The RON Receptor Tyrosine Kinase Promotes Metastasis by Triggering Epigenetic Reprogramming through the Thymine Glycosylase MBD4

Stéphanie Cunha¹, Yi-Chun Lin¹, Elizabeth A. Goossen¹, Christa I. DeVette¹, Mark R. Albertella², Stuart Thomson², Mark J. Mulvihill², and Alana L. Welm¹*
¹Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah, 2000 Circle of Hope, Salt Lake City, UT 84112, USA
²OSI/Astellas, Bioscience Park Drive, Farmingdale, NY 11735, USA
*Correspondence: alana.welm@hci.utah.edu
Figure S1
Figure S2
Figure S3
Figure S4
Supplemental Figure Legends

Figure S1. (A) Tumor growth curves following orthotopic transplantation of MCF7 or T47D cells (blue) or MCF7-RON/MSP or T47-RON/MSP (red) into cleared inguinal mammary fat pads of 3-week old NOD/SCID mice. (B) Representative lung sections from mice carrying MCF7-RON/MSP or T47D-RON/MSP tumors, stained with H&E or immunostained with an antibody specific for human cytokeratin to positively identify metastasis.

Figure S2. (A) Real-time quantitative RT-PCR for various DNMT mRNA expression levels, normalized to β-actin mRNA expression, in MCF7 and MCF7-RON/MSP cells. (B) DNA methylation activity in MCF7 and MCF7-RON/MSP cells, as assessed by EpiQuick DNA methyltransferase assay. (C) Real-time quantitative RT-PCR for various TET mRNA expression levels, normalized to β-actin mRNA expression, in MCF7 and MCF7-RON/MSP cells.

Figure S3. (A) Tumor growth following orthotopic implantation of MCF7 (dark blue), MCF7-RON/MSP (red), MCF7-RON/MSP-shScr (green), MCF7-RON/MSP-shMBD4 (yellow) and MCF7-RON/MSP-shMBD4R cells (light blue) in MCF7 (left) and T47D (right) models. (B) Metastasis frequencies for MCF7, MCF7-RON/MSP, MCF7-RON/MSP-shScr, and MCF7-RON/MSP-shMBD2 tumors. (C) Representative images showing lack of effect of sh-MBD2 on spontaneous lung, bone, liver and brain metastasis of RON/MSP-expressing MCF7 tumors.

Figure S4. (A) The expression of MET, MBD4 and ACTB in MCF7 cell line infected with retroviruses carrying cDNA of MET. (B) Western blots showing expression of EGFR, p-EGFR, MBD4, p-AKT, p-ERK, AKT, ERK and β-actin in MDA-MB-231 cells treated with EGF (15 nM) for 1 hour. (C) MCF7 and MCF7-RON/MSP cells were treated with or without alkaline phosphatase for 30 min. The lysates were subjected to (a) conventional SDS-PAGE on 8% (wt/vol) polyacrylamide gel or (b) SDS-PAGE on 6% (wt/vol) polyacrylamide gel containing 20 µM Mn²⁺-Phos-tag, followed by immunoblotting with the anti-MBD4 antibody. By treatment with alkaline phosphatase, shifts in the mobility of MBD4 were suppressed.
Supplemental table

**Table S1.** Differentially expressed genes (DEGs) identified from both RNA-sequencing and microarray

**Supplemental Experimental Procedures**

**Cell culture**

MCF-7 cells were cultured in Dulbecco's modified Eagle's medium (DME/F-12, Thermo Scientific) supplemented with 1% penicillin/streptomycin (Thermo Scientific), 10 µg/ml insulin (Gibco), and 10% heat inactivated fetal bovine serum (Thermo Scientific) at 37°C in 5% CO2. T47D cells were cultured in the same conditions in RPMI1640 (RPMI1640, Thermo Scientific) supplemented with 1% penicillin/streptomycin (Thermo Scientific) and 10% heat inactivated fetal bovine serum. RON and MSP were stably expressed in MCF7 cells by retroviral infection (Liu et al., 2011). MCF7-RON/MSP cells were maintained in MCF7 medium supplemented with 0.2 µg/ml puromycin and 50 µg/ml hygromycin. MCF7 and T47D cell lines were infected with lentiviral constructs that direct the synthesis of shRNA (based on PLKO.1 but with a neo resistant cassette substituted for the puromycin resistant cassette) and selected by the addition of 1 mg/mL G418 48 hours later.

