Supplemental Methods

Defining ER and PR status from gene expression data

We first analyzed association of known IHC status (n=475 patients) with expression level of *ESR1* and *PGR* genes. As shown on Fig. S1, for both ER and PR status, IHC-negative and IHC-positive groups were clearly different by the expression levels for corresponding genes.

Based on those data, we selected the threshold values for *ESR1* and *PGR* genes giving the best agreement (i.e. minimal number of misclassified patients) with IHC data (Fig. S2). This threshold was then applied for all 606 cases. Among the 475 patients with known IHC status, ESR1 expression-derived status was discordant from IHC-derived ER status for 39 patients; PGR expression-derived status was discordant from IHC-derived PR status for 75 patients; overall hormonal receptor status (i.e. positive when any of the two receptors positive) was discordant for 32 patients.

ActMiR procedure

We previously developed ActMiR, a method for inferring miRNA activity based on expression levels of miRNAs and their predicted target genes ([10](#_ENREF_10)). Three pieces of information were used: (i) miRNA expression levels of samples; (ii) mRNA expression levels of samples; (iii) the predicted target lists of each miRNA. For the predicted target list of miRNAs, we used a collection of predicted target genes for 1537 unique mature miRNAs from TARGETSCAN (www.targetscan.org) that considers all conserved miRNA binding sites inherited from 23-way alignments of UTR sequences ([48](#_ENREF_48)). In order to obtain robust results, we filtered out miRNAs whose number of target genes is smaller than 10. Among these miRNAs, we further focused on miRNAs whose predicted target genes’ expression levels and their own expression levels are available. The ActMiR method consists of three steps. First, for each miRNA, we estimated the “baseline” expression levels of miRNA’s target genes at the state where the miRNA had no impact. As sufficient miRNA concentration is essential for its functional activity, we defined baseline expression level *Y*bφt of the target gene t of miRNA φ as the average expression level of the samples with low miRNA expression level. Next, we defined the “degradation” levels as the difference between the observed expression levels of targeted genes for each sample, which is affected by the miRNA, and the baseline expression level, which is unaffected by the miRNA. For each sample s, degradation levels *Y*bφts of predicted target *t* of miRNAs φ is determined as follows: *Y*dφts = *Y*bφt - *Y*φts, where *Y*bφts is the observed expression level of the predicted target. The expression degradation level allows us to measure how much expression level change of the target gene is potentially affected by each miRNA. Finally, based on the assumption that the impact of a miRNA on its target genes depends on its expression level, we used a linear model representing the relationship between the degradation levels and baseline expression levels of target genes for each sample, in which the coefficient represents miRNA activity in each sample. Positive miRNA activity corresponds to the high degradation effect on their targets.

We calculate Pearson correlation between activity level of miRNA and expression level of mRNA for ER+ and ER-/Her2- tumors, separately. We also calculate correlation between expression level of miRNA and expression level of mRNA for both subtypes.

The functional target genes are defined as genes whose expression levels are significantly correlated with miRNA activity and are predicted target genes of the miRNA based on TARGETSCAN([48](#_ENREF_48)). The significant connection between activity of miRNA and mRNA was assessed by permutation tests. For each permutation, sample labels were randomly assigned to expression profiles in order to maintain the correlation structure among transcripts. The permutation was performed 20 times and the average number of significant mRNA-miRNA connections over 20 permuted data sets at a specific absolute correlation was used to calculate a FDR. The absolute correlation coefficient 0.322 and 0.514 corresponded to a highly conservative false discovery rate (FDR<1x10-4) for ER+ and ER-/HER2- samples, respectively.