Supplementary Methods

ChIP assay. The sequence of miR-200c promoter was obtained from UCSC Genome Database. Analysis of putative transcription factor binding sites on miR-200c promoter was done by TRED (Cold Spring Harbor Laboratory) and MatInspector (Genomatix). Chromatin immunoprecipitation (ChIP) experiment was modified from the EZ-CHIP (Upstate) protocol using STAT3, G9a (Cell Signaling) and H3K9Me2 antibodies (Abcam). Sequential ChIP was performed as described previously (Chang et al., 2011b).

Genome-wide ChIP-sequencing analysis. Following the ChIP procedure as described above, 10 ng DNA of each sample was converted to phosphorylated blunt-ended with T4 DNA polymerase, Klenow polymerase, and T4 polymerase (NEB). An 'A' base was added to the 3' end of the blunt phosphorylated DNA fragments using the polymerase activity of Klenow (exo minus) polymerase (NEB). Illumina's genomic adapters were ligated to the A tailed DNA fragments. PCR amplification was performed to enrich ligated fragments using Phusion High Fidelity PCR Master Mix with HF Buffer (Finnzymes Oy). The enriched product of ~200-700 bp was cut out from gel and purified. The library was denatured with 0.1M NaOH to generate single-stranded DNA molecules, and loaded onto channels of the flow cell at 8pM concentration, amplified in situ using TruSeq Rapid SR cluster kit (#GD-402-4001, Illumina). Sequencing was carried out by running 100 cycles on Illumina HiSeq 2000 according to the manufacturer's instructions. Image analysis and base calling were performed using Off-Line Basecaller software (OLB V1.8). Aligned reads were used for peak calling of the ChIP regions using MACS V1.4.0. Statistically significant ChIP-enriched regions (peaks) were identified by comparison to a Poisson background model, using a p-value threshold of 10^{-4}. The peaks in samples were annotated by the nearest gene using the newest UCSC RefSeq database. The annotation of the peaks which were located within -2Kb to +2Kb around the corresponding gene TSS. UCSC genomic browser (hg19) and DAVID bioinformatics database were used to generate ChIP-seq plots and Gene Ontology functional annotations.

Luciferase reporter assay. *OBR*-3'UTR luciferase construct was purchased from Switchgear Genomics. *MiR-200c* luciferase construct was generated from our previous study in Chang et al.(Chang et al., 2011) PCR-based site-directed mutagenesis (Stratagene) was used to generate the mutants. All of these constructs were verified by DNA sequencing. Luciferase assay was performed according to the

manufacturer's instructions (Promega and Switchgear Genomics). The luciferase activity was measured with an AutoLumat LB953 luminometer (Berthold).

Antibodies. The following antibodies were used for immunoblotting according to standard protocol: anti-ZEB1 (Cell Signaling, 1:1000), anti-N-Cadherin (Abcam, 1:2000), anti-STAT3 (Cell Signaling, Dilutions 1:1000), anti-phospho-STAT3 (Tyr705) (Cell Signaling, 1:2000), anti-E-cadherin (Santa Cruz, 1:2000), anti-OBR (Santa Cruz, 1:200), G9a (Cell Signaling, 1:1000).

Mass spectrometry analysis. HPLC/MS/MS analysis was performed by Agilent 1100 liquid chromatography system and LTQ two-dimensional ion trap mass spectrometer. Database search was conducted with NCBInr database.

GST pull-down and co-immunoprecipitation assay. For GST pull-down, bacterial lysate for GST–STAT3 and recombinant His-G9a purified from baculovirus-infected Sf9 cells were incubated for 3 hrs and washed three times with pull-down buffer then prepared for immunoblotting. For co-immunoprecipitation, nuclear extract was prepared according to the manufacturer's instruction (Active Motif) and the same amounts of input and immunoprecipitation eluates by STAT3 and G9a antibodies were subjected to immnublotting.

Immunohistochemistry and *in situ* hybridization. Breast cancer tissue microarrays were purchased from Pantomics (BRC1021) and immunostained with anti-phospho-STAT3 (Tyr705) (Cell Signaling, 1:200), anti-OBR (Santa Cruz, 1:100), anti-LEP (Santa Cruz, 1:200). The histological grading and pathological annotation (tumor grade and subtype) were provided by the pathologists at Pantomics. Detection of the antibody signal was performed with the LSAB2 System-HRP (K0672, Dako) and liquid DAB+ Substrate Chromogen System (K3468, Dako) according to the manufacturer's instructions. In situ hybridization was performed using the miRNA-200c probe from Exiqon (miRCURY LNA detection probe 5'-DIG-labeled). Detection of the probe was carried out using digoxigenin antibody (Abcam ab420, 1:200), LSAB2 System-HRP (K0672, Dako) and liquid DAB+ Substrate Chromogen System (K3468, Dako) according to the manufacturer's instructions. Cell nuclei were counterstained with Mayer's hematoxylin stain. The correlation between miR-200c, LEP, p-STAT3, and OBR was analyzed using Chi-Square test (Figure 7B) or Fisher's exact analysis (Table S6).

miRNA expression array, gene expression array, and antibody array. The genome-wide human miRNA-PCR array was purchased from SABiosciences. 4x44K whole human genome oligonucleotide microarray was purchased from Agilent, and analyzed by GeneSpring GX software v11.5.1. The PathScan

Intracellular Signaling Array was purchased from Cell Signaling. All procedures and analyses were performed according to the manufacturer's instructions.

Real-time PCR. Total RNAs were extracted from cells by using RNeasy Plus kit (Qiagen). RNAs were reverse-transcribed by using Superscript II kit (Invitrogen). The results were analyzed by the iCycler (Biorad), and the quantification of RNA levels was normalized to GAPDH as CT (difference of cycling threshold) = CT (target) – CT (control). Higher CT values indicate relatively lower expression RNA levels.

Flow cytometry analysis (FACS). The APC-conjugated anti-CD44 (BD Biosciences), PE-conjugated anti-CD24 (BD Biosciences), and FITC-conjugated anti-OBR (Lifespan Biosciences) antibodies were used for FACS analysis as described previously (Chang et al., 2011). Lineage depletion was performed in primary tumor cells as described previously (Al-Hajj et al., 2003). Data was analyzed by FCS3 express (Denovo Software).

Microarray dataset. Array data is deposited in NCBI GEO database (GSE58059).

Statistical analysis. All data were presented as means± the standard deviation of the mean (SD). Statistical calculations were performed with Microsoft Excel analysis tools. Differences between individual groups were analyzed by two-tailed paired t test. P values of <0.05 considered statistically significant. Genome-wide microRNA PCR array analysis was performed according to the manufacturer’s instructions (SABiosciences). Chi-square analysis for tumor samples were performed as described previously (Chang et al., 2011).