**In vivo metastasis**

All animal procedures were reviewed and approved by the University of Utah Institutional Animal Care and Use Committee. Breast cancer cells (2 X 10^6 cells in 25 µl matrigel) were implanted into the cleared right inguinal mammary fat pads of 3 week-old female NOD/SCID mice as previously described (Welm et al., 2007). An intrascapular estrogen pellet (DeRose et al 2011) was subcutaneously implanted into the mice at the same time. For bioluminescence imaging, mice were anesthetized and given 150 mg/kg of D-luciferin in PBS by i.p. injection. Five minutes after injection, mice were killed and the organs extracted and imaged *ex vivo* to detect metastasis foci. Bioluminescence was imaged with a charge-coupled device camera (IVIS; Xenogen). Bioluminescence images were obtained with a 15 cm field of view, binning (resolution) factor of 8, 1/f
stop, open filter, and an imaging time of 8 sec. We defined a metastatic event as any detectable luciferase signal above background and validated metastasis histologically. For drug treatment studies, female NOD-SCID mice were each orthotopically transplanted with a single primary breast cancer tumor fragment from patient lines HCI-003 and HCI-011 (DeRose et al., 2011). Mice bearing these established xenograft tumors (2-4 mm in diameter) received vehicle (40% Trappsol) or OSI-296 (200mg/kg) by oral gavage every other day until a tumor diameter of 1.2-1.5 cm was reached. Mice were killed and analyzed for metastasis histologically.

**DNA and RNA isolation procedures**

Genomic DNA for bisulfite modification was isolated using the DNeasy Blood Kit (Qiagen). Total RNA for gene expression analysis by microarray and quantitative RT-PCR was isolated with the RNeasy Kit (Qiagen) and tested for integrity on RNA 6000 NanoChips using an Agilent 2100 Bioanalyzer.

**DNA methyltransferase (Dnmt) activity assay.**

We quantified the total Dnmt activity levels using a colorimetric assay with the EpiQuik DNA Methyltransferase Activity Assay kit (Epigentek Group). In this assay, Dnmt transfers a methyl group from S-adenosylmethionine (AdoMet) to a cytosine in the DNA substrate. Methylated DNA was probed using an anti-5-methylcytosine antibody. The levels of methylated DNA, which were proportional to the enzymatic activity, were then colorimetrically quantified and expressed as absorbance units 450 nm•h−1•mg protein−1.

**Quantitative real-time RT-PCR**

Total RNA (1µg) was primed with oligo (dT) primers and converted into cDNA with RevertAid First Strand cDNA Synthesis Kit (Fermentas). SYBR green-based real time PCR was performed with a Roche LightCycler system, and the reaction mix contained 1X SYBR green master mix (SABiosciences) and 0.5 µM each of the forward and reverse primers in a volume of 25 µl. PCR cycling consisted of 95°C for 10 min, then 30
cycles of 95°C for 10 sec, 60°C for 30 sec, 72°C for 10 sec, followed by a melting-curve analysis. All PCR primers used are available on request.

Expression array hybridization and data analysis

Triplicate biological replicates of hybridization were performed for each cell line. The Agilent Quick Amp Labeling Kit is used to generate fluorescently labeled cRNA for one-color microarray hybridizations. Agilent RNA spike-in controls are combined with input total RNA samples (50 to 1000 ng). The polyadenylated fraction of the total RNA sample is primed with oligo dT/T7 RNA polymerase promoter oligonucleotide sequences and cDNA synthesis is accomplished through the addition of MMLV-RT. Following cDNA synthesis, T7 RNA polymerase and dye-labeled nucleotides are combined with the reaction mixture to simultaneously amplify the target material through the generation of cRNA and incorporate cyanine 3-CTP. Fluorescently labeled, cRNA molecules are purified from the reaction mixture using the Qiagen RNeasy mini kit. The concentration of the purified samples is determined using a NanoDrop ND-1000 spectrophotometer. Fluorescently labeled cRNA samples (825 ng each) were fragmented and combined with Agilent Hi-RPM Hybridization Buffer. Microarray hybridizations were performed using Agilent SureHyb Hybridization chambers. Hybridization chambers were loaded onto a rotisserie in an Agilent Hybridization oven and were incubated at 65°C for 17 hours with a rotational speed of 10 rpm. Following incubation, the microarray slide was washed for 1 minute each in Gene Expression Wash Buffer 1 (6X SSPE, 0.005% N-lauroylsarcosine; room temperature) and Gene Expression Wash Buffer (0.06X SSPE, 0.005% N-lauroylsarcosine; 31°C) for 1 minute each. Microarray slides were briefly dipped in a solution of acetonitrile and dried. Microarray slides were scanned in an Agilent Technologies G2505C Microarray Scanner at 5 µm resolution. The scanner performs detection of Cyanine-3 signal on the hybridized slide. TIF files generated from the scanned microarray image are loaded into Agilent Feature Extraction Software version 10.5. The software automatically positions a grid and finds the centroid positions of each feature on the microarray. This information is used to perform calculations that include feature intensities, background measurements and statistical analyses. Data generated by the software is recorded as
a tab-delimited text file. The text data were analyzed using the AgilentFilter and Genesifter software. Differential gene expression was evaluated using the t-test (p<0.01) and Benjamini and Hochberg correction. The threshold was set at 2-fold change for both upregulated and downregulated genes. Specific differentially expressed genes were confirmed by RT-qPCR.

