Stem Cell Reports, Volume *10*

Supplemental Information

HN1L Promotes Triple-Negative Breast Cancer Stem Cells through *LEPR-STAT3* Pathway

Yi Liu, Dong Soon Choi, Jianting Sheng, Joe E. Ensor, Diana Hwang Liang, Cristian Rodriguez-Aguayo, Amanda Polley, Steve Benz, Olivier Elemento, Akanksha Verma, Yang Cong, Helen Wong, Wei Qian, Zheng Li, Sergio Granados-Principal, Gabriel Lopez-Berestein, Melissa D. Landis, Roberto R. Rosato, Bhuvanesh Dave, Stephen Wong, Dario Marchetti, Anil K. Sood, and Jenny C. Chang

Supplementary Materials

Table S1: List of 13 CSC genes

Gene Symbols	Gene name
RPL39	ribosomal protein L39
MLF2	myeloid leukemia factor 2
HN1L	hematological and neurological expressed 1-like
MAGI3	membrane associated guanylate kinase, WW and PDZ domain containing 3
GNAZ	G protein subunit alpha z
HMGXB3	HMG-box containing 3
ZBTB16	zinc finger and BTB domain containing 16
KIF16B	kinesin family member 16B
TRBV19	T cell receptor beta variable 19
MARVELD2	MARVEL domain containing 2
MAP7	microtubule associated protein 7
SHB	Src homology 2 domain containing adaptor protein B
PLCH1	phospholipase C eta 1

Table S2: List of mutations identified using RNA-seq in patients with lung metastasis.

Gene	AA. variant	Chr.	Base Position	refBase	altBase
HN1L	p.P20L	16	1735454	C	T
HN1L	p.A106V	16	1741967	C	T

Table S3: siRNA sequences targeting the 11 BCSC genes.

Name	Sequence (Sense)	Sequence (Antisense)
HN1L	CGCCUGUAUUUGGAAGAUUUAA	UUAAAUCUUCCAAAUACAGGCA
MAGI3	CCCUUCUGAGGUCUACCUGAAA	UUUCAGGUAGACCUCAGAAGGA
SHB	ACCUUCUUUGCUGGCUUUAUUA	UAAUAAAGCCAGCAAAGAAGGG
KIF16B	CGGCUGAGAAGUUUCAGAUAUU	AAUAUCUGAAACUUCUCAGCCU
GNAZ	CGCUAAGUGUCUUGGUAUUUAA	UUAAAUACCAAGACACUUAGCU
PLCH1	CGCUCAGUACCUGAAAGGAAUA	UAUUCCUUUCAGGUACUGAGCA
ZBTB16	ACCCUUCAGUCUCCACUUCAUU	AAUGAAGUGGAGACUGAAGGGC
MAP7	AUCUUACAUAAUGUAUUUAUAA	UUAUAAAUACAUUAUGUAAGAG
MARVELD2	AUGCUACUAUCCGUUAUUUAAU	AUUAAAUAACGGAUAGUAGCAG
TRBV19	AACCCUGAGUUGUGAACAGAAU	AUUCUGUUCACAACUCAGGGUC
HMGXB3	GCCUGUCUAUGUGGUAGAU	AUCUACCACAUAGACAGGC
Scrambled	AUCUCGCUUGGGCGAGAGUAAG	CUUACUCUCGCCCAAGCGAGAG

Supplementary Experimental Procedures

Screening BCSC candidate genes.

MDA-MB-231 cells were screened for mammosphere forming ability against 11 genes from the 13 genes (Table S1) using each specific siRNA (Table S3). Cells (200,000 cells/well) in an ultralow-attachment 24-well plate were transfected in 6 replicates using siPORT (Invitrogen) according to the manufacturer's instruction. The final concentration of each siRNA was 30nM per well. Methyl cellulose (1%) was used to for the mammosphere formation assay. For the secondary mamm0sphere formation, 5000 cells were used for each well in 6 replicates per gene. We repeated the screening twice with similar gene silencing effects on the mammosphere formation of MDA-MB-231

siRNA and shRNA knockdown.

