Supplementary table legends

**Supplementary Table S1:** Chi-Square tests to assess difference in distribution between CTC-positive and CTC-negative patients with respect to clinico-pathological features. #Wilcoxon rank sum test, *logrank test, data corresponding to the primary tumour, median could not be calculated as number of evens are too low. Abbreviations: ILC- invasive ductal carcinoma, ILC- invasive lobular carcinoma, ER- oestrogen receptor, PR- progesterone receptor, HER2- human epidermal growth factor receptor 2, PFS- Progression-free survival, OS- overall survival, NA- data not available, CI- confidence interval.

**Supplementary Table S2:** Cp value of the two run controls used to check for inter-run variation across all miRNAs and the corresponding standard deviation (SD).

**Supplementary Table S3:** Limma results of the seventeen candidate miRNAs selected from the array analysis for the validation round, with a mean fold change (FC= \(2^{\Delta Ct}>4\) or \(<0.25\)) and FDR < 0.1, listed here for each comparison.

**Supplementary Table S4:** Interaction of miRNA expression and age in cases versus control comparison. Method of Smith and Thompson (1996) to control for confounding effects; \(P < 0.1\) indicates a significant interaction. Five miRNAs interact with age with the predictive value of these miRNAs stronger in younger woman.

**Supplementary Table S5:** Age-adjusted \(P\) values for comparison of CTC-positive and CTC-negative MBC computed from Wald tests in logistic regression models including age as a co-variable.
**Supplementary figure legends**

**Supplementary Fig. S1**: Unsupervised hierarchical complete linkage clustering with Euclidean distance metric, based on the 216 miRNA features remaining after initial filtering (red: CTC-positive, green: CTC-negative, black: controls).

**Supplementary Fig S2**: Partial correlations between miRNAs based on their expression with the corresponding correlations depicted along the lines ($P < 0.00001$ for all).

**Supplementary Fig. 3**: Prediction error curves up to 15 months (PFS) or 7.5 months (OS) for the null model (Kaplan-Meier model without any covariate information), CTC, miRNA multivariable model, miRNA model + CTC for (a) PFS and (b) OS.