**Whole-Genome Bisulfite Sequencing (WGBS)**

Library construction was performed using the Illumina TruSeq. DNA Sample Preparation Kit (cat# FC-121-2001, FC-121-2002) using the following methods: Genomic DNA (approximately 50 ng to 2 µg) was sheared to an average size range of 275 bp in a volume of 52.5 µl using a Covaris S2 Focused-ultrasonicator with the following settings: Intensity 5.0; Duty cycle 10%; 25 cycles per Burst 200; Treatment Time 60 seconds. Sheared DNA was converted to blunt-ended fragments with 5'-phosphates and 3'-hydroxyl groups using a combination of enzymes that perform fill-in reactions and exhibit exonuclease activity. Size selection of the blunt-ended DNA (200-450 bp average) was accomplished using bead-based methodologies. An A-base was added to the blunt ends as a means to prepare the fragments for adapter ligation and block concatamer formation during the ligation step. Adapters containing a T-base overhang were ligated to the A-tailed DNA fragments. Adapter-ligated molecules were purified by bead based methodologies and were bisulfite converted using the Qiagen EpiTect Bisulfite Kit (cat# 59104). Bisulfite treated DNA was PCR amplified for 12 cycles using Agilent Pfu Turbo Cx (Agilent cat# 600410) to enrich those fragments that have adapter molecules ligated to both ends. The concentration of the amplified library was measured using the Invitrogen Qubit dsDNA HS Assay (Q32851) and an aliquot of the library was resolved on an Agilent 2200 Tape Station using a D1K (cat# 5067-5361 and 5067-5362) or a High Sensitivity D1K (cat# 5067-5363 and 5067-5364) assay to define the size range. Libraries were adjusted to a concentration of approximately 10 nM and quantitative PCR was performed using the KapaBiosystems Kapa Library Quant Kit (cat# KK4824) to quantify adapter-ligated library molecules. The concentration was further adjusted following qPCR to prepare the library for Illumina sequence analysis.
Computational analytical method for WGBS

The software packages used in this analysis are open source and available from the USeq (http://useq.sourceforge.net/usageBisSeq.html) project website (Nix et al., 2010). SAM alignments were generated from Illumina Fastq files aligned to the human hg19 genome using Novocraft’s novoalign aligner (http://www.novocraft.com) in bisulfite mode with the following parameters: -r Random –t 240 –h 120 –b 2 –p bisulfite. An in silico chrLambda sequence was used to align the fully methylated lambda sequence that was spiked into the samples in order to measure the bisulfite conversion efficiency and displayed very low non-conversion/sequencing error rates of 0.00348 (MCF7) and 0.00245 (MCF7-RON/MSP). These rates were used for scoring individual cytosines for significant methylation with a binomial test. A chrPhiX sequence was added to estimate the read quality and a chrAdapter sequence containing all permutations of the adapter sequences was used to remove these artifacts from the data. Differentially-methylated regions (DMRs) were determined by comparing methylation levels in each cell line using a sliding window of 150 base pairs. Overlapping windows passing two thresholds, a Benjamini–Hochberg corrected a chi-square test of independence FDR of < 0.01, and fold difference in methylation of > 2, were merged into continuous DMRs and assigned the best window scores (See http://useq.sourceforge.net/usageBisSeq.html for details). Next, DMRs were used to score mean fraction methylation in every sample.