HN1L shRNA sequences are: 5'-CGCCTGTATTTGGAAGATTTAA-3' and 3'-TTAAATCTTCCAAATACAGGCA-5'. siRNA was tranfected *in vitro* using Lipofectamine 2000 (Invitrogen) or TransIt-TKO (Mirus Bio, Madison, WI). The optimal *HN1L* siRNA sequences delivered were obtained from Ambion; *HN1L* siRNA1 :5'-CCAAGGAUCAUGUUUUCUU-3' and 3'- AAGAAAACAUGAUCCUUGG-5', HN1L siRNA2: 5'-CCUCAGAACAUACCCAAGA-3' and 3'-UCUUGGGUAUGUUCUGAGG-5'. Scrambled siRNA sequences are 5'- CGUGAACACGCAACUAAGG-3' and 3'-CCUUAGUUGCGUGUUCACG-5'. *LEPR* siRNA were also purchased from Ambion, and sequences are 5'-GAGUGAUCAUGUUAGCAAA-3' and 3'-UUUGCUAACAUGAUCACUC-5'. For *in vivo* delivery, we used *HN1L* siRNA1. For testing gene silencing effects on MSFE, cells were treated with siRNAs (25 to 50nM) for 48 hours in advance. When plated into the assay

plates, cells were treated again with 25nM of siRNAs. For the secondary mammosphere assay, no siRNA transfection was used.

Fluorescence-activated cell sorting (FACS) analysis.

The gating was performed as previously described (Li, Lewis et al. 2008, Creighton, Li et al. 2009, Choi, Blanco et al. 2014, Dave, Granados-Principal et al. 2014). Briefly, side scatter and forward scatter were used to eliminate debris and doublets, and Sytox-Blue staining was used to differentiate live and dead cells. The remaining tumor cells were further analyzed using antibodies, CD44-APC and CD24-FITC or -PE-Cy7 or by measuring aldehyde dehydrogenase activity. For In Vivo tumors, tumor cells negative for H2kD were analyzed for CSCs. Data analysis was performed with FACS Diva (BD Biosciences, San Jose, C.A., USA). $CD44^{+}/CD24^{-/low}$ (BCSC) and other cells (non-BCSC) were sorted as previously described (Creighton, Li et al. 2009). Both flow analysis and sorting were performed at the Houston Methodist Research Institute Flow Cytometry Core, using BD FACS Fortessa for analysis and BD FACS Aria II for cell sorting. All in vitro experiments were repeated three times with 3 replicates. The average values of a single experiment were shown in the figure.

Mammosphere Assay

Cells treated with either siRNA or plasmid for different durations were trypsinized, collected and counted for mammosphere assay. Mammospheres were growing in MammoCult medium with 0.5% methylcellulose. Cells were seeded at 1,000 to 5,000 cells per well in 500µl Mammocult medium in 24-well ultra-low attachment plate (Corning, Lowell, MA). Every three days, 100 µl of fresh Mammocult medium was added into the well. After 5-14 days, depending on the cells,

mammospheres formed and were counted with GelCount (Oxford Optronix, Oxford, UK) and its bundled software. For the secondary mammosphere assay, cells were collected with 0.05% trypsin for 5 minutes followed by neutralization with 10% FBS. The cells were then re-suspended in MammoCult medium and seeded at 1,000 cells per well. Secondary mammosphere formed in SUM159 cells were counted on day 10 and in MDA-MB-231 cells were counted on day 14. Mammosphere assays were repeated with 6-12 replicates for each treatment group.

Western blot

Cells treated with siRNA or plasmids with different durations were lysed in a lysis buffer (1.5% Triton X-100 and 10% glycerol in DPBS) containing a proteinase and phosphatase inhibitor cocktail (Thermo Scientific Pierce Protein Biology, Rockford, IL). 30-500 μg of protein extracts were loaded for western blot.

Cell migration and invasion assays.

50,000 cells were seeded in each well after overnight starvation in serum-free medium, and migration was measured 6 hours after adding DMEM+10%FBS as a chemoattractant. For the invasion assay, the Boyden Chamber was coated with 0.1x BME solution and invasion was measured after 24 hours. In the 3D invasion assay, 5,000 cells were plated in each well and cells invaded for 4 days in invasion matrix. Data was quantified using the Image J Program.

