Supplementary methods:

Cellular growth assay and cell cycle analysis

To determine MRK-003 IC50 values, cell lines were plated in triplicate in 96-well plates at 3 x 10^3 cells/well (except for TALL1 and REC-1 cells, which were plated at 1 x 10^4 cells/well) and treated with increasing concentrations of MRK-003 or vehicle (DMSO). Cell numbers were determined on day 7 using the Cell Titer-Glo Luminescent Cell Viability assay (Promega, Madison, WI) per the manufacturer's instructions. Luminescence was measured using a SpectraMax plate reader (Molecular Devices, Sunnyvale, CA). We calculated relative proliferation rates in control (DMSO) or MRK-003 treated cells by evaluating and normalizing doubling time for each cell line. We first determined exponential growth rates (μ) in drug treated or control (μmax) treatment conditions and computed relative (normalized) proliferation rates by using the following equation:

$$e^{1.4(\frac{\mu}{\mu_{\max}}-1)}$$

This method makes it possible to compare drug effects on cell lines with significantly different doubling times to be compared.

For cell cycle analysis, 2 x 10^5 cells were plated in triplicate in 6-well plates and treated with 1μM MRK-003 or vehicle (DMSO) for 5 days. Cells were stained using the Guava cell cycle kit according to manufacturer's protocol. Stained cells were analyzed on a FACSCalibur flow cytometer (BD Biosciences, San Diego, CA), and collected data sets were analyzed using FlowJo software (Tree Star Inc., Ashland, OR).

CD24/CD44 Expression

Cells were plated in 6-well plates in triplicate and treated with 1μM MRK-003 or vehicle (DMSO) for five days. Cells were washed with ice-cold phosphate-buffered saline (PBS), detached from wells with 0.05% trypsin/0.025% EDTA (Cellgro, Manassas, VA), and washed in HBSS (Cellgro, Manassas, VA) containing 1% BSA. Fluorochrome-conjugated monoclonal antibodies from BD Biosciences (San Diego, CA) against human CD44 (FITC, #555478) and CD24 (PE, #555428) or their respective isotype controls were added to the cell suspension at concentrations recommended by the manufacturer and incubated at 4°C for 45 min. Cells were then washed, resuspended in HBSS containing 1% BSA, and analyzed by flow cytomtery using FlowJo software (Tree Star Inc., Ashland, OR).

ALDEFLUORTM stem and progenitor cell identification

5 x 10^5 MB-157 or NCIH226 cells were plated into 6-well plates in their recommended growth medium and treated with 500nM MRK003 or DMSO as vehicle control in duplicates. Cells were incubated for 72hr, harvested, and stained with the ALDEFLUORTM kit to detect cells with high enzymatic ALDH activity (Stem Cell Technologies) according to manufacturer's protocol and samples were analyzed on FACSCANTO II (BD Biosciences).

Luciferase reporter assay

Construction of human NOTCH1 expression plasmids, lipofection of U2OS cells, and Notch luciferase reporter gene assays were performed as described (5). In brief, cells in 24-well dishes were cotransfected in triplicate with 10 ng of various pcDNA3-NOTCH1 expression constructs, a Notch-sensitive firefly luciferase reporter gene (15), and an internal control *Renilla* luciferase plasmid (Promega, Madison, WI). Total introduced DNA was kept constant by adding empty pcDNA3 plasmid. Normalized firefly luciferase activities were measured in whole-cell extracts prepared 44 to 48 hr after transfection using the Dual Luciferase kit (Promega, Madison, WI) and luminometer configured for dual assays (Turner Systems). Cells were treated post-transfection with MRK-003 at 1μM or vehicle control (0.01% DMSO). Site-directed mutagenesis

of the HD domain or PEST domain was conducted using the QuikChange kit (Stratagene, La Jolla, CA). Mutagenic PCR primers were designed using the Stratagene web tool.

Whole exome sequencing and exon imbalance analysis

Whole exome sequencing (WES) or targeted exome sequencing (TES) on 4000 genes was done at BGI using standard Illumina and Agilent Sureselect V4 protocols. Targeted exome sequencing was done using the same set of capture probes as in Agilent Sureselect V4. Paired-end sequencing (100 base pairs) with 200 base pair insert size was performed using the Illumina Hiseq 2000 platform. Analysis of sequencing reads revealed an average coverage of 100x and >20x coverage for 80% of target bases. Sixty-six and 154 triple negative breast cancers were subjected to WES and TES, respectively, while 608 cell lines were sequenced using TES. The GATK1 toolbox (19) was used to process sequencing reads, call variants, and summarize exon coverage using the HG19 reference genome sequence and default parameters.

Analysis of exon imbalance was done as follows. Exon coverage for a gene of interest, for example *NOTCH1*, was scaled by the average exons coverage of the gene to reduce variation due to amount of sequencing and global alternations, such that scaled exon coverage

$$\boldsymbol{C_i} = \frac{N_i * M}{\sum_{i=1:M} N_i}$$ (equation A)

where Ni is coverage of exon i; M is the number of exons in the gene. Adjustments for capture efficiency of specific exon were made by assuming that capturing efficiency is the same for the exon i in all samples within a single batch of Sureselect reagents and that coverage is proportional to true abundance multiply by capturing efficiency.

Therefore, coverage μ_{ji} of exon *i* in sample *j* can be expressed as

$$\mu_{ji} = \frac{C_{ji} * K}{\sum_{j=1:K} C_{ji}}$$ (equation B)

where K is number of samples, C_{ji} is coverage of exon *i* in sample *j* calculated as in equation A. To detect known intragenic deletions, average coverage within deleted and non-deleted regions is then compared using one-way ANOVA (Matlab or SAS) to test for significant (p-value < 10^{-5}) differences in exon read coverage. Statistical tests were done on log transformed coverage values. These analyses were done on TES data sets from 608 cell lines and WES data sets from 66 triple negative breast cancers. In doing so, we identified two critical processing constraints. Sequencing of cell lines and all WES were carried out with the same batch of Sureselect capture probes; however, the TES carried out on 4000 primary samples showed significant interbatch variation, which has a limited impact on mutation calling but confounded quantitative analysis of coverage data. Secondly, we noted that whole genome amplification introduces exon-specific biases that cannot be adjusted by a simple scaling; therefore, we limited the analysis to sequencing data obtained on DNAs that were not pre-amplified. A summary of the NOTCH gene coverage data from human tumors are provided (See Supplemental Material; NOTCH gene coverage; Cell line sequencing data can be accessed at SRP044150).

RNA-seq analysis

Total RNA was prepared using Trizol extraction and the RNeasy Mini kit (Qiagen). Integrity and concentration of total RNA was analyzed by Agilent bioanalyzer and RNA was then enriched for mRNA on oligo-dT beads. RNA-seq libraries were prepared with TruSeq RNA Sample Prep V2 kit (Illumina RS-122-2001) as per the instructions. Sequencing was performed on the Hiseq2000 platform. RNA-seq reads were aligned to reference human genome hg19 using TopHat (34) and the expression level of each gene was calculated using Cufflinks (35). Differentially expressed genes were identified using Cuffdiff (36), with the thresholds for differential expression set as p-value < 0.05 and fold-change > 2. Expression levels of NOTCH pathway

genes in human tumors are provided (See supplemental material: human breast tumor Notch GE pathway)