NovoalignBisulfiteParser(http://useq.sourceforge.net/cmdLnMenus.html#NovoalignBisulfiteParser) was used to parse the text based novoalignments into four binary “PointData” sets containing the number of observed converted Cs (Ts - non methylCs) and non-converted Cs (methylCs) at each reference C sequenced in the genome for both the plus and minus strands. The PointData was then parsed into mCG context using the ParsePointDataContexts (http://useq.sourceforge.net/cmdLnMenus.html#ParsePointDataContexts) application.

BisStat (http://useq.sourceforge.net/cmdLnMenus.html#BisStat) application was used to calculate per base fraction methylation scores for bases with five or more reads from both strands and generate tracks for visualization in IGB. BisStat also calculated fraction methylation for 500bp windows containing a minimum of 5CpGs.
Mean fraction methylation at each DMR was scored using the ScoreMethylatedRegions (http://useq.sourceforge.net/cmdLnMenus.html#ScoreMethylatedRegions) application. Only regions with a minimum of 5Cs and a minimum read coverage of 5 in all samples were used in further analysis.

Differentially methylated regions were identified using the BisSeq (http://useq.sourceforge.net/cmdLnMenus.html#BisSeq) application. Overlapping 150bp windows that met or exceeded an FDR of 0.01 and a 2-fold difference were merged and included in a spreadsheet output of differentially methylated regions. This list was further filtered in Excel for including regions with sample having > 0.25 fraction methylation difference.

Human annotation and genomic sequence (Feb. 2009, GRCh37/hg19) were obtained from the UCSC Genome Bioinformatic website.

**RNA-sequencing**

Alignments were generated from Illumina Fastq files to the hg19 human genome with all known and theoretical splice junctions using Novocraft’s novoalign aligner. NovoalignParser (http://useq.sourceforge.net/cmdLnMenus.html#NovoalignParser) application was used to parse the alignment files into binary point data. DefineRegionScanSeqs (http://useq.sourceforge.net/cmdLnMenus.html#DefinedRegionScanSeqs) application was used to define differentially expressed genes.

Sequencing was performed with 3 biological replicates for each cell line. Sequencing reads were aligned to the hg19 human genome with all known and theoretical splice junctions using Novocraft’s novoalign aligner.

**Pathway and functional analyses of differentially methylated and differentially expressed genes**

Differentially expressed genes (p < 0.05) were functionally annotated and classified by using the functional annotation tool database for annotation, visualization, and integrated discovery (DAVID), which provides integrated annotation and analysis of genome-scale datasets derived from high-throughput technologies.
Human breast tumor gene expression data pre-processing

We imported expression data from the referenced studies. Raw data was downloaded from the NCBI GEO website or from websites indicated in the original publications and processed as previously described (Segal et al., 2004). Briefly, we first log2 transformed the expression values and then calculated the mean expression level for each gene across all samples in a given dataset. These mean values were subtracted from all data points, such that expression was represented relative to the mean of each gene, negative values representing below-mean expression and vice versa. To construct the breast cancer compendium, we first normalized each of the five-included studies independently, and then concatenated these sets. In cases where the same individuals were included in more than one of the eight studies comprising the compendium, such redundancy was eliminated so that each individual was included only once in the compendium. Specifically, individuals of the Uppsala cohort analyzed in the Miller set were eliminated from the Desmedt set.

Analysis of gene-set enrichment patterns

To identify gene-set enrichment patterns we used methods described previously (Segal et al., 2004), applied using the Genomica software. For each sample (array), we first scored the genes whose expression was at least twofold above or below the mean expression level. We then assessed the fraction of over- or under-expressed genes that belong to each tested gene set, calculating a $P$ value according to the hypergeometric distribution. This was repeated for every sample, using a threshold of $P < 0.05$ for significant enrichment. To compare enrichment patterns across sample groups, we included clinical annotations for each individual sample (such as tumor grade, size and ER status) derived from the original publications. For all samples showing enrichment for the signature, we calculated the fraction of samples that possessed each annotation, and assigned a $P$ value according to the hypergeometric distribution. We used a more stringent threshold, $P < 0.01$, for this calculation. Heat maps showing gene-set enrichments in individual samples include only those samples enriched for the signature.