***In vivo* Experiments.**

MDA-MB-231 cell line tumor model: $3x10^6$ MDA-MB-231 cells injected into the mammary fat pad of SCID-Beige mice were grown to 150-300 mm^3. Then, mice were randomized into 5 groups: (*i*)scrambled siRNA-DOPC, (*ii*)*HN1L* siRNA-DOPC, (*iii*)docetaxel+PBS, (*iv*)docetaxel +scrambled siRNA-DOPC, (*v*)docetaxel+*HN1L* siRNA-DOPC. Groups *(i), (ii), (iv), and (v)* (n=10) were treated with 5 µg/mouse DOPC nanoliposomal siRNA intraperitoneal (IP) injection twice a week for 3 weeks. Mice were sacrificed on day 21. Tumors were harvested and analyzed for BCSCs using FACS, MSFE and limiting dilution assays as previously described (Schott, Landis et al. 2013). Limiting dilution assays were analyzed by Extreme Limiting Dilution Analysis (ELDA) (Hu and Smyth 2009). Groups (*iii*) - (*v*) (n=13) were given 20 mg/kg docetaxel IP injection every 2 weeks for a total of 6 weeks. Respective 5 µg/mouse DOPC-liposomal siRNA was also given twice a week for a total of 6 weeks. Treatment ended on day 42, but mice were continuously monitored without any treatment until day 58.

In order to investigate *HN1L* siRNA treatment effects on lung metastasis, $3x10^6$ MDA-MB-231 cells transfected with luciferase were injected into the mammary fat pad of each SCID-Beige mouse, and the mice were randomized into 4 groups (n=10) when the primary tumor reached 150-300mm^3: (*i*)scrambled siRNA-DOPC, (*ii*)*HN1L* siRNA-DOPC, (*iii*)docetaxel+scrambled siRNA-DOPC, (*iv*)docetaxel+*HN1L* siRNA-DOPC. Treatment schedule was the same as above. At the end of the study, mice were sacrificed and lungs were harvested and imaged as previously described (Choi, Blanco et al. 2014).

SUM159 cell line tumor model: $1x10^6$ SUM159 cells injected into the mammary fat pad of SCID-Beige mice were grown to ~200 mm^3. Then, mice were randomized into 2 groups: (*i*)scrambled siRNA-DOPC, (*ii*)*HN1L* siRNA-DOPC. Groups *(i) and (ii),* (n=9), were treated with 5 µg/mouse DOPC nanoliposomal siRNA intraperitoneal (IP) injection twice a week for 3 weeks. Mice were sacrificed on day 21. Tumors were harvested and analyzed for BCSCs using FACS, and MSFE analysis in vitro (Schott, Landis et al. 2013).

BCM2665 tumor model: In docetaxel-resistant BCM2665 PDX xenografts, tumors were transplanted into the mammary fat pad of SCID-Beige mice. Mice were randomized into 5 groups when tumor volume reached 150-200mm^3: (*i*)scrambled siRNA-DOPC, (*ii*)*HN1L* siRNA-DOPC, (*iii*)docetaxel+PBS, (*iv*)docetaxel+scrambled siRNA-DOPC, (*v*)docetaxel+*HN1L* siRNA-DOPC. Groups (*i*) and (*ii*) (n=10) were treated with 5 µg/mouse DOPC nanoliposomal siRNA IP injection twice a week for 3 weeks. Mice were sacrificed on day 21. Tumors were harvested and analyzed for BCSC using FACS and MSFE. Groups (*iii*) - (*v*) (n=15) were given 33.3 mg/kg docetaxel IP injection on day 1, 8 and 22. DOPC-liposomal siRNA (5 µg/mouse) was given twice a week for 5 weeks. At the end of the study, mice were sacrificed and tumors were harvested for limiting dilution assays as previously described (Chen, Iliopoulos et al. 2014). Tumor pieces containing $6x10^5$, $3x10^5$ or $1x10^5$ cells were transplanted with basal membrane extract into mammary glands on both sides of nude mice. Tumor incidence was reported at 4 weeks after transplantation on the counts of established tumors (equal or larger than 50mm^3). Limiting dilution assays were analyzed by Extreme Limiting Dilution Analysis (ELDA) (Hu and Smyth 2009). DOPC nanoliposomal siRNA was prepared as previously described (Landen, Chavez-Reyes et al. 2005, Tanaka, Mangala et al. 2010).

ChIP and ChIP-seq.

