905	0.008538
Basal Subtype	hsa.mir.200c	27	27	54	0.230082	0.228018	0.644071	0.040227	0.082286	0.106288	0.008825	0.017671
Basal Subtype	hsa.mir.429	27	27	54	0.263462	0.386335	0.055508	0.532915	0.416143	0.134729	0.065826	0.001763
Luminal Subtype	hsa.mir.141	116	116	232	0.639644	<0.00001	0.002873	<0.00001	0.000977	<0.00001	0.000438	0.000482
Luminal Subtype	hsa.mir-200a	116	116	232	0.271001	0.192737	0.102059	0.378931	0.199846	0.162418	0.038298	0.217797
Luminal Subtype	hsa.mir.200b	116	116	232	0.782774	0.070269	0.054018	0.011846	0.174735	0.001714	0.006549	0.03684
Luminal Subtype	hsa.mir.200c	116	116	232	0.760363	0.000446	0.000239	<0.00001	0.000411	<0.00001	0.000933	<0.00001
Luminal Subtype	hsa.mir.429	116	116	232	0.159608	0.834481	0.226893	0.024526	0.004139	0.001597	0.014674	0.292678

Supplementary Methods

Cell viability assay. Cell viability assays were performed by testing cell's ability to reduce the tetrazolium salt [3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium, inner salt] to a formazan. Briefly, RF24 cells were seeded in a 96-well plate and were reverse transfected with either control miRNA, miR-200a or miR-200b as described before. At 24, 48 and 72 hours following transfection, cells were incubated with 0.15% 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) for 2 hours at 37°C. The supernatant was removed, cells were dissolved in 100 µL DMSO and the absorbance at 570 nm was recorded.

Bioinformatic Analysis. We used clinically annotated data from The Cancer Genome Atlas (TCGA) obtained from the Open-Access and Controlled-Access tiers of the TCGA Data Portal (http://tcga-data.nci.nih.gov/tcga/findArchives.htm) with NIH approval. Our preliminary analysis focused on nine different cancer types, subsequently reduced to four that had a large number of samples (>200). Expression data for all five miR-200 members (miR-141, -200a, -200b, -200c, and -429) were obtained from either Agilent miRNA microarrays (8x15K; ovarian cancers) or miR-Seq (Illumina GA and HiSeq; breast and clear cell renal cancers). The data we downloaded from the TCGA website and used for analysis can be found in Supplementary Data 1. Approximately 2/3 of the samples for each tumor type were used as a training cohort to obtain the miRNA expression thresholds yielding the most significant log rank test p-values (one-tailed test), thus dividing the data into good and poor prognosis groups. These thresholds were then used to divide the data in the remaining 1/3 of the samples that functioned as independent validation cohorts, and p-values were computed for those cohorts. The same procedure was repeated again to look for significant differences in progression free survival. Kaplan-Meier curves were also plotted for all datasets, with a p-value < 0.05 considered significant.

Algorithms for predicting miRNA binding sites were utilized and are publically available (see luciferase experiments). For each putative miRNA-mRNA regulatory pair (mRNA with corresponding miRNA predicted binding sites in 3'-UTR), the person correlation coefficient was further calculated using the paired mRNA and miRNA expression data in TCGA database. The genes, which have binding sites in the 3'-UTRs and are inversely correlated (FDR<0.01, Benjamini-Hochberg multiple testing correction), were predicted to be the targets of miRNAs. The pathway analysis was performed by ingenuity pathway analysis (IPA) software.

Scoring of network components (nodes, edges) for their relevance to a given dataset is done by a random walk–based scoring strategy by using the data values of nodes as transition probabilities. Briefly, each interaction *i–j* is assigned a probability value (p_{ij}) based on the data values of nodes in the neighborhood, **Equation S1:**

$$p_{ij} = \frac{w_j}{\sum_{k \in N_i} w_k}, \qquad [S1]$$

where w_j is the experimental value for node *j* and N_i is the set of immediate downstream neighbors of node *i*. If there are no downstream nodes of the node *i* ($|N_i| = 0$), p_{ij} is set to $p_{ij} = 1/n$ for all *j*, where n is the total number of nodes in the network. Also at each step, we assign a small probability ($q = .01$) that the random walk will "jump" to any other node in the network. Final relevance scores of nodes are given by their visitation frequencies by the random walk in the end of infinite iterations (ie, value at the stationary distribution of the random walk).