ChIP DNA was prepared into libraries and sequenced by the Epigenomics Core of Weill Cornell Medical College using SR50 lane. Antibodies used are anti-FLAG (Sigma, #3165) and anti-mouse IgG (EMD Millipore, #12-371). ChIP-Seq analysis began with mapping the sequenced reads to the genome. We utilized the Burrows-Wheeler Aligner (BWA) MEM algorithm to align the sequence reads against the human genome GRCh37/hg19 Assembly (Li 2013). We next used the Hypergoemetric Optimization of Motif Enrichment (HOMER) suite of tools to find and annotate peaks, and identify enriched motifs. First, we utilized HOMER's findPeaks tool to perform peak calling. Peak calling identifies the regions in the genome where a significant number of sequencing reads are found. These peaks were visualized in bigWig track file format in the UCSC genome browser. This UCSC-accepted file was created by first running HOMER's makeUCSCfile tool followed by UCSC's bedGraphToBigWig script. Next, HOMER's annotatePeaks.pl program was used to associate peaks with nearby genes. From here we compared the list of nearby genes with the HN1L and BCSC gene signatures, as well as CSC TF's. The final stage of the ChIP-Seq analysis involved using HOMER's findMotifsGenome.pl program to find enriched motifs, and ultimately, consensus sequences in the ChIP-Seq peaks. The results include a ranked list of de novo and known motifs. The "best" motifs are those with p-values significantly smaller than 1e-50. The p-value in this application is a measure of the ratio of target peaks containing the motif to background peaks containing the motif (Heinz, Benner et al. 2010). Using the chosen motif file from HOMER, we used the R program "seqLogo" to create a visually informative motif logo (Bindewald, Schneider et al. 2006). To confirm the peaks found by HOMER's findPeaks tool, and

to check for additional peaks, we used another peak-finding algorithm, the Model-based Analysis for ChIP-Seq (MACS) program, to find peaks in our sample. Following this, we used the online tool PAVIS to annotate these peaks (Zhang, Liu et al. 2008, Huang, Loganantharaj et al. 2013). The GEO accession number for this ChIP-seq data is GSE105446

Gene expression microarray analysis.

Microarrays were performed, using Affymetrix genechip U133plus 2.0. Normalization and evaluation of the data was performed as previously described (Dave, Granados-Principal et al. 2014). Differentially expressed genes were identified as the *HN1L* knockdown gene signature with cutoff *p* value less than 0.05 and fold-change greater than 1.5. Further functional and pathway analyses were done by Ingenuity Pathway Analysis (IPA) tools. Gene Set Enrichment Analysis (GSEA) was used to determine the alternation of the *JAK-STAT* pathway in *HN1L* siRNA-treated tumors compared with scrambled siRNA-treated tumors.

MicroRNA array analysis.

miRNA expression values obtained from affymetrix – miRNA array 4.0 was used to perform differential expression analysis across the 10 tumor samples. Bioinformatic analysis was performed to study changes in miRNA profiles between scrambled siRNA-treated mice and *HN1L* siRNA-treated mice tumor samples. Expression values were processed and normalized using the *affy* library in R statistical software. Only human probes were isolated, from the multispecies miRNA profiling by the miRNA array 4.0, for the purpose of the analysis. Given the 6631 human probes, supervised expression based clustering was performed to remove outlier samples present in the sample groups. Differential expression analysis was then performed using *limma* package

in R. The most significantly differentially expressed miRNAs were selected based on a cut off log fold change > 1 and FDR < 0.2. The GEO SuperSeries accession number for this microRNA array data is GSE106200.

Real-time PCR analysis.

All primers used are listed below and were designed using Primer3 and synthesized by Sigma. *STAT3* forward 5'-CCAAGATAGCGCCACTGC-3'; reverse 5'-ACATGTATCCTGTTAATTGACTTGC-3'. *FGFR2* forward 5'-TGCACTATTCACCCAACTTTCT-3'; reverse 5'-AGGAATGTGTTTGTGGCCAC-3'. All TaqMan Gene Expression Assays were purchased from Invitrogen; *STAT3* (Assay ID: Hs01051722_s1), *LEPR* (Assay ID: Hs00174492_m1), *LEP* (Assay ID: Hs00174877_m1), and *HN1L* (Assay ID: Hs00375909_m1). Eukaryotic 18S rRNA Endogenous Control (Invitrogen) was used as an internal control. Gene expression was analyzed using a standard curve for each gene as described previously (Choi, Blanco et al. 2014, Dave, Granados-Principal et al. 2014), PCR were repeated two to three times with three technical repeats. Representative results are shown in the Fig. with standard deviation.

Rescue experiment by constitutively active *STAT3*.

EF.STAT3DN.Ubc.GFP (Addgene plasmid # 24984) and EF.STAT3C.Ubc.GFP (Addgene plasmid # 24983) were gifts from Linzhao Cheng. SUM159 cells ($2x10^6$ cells) were co-transfected with these plasmids and siRNAs (50nM) using Lipofectamine 2000 (Invitrogen) according to the manufacturer's manual. After 48 hours later, cancer cells were analyzed for $CD44^+/CD24^{low/-}$ breast cancer stem cells by FACS or for MSFE as described earlier. For FACS analysis, the same

gating strategy was used as described in the Fluorescence-activated cell sorting (FACS) analysis section with an additional EGFP positive gate. The FACS data were the average values of three biological repeats. The mammosphere experiments were repeated twice with 6 replicates. The average values of one experiment were presented in the figure.

MicroRNA array analysis.

MicroRNA arrays were performed on 10 snap-frozen tumor samples from scrambled siRNA-treated and *HN1L*-siRNA treated BCM2665 xenografts (5 samples from each treatment group), as described in the Supplemental Experimental Procedures.

Supplementary Figure Legends

Fig. S1. *HN1L* is one of the top candidate genes critical for breast cancer mammosphere forming ability. MDA-MB-231 cells were transfected with each specific siRNAs (30nM) prior to the primary mammosphere assay. Scrambled siRNA was used as the control. After the primary mammosphere formation, the cancer cells (5000 cells/well) were re-seeded for the secondary mammosphere. Each MS sample included 6 replicants, and experiments were repeated 3 times independently.

Fig. S2. Non-TNBC patients show no survival correlation with the expression levels of HN1L. (A and C) The expression *HN1L* is not correlated with the overall survival of patients with non-TNBC. TNBC patients with the higher expression level of *HN1L* tend to show the shorter overall survival than those with lower *HN1L* expression (B). (A) TCGA; (B) and (C): Curtis.

Fig. S3. *HN1L* silencing has anti-CSC effects.

(A) Western blot showing *HN1L* knockdown by two different *HN1L* siRNAs in TNBC cell lines, SUM159, MDA-MB-231, and MDA-MB-468. (B) *HN1L* siRNA silencing decreased $CD44^{+}/CD24^{-/low}$ population by flow cytometry analysis in SUM159, MDA-MB-231, and MDA-MB-468 cell lines. (C) *HN1L* siRNA silencing reduced primary and secondary MSFE in SUM159 and MDA-MB-231 cell lines. Each experimental group had 6 replicates, and all experiments were repeated three times. (D) Immunohistochemistry staining of tumor samples using HN1L antibody to show target engagement. (E) A statistically non-significant decrease in ALDF+ cells was apparent in MDA-MB-231 cells with *HN1L* knockdown. (F-H) SUM159 cells were injected into SCID Beige mice and allowed tumors to grow to ~200mm3 before were randomized into different groups. These mice were injected with the respective DOPC liposomal

siRNA by i.p. injection at 5μg/injection twice a week for 3 weeks. Two treatment groups (n=10): scrambled siRNA, *HN1L* siRNA. (F) Tumor volume was measured. Mice were sacrificed 3 weeks later, and tumors were collected and processed for flow analysis of CD44+/CD24-/low cells (G), and ALDF+ cells (H). (*p<0.05, **p<0.01, ***p<0.001). All experiments were repeated three times with three technical repeats. For the purpose of publication, representative data of a repeat is presented. For the multiple comparison, Tukey's multiple comparison tests for one-way ANOVA was performed with Graphpad Prism 5.0 (Graphpad Software Inc., La Jolla, CA, USA).