To determine clinical relevance of IL-8 and CXCL1 expression in breast and lung cancer subtypes, we utilized large public microarray databases using Affymetrix HGU133A and HGU133+2 data that have been previously described[43]. Samples were classified as having high or low expression based on being above or below the median expression level, respectively. Each molecular subtype of breast cancer (basal-like, luminal A or B) was considered individually

for overall survival (n=185, 459 and 303, respectively), relapse-free survival (n=478, 1370 and 869, respectively) and distant metastasis-free survival (n=215, 707 and 321, respectively). Similarly, non-small cell lung cancers (n=1,404), lung adenocarcinomas (n=486) and lung squamous carcinomas (n=421) histologies were considered separately for overall survival. The following cohorts were included for the lung cancer analysis (CAARRAY, GSE14814, GSE19188, GSE29013, GSE31210, GSE3140, GSE37745, GSE4573, GSE8894 and TCGA).

mRNA microarray. Total RNA was extracted from the HeyA8 and 344SQ lentiviral clones using a mirVana RNA Isolation labeling kit (Ambion, Inc., Austin, TX). RNA purity was assessed by a Nanodrop spectrophotometric measurement (Thermo Scientific) of the OD260/280 ratio with acceptable values falling between 1.9 and 2.1. Five hundred nanograms of total RNA were used for labeling and hybridization on a Human HT-12 v4 Beadchip (HeyA8; Illumina, San Diego, CA) or a Mouse WG-6 v2 Beadchip (344SQ; Illumina, San Diego, CA) according to the manufacturer's protocols. After the bead chips were scanned with an Illumina BeadArray Reader (Illumina, San Diego, CA), the microarray data were normalized using the quantile normalization method in the Linear Models for Microarray Data (LIMMA) package in the R language environment. The expression level of each gene was transformed into a $\log_2$ base before further analysis.

miRNA *in situ* hybridization. Tissue microarray samples for lung (Vanderbilt University), ovarian (Wayne State), and basal-like breast (MD Anderson Cancer Center) cancers were obtained and prepared following institutional review board approval for each institution. The formalin-fixed paraffin embedded tissue sections were dewaxed in xylenes, and rehydrated through an ethanol dilution series. Tissue sections were digested with 15 µg/mL proteinase K for 20 minutes at RT, then loaded onto Ventana Discovery Ultra for *in situ* hybridization analysis. The tissue slides were incubated with double-DIG labeled mercury LNA miR-200b (lung cancers) or miR-200c (ovarian and basal-like breast cancers) microRNA probe (Exiqon) for 2

hrs at 52 C. The double-DIG labeled control U6 snRNA probe (Exiqon) was used as a positive control. The digoxigenins were then detected with a polyclonal anti-DIG antibody and Alkaline Phosphatase conjugated second antibody (Ventana) using NBT-BCIP as the substrate. Representative light field images were obtained using a Nikon Microphot-FXA microscope and Leica DFC320 digital camera. The expression levels for miR-200b and miR-200c were determined using CellProfiler 2.0 software[44] to establish staining intensity threshold levels and to quantify number of positively staining cancer cells per high-powered field (200x magnification).

Target Gene Binding Sites, Luciferase Reporter Assays and 3' URT Site Mutagenesis. The putative binding sites for miR-200a and miR-200b were predicted bioinformatically using several algorithms for predicting miRNAs targets and binding sites for IL-8 and CXCL1. This was done utilizing the following publically available sites: http://www.microrna.org for the miRanda algorithm, http://www.targetscan.org for the TargetScan algorithm, http://genie.weizmann.ac.il/pubs/mir07 for the PITA algorithm, http://cbcsrv.watson.ibm.com for the RNA22 algorithm, http://diana.cslab.ece.ntua.gr/microT for microT algorithm, and http://genome.ucsc.edu/cgi-bin/hgTables?command=start together with http://pictar.mdc-berlin.de/ for the PicTar algorithm. We used Perl to retrieve and sort information available through these sources, and Latex to present the miRNA binding sites most probable to interact. GoClone pLightSwitch luciferase reporters for the 3' UTR regions of IL-8 and CXCL1 were obtained form SwitchGear Genomics (Menlo Park, CA). SKOV3 cells were transfected with FuGENE HD TFX reagent in a 96-well plate with scrambled control, miR-200a or miR-200b mimics (Ambion, 100 nM) along with the 3' UTR reporter gene and Cypridina TK control construct (pTK-Cluc). After 24 hours of transfection, luciferase activity was obtained with the LightSwitch Dual Luciferase assay kits using a microplate luminometer per manufacturer guidelines. Two independent experiments were performed with 3 replicates each. Luciferase

activity was normalized with the Cypridina TK control construct, and an empty luciferase reporter vector was used as a negative control. The ratios obtained were further normalized according to the scrambled control. Mutants of the IL-8 and CXCL1 3' UTR were generated using the QuikChange Lightning Multi Site-direct Mutagenesis kit (Agilent Technologies, La Jolla, CA) using the following primers to mutate 5 base pairs within the miR-200a and miR-200b binding sites:

IL-8: (forward) actcccagtcttgtcattgccagctcacaaggtagtgctgtgttgaattacg;

(reverse) cgtaattcaacacagcactaccttgtgagctggcaatgacaagactgggagt

CXCL1: (forward) tgtaaggcatactgccttgtttaatggtagttttacacacaatctggcttagaacaaaggg;

(reverse) ccctttgttctaagccagattgtgtgtaaaactaccattaaacaaggcagtatgccttaca

Proper site mutagenesis was confirmed with sequencing prior to luciferase assays.

Immunostaining. Staining was performed in formalin-fixed, paraffin embedded tumor sections (8 µm thickness) or from OCT embedded frozen tissue sections. After deparaffinization, rehydration and antigen retrieval or fixation, 3% H_2O_2 was used to block the endogenous peroxidase activity for 10 minutes. Protein blocking of non-specific epitopes was done using either 5% normal horse serum, 1% normal goat serum or 2.8% fish gelatin in either PBS or TBS-T for 20 minutes. Slides were incubated with primary antibody for ZEB1 (Sigma HPA027524, 1:600), E-cadherin (BD Transduction Laboratories 610181, 1:50), CD-31 (rat monoclonal anti-mouse, 1:800, Pharmingen), Desmin (rabbit anti-mouse, Abcam 15200), Ki-67 (rabbit anti-mouse, 1:200, Abcam ab15580) or IL-8 (rabbit polyclonal anti-human, 1:25, Biosource) overnight at 4 ºC. For immunohistochemistry, after washing with PBS, the appropriate amount of horseradish peroxidase-conjugated secondary antibody was added and visualized with 3,3'-diaminobenzidine chromogen and counterstained with Gill's hematoloxylin #3. For immunofluorescence, secondary antibody staining was performed with either Alexa 594

(Molecular Probes) or DyLight (Jackson ImmunoResearch). Nuclear staining was performed with Hoechst 33342 (1:10,000, Molecular Probe H3570). Light field images were obtained using a Nikon Microphot-FXA microscope and Leica DFC320 digital camera, while immunofluorescent images were obtained using a Zeiss Axioplan 2 microscope and Hamamatsu ORCA-ER digital camera. The expression levels for IL-8 in an ovarian tissue microarray was determined using CellProfiler 2.0 software[44] to establish staining intensity threshold levels and to quantify number of positively staining cancer cells per high-powered field (200x magnification). To quantify microvessel density (MVD), we examined 5-10 random fields at 100x magnification for each tumor (5 tumors per group) and counted the microvessels within those fields. A vessel was defined as an open lumen with at least one adjacent CD31-positive cell. Multiple positive cells beside a single lumen were counted as one vessel, and quantification was performed by two investigators in a blinded fashion. Proliferation indices were determined using three representative fields at 200x magnification for each tumor (5 tumors per group). All Ki-67 positive cells per high-powered field were enumerated. For pericyte coverage analysis, 5 random fields at 200x magnification were taken per slide (5 tumors per group), and the percentage of CD-31 staining blood vessels with at least 50% pericyte coverage (positive desmin staining) were enumerated per high powered field. Quantification was performed in a blinded fashion.

SUPPLEMENTARY REFERENCES:

42. Landen, C.N.*, et al.* Therapeutic EphA2 Gene Targeting In vivo Using Neutral Liposomal Small Interfering RNA Delivery. *Cancer Research* **65**, 6910-6918 (2005).

43. Gyorffy, B.*, et al.* An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast cancer research and treatment* **123**, 725-731 (2010).

44. Lamprecht, M.R., Sabatini, D.M. & Carpenter, A.E. CellProfiler: free, versatile software for automated biological image analysis. *BioTechniques* **42**, 71-75 (2007).