Fig. S4. *HN1L* silencing effect *in vivo*. (A) Western blot data assessing HN1L, LEPR and STAT3 signaling in BCM2665 xenograft tumors under different treatment conditions; siScr: scrambled siRNA. β-actin serves as a loading control. (B) Tumor volume of mice from 3 treatment groups (n=13): docetaxel+PBS, docetaxel+scrambled siRNA, docetaxel+*HN1L* siRNA. Docetaxel was given every 2 weeks for 3 cycles. Liposomal siRNA was delivered twice a week for 6 weeks. Tumor volumes were still monitored for 2 more weeks after treatment was stopped. (C) Effects of *HN1L* knockdown on metastasis *in vivo*. *Ex vivo* imaging on lungs from each group on day 21 were presented on the upper panel when *HN1L* siRNA was used as single agent. *Ex vivo* imaging on lungs from each group on day 58 were shown on the lower panel when *HN1L* siRNA was combined with Docetaxel. (*p<0.05)

Fig. S5. HN1L regulates expression of STAT3-LEPR signaling pathway (A) Overexpression of HN1L cells upregulated phosphorylation of STAT3 and LEPR protein expression in SUM159. (B) Conversely, *HN1L* gene silencing reduced the mRNA expression of *STAT3*. (C) Overexpression of *HN1L* increases the mRNA levels of *STAT3* and *LEPR* in MDA-MB-231 and

SUM159. (D-E) Co-transfection with *HN1L* siRNA and a plasmid with constitutively active *STAT3-GFP* gene rescues the anti-CSC effects of *HN1L* silencing. For the mammosphere formation assay, each experimental group had 6 replicates, and all experiments were repeated three times. All experiments were repeated three times with three technical repeats. For the purpose of publication, representative data of a repeat is presented. For the multiple comparison, Tukey's multiple comparison tests for one-way ANOVA was performed with Graphpad Prism 5.0 (Graphpad Software Inc., La Jolla, CA, USA).

Fig. S6. ChIP peaks in *LEPR* called by MACS. (A) Identification of *HN1L* motif in HN1L overexpressed SUM159 ChIP-seq data. Anti-FLAG antibody was used for ChIP. Matrices predicted by HOMER Motif Analysis. (B)Visualized peaks shown in both input and anti-FLAG samples. The peak found by MACS and validated by QPCR (C) was indicated by the red arrow. (D) Pathway analysis by STRING 10 revealed top pathways regulated by the overlapped genes from Fig. 7.

Fig. S7. HN1L regulates *LEPR* and *miR150* pathways which converge into *STAT3*. (A) Top most differentially expressed miRNAs between scrambled siRNA-treated mice tumors and *HN1L* siRNA-treated tumors in BCM2665 xenografts (n=4 for each treatment arm). miRNA were selected based on a cut-off log fold change >1 and FDR <0.2. (B) Table showing the ID, log fold change and p-value of the top 4 most upregulated (positive value in logFC) or downregulated (negative value) miRNA. (C) Validation of reduced miR-150 in *HN1L* siRNA-treated tumors by qPCR. (D) Protein-protein interactions of known miR-150 targets with LEPR-STAT3 are presented by STRING 10. Thicker lines indicate the stronger confidence of associations.

References

Bindewald, E., T. D. Schneider and B. A. Shapiro (2006). "CorreLogo: an online server for 3D sequence logos of RNA and DNA alignments." Nucleic Acids Res **34**(Web Server issue): W405-411.

Chen, X., D. Iliopoulos, Q. Zhang, Q. Tang, M. B. Greenblatt, M. Hatziapostolou, E. Lim, W. L. Tam, M. Ni, Y. Chen, J. Mai, H. Shen, D. Z. Hu, S. Adoro, B. Hu, M. Song, C. Tan, M. D. Landis, M. Ferrari, S. J. Shin, M. Brown, J. C. Chang, X. S. Liu and L. H. Glimcher (2014). "XBP1 promotes triple-negative breast cancer by controlling the HIF1alpha pathway." Nature **508**(7494): 103-107.

Choi, D. S., E. Blanco, Y. S. Kim, A. A. Rodriguez, H. Zhao, T. H. Huang, C. L. Chen, G. Jin, M. D. Landis, L. A. Burey, W. Qian, S. M. Granados, B. Dave, H. H. Wong, M. Ferrari, S. T. Wong and J. C. Chang (2014). "Chloroquine eliminates cancer stem cells through deregulation of Jak2 and DNMT1." Stem cells **32**(9): 2309-2323.

Creighton, C. J., X. Li, M. Landis, J. M. Dixon, V. M. Neumeister, A. Sjolund, D. L. Rimm, H. Wong, A. Rodriguez, J. I. Herschkowitz, C. Fan, X. Zhang, X. He, A. Pavlick, M. C. Gutierrez, L. Renshaw, A. A. Larionov, D. Faratian, S. G. Hilsenbeck, C. M. Perou, M. T. Lewis, J. M. Rosen and J. C. Chang (2009). "Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features." Proc Natl Acad Sci U S A **106**(33): 13820-13825.

Dave, B., S. Granados-Principal, R. Zhu, S. Benz, S. Rabizadeh, P. Soon-Shiong, K. D. Yu, Z. Shao, X. Li, M. Gilcrease, Z. Lai, Y. Chen, T. H. Huang, H. Shen, X. Liu, M. Ferrari, M. Zhan, S. T. Wong, M. Kumaraswami, V. Mittal, X. Chen, S. S. Gross and J. C. Chang (2014). "Targeting RPL39 and MLF2 reduces tumor initiation and metastasis in breast cancer by inhibiting nitric oxide synthase signaling." Proc Natl Acad Sci U S A **111**(24): 8838-8843.

Heinz, S., C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh and C. K. Glass (2010). "Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities." Mol Cell **38**(4): 576-589.

Hu, Y. and G. K. Smyth (2009). "ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays." J Immunol Methods **347**(1-2): 70-78.

Huang, W., R. Loganantharaj, B. Schroeder, D. Fargo and L. Li (2013). "PAVIS: a tool for Peak Annotation and Visualization." Bioinformatics **29**(23): 3097-3099.

Landen, C. N., Jr., A. Chavez-Reyes, C. Bucana, R. Schmandt, M. T. Deavers, G. Lopez-Berestein and A. K. Sood (2005). "Therapeutic EphA2 gene targeting in vivo using neutral liposomal small interfering RNA delivery." Cancer Res **65**(15): 6910-6918.

Li, H. (2013). "Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM." eprint arXiv:1303.3997.

Li, X., M. T. Lewis, J. Huang, C. Gutierrez, C. K. Osborne, M. F. Wu, S. G. Hilsenbeck, A. Pavlick, X. Zhang, G. C. Chamness, H. Wong, J. Rosen and J. C. Chang (2008). "Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy." J Natl Cancer Inst **100**(9): 672-679.

Schott, A. F., M. D. Landis, G. Dontu, K. A. Griffith, R. M. Layman, I. Krop, L. A. Paskett, H. Wong, L. E. Dobrolecki, M. T. Lewis, A. M. Froehlich, J. Paranilam, D. F. Hayes, M. S. Wicha and J. C. Chang (2013). "Preclinical and clinical studies of gamma secretase inhibitors with docetaxel on human breast tumors." Clin Cancer Res **19**(6): 1512-1524.

Tanaka, T., L. S. Mangala, P. E. Vivas-Mejia, R. Nieves-Alicea, A. P. Mann, E. Mora, H. D. Han, M. M. Shahzad, X. Liu, R. Bhavane, J. Gu, J. R. Fakhoury, C. Chiappini, C. Lu, K. Matsuo, B. Godin, R. L. Stone, A. M. Nick, G. Lopez-Berestein, A. K. Sood and M. Ferrari (2010). "Sustained small interfering RNA delivery by mesoporous silicon particles." Cancer Res **70**(9): 3687-3696.
Zhang, Y., T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li and X. S. Liu (2008). "Model-based analysis of ChIP-Seq (MACS)." Genome Biol **9**(9): R137.

Fig. S1

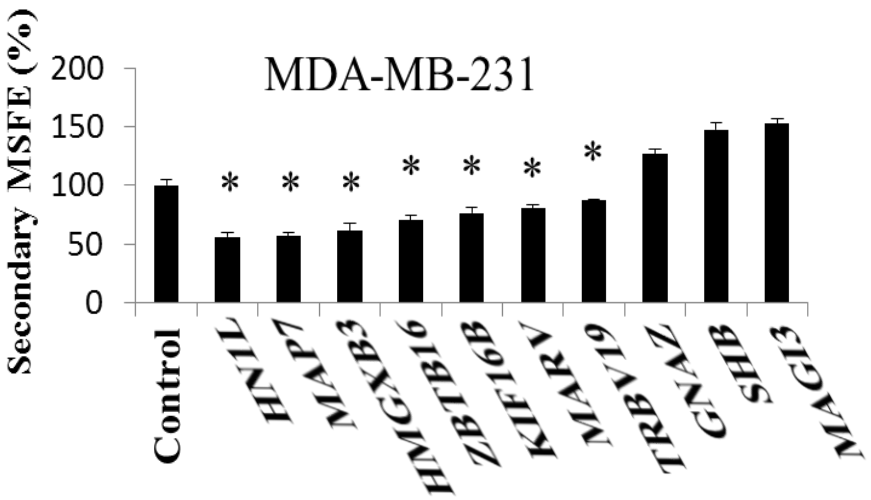
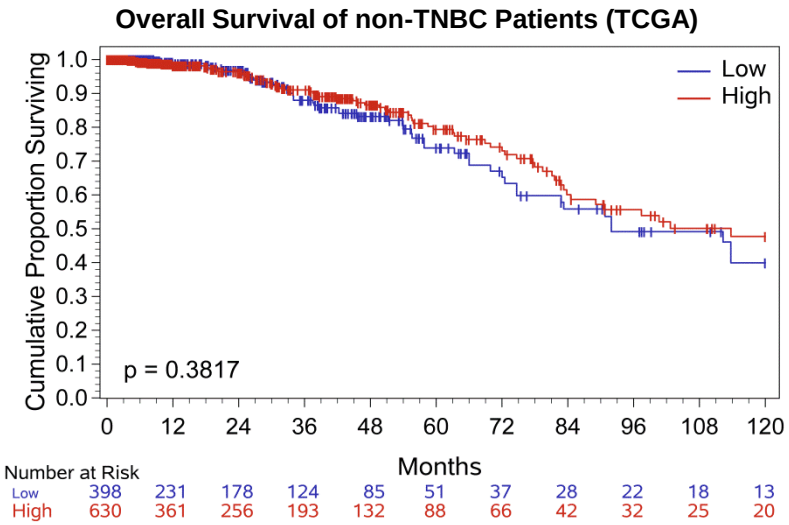
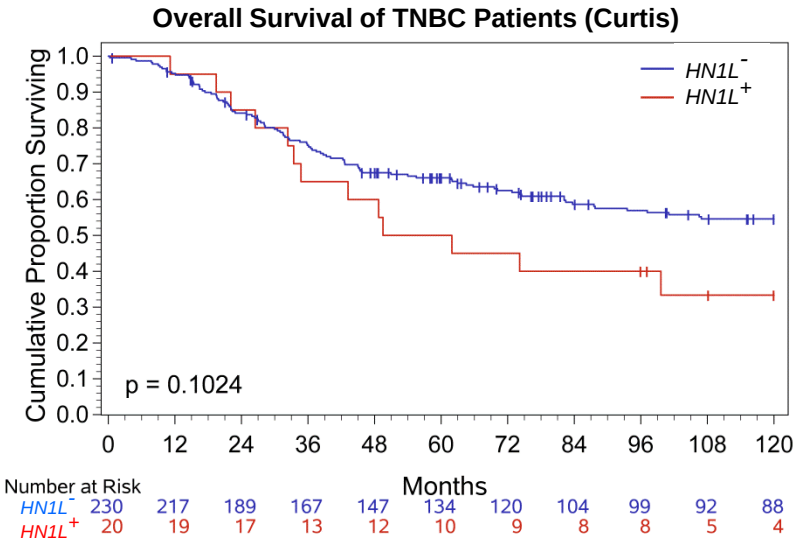

MDA-MB-231
Secondary MSFE (%)
200
150
100
50
0
*
*
*
*
*
*
*
Control
HN1L
MAP7
MAJN
HMGXB3
ZBTB16
KIF16B
MARV
TRBV19
GNAZ
SHB
MAGI3

Fig. S2

A


Overall Survival of non-TNBC Patients (TCGA)
Cumulative Proportion Surviving
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Low
High
p = 0.3817
0
12
24
36
48
60
72
84
96
108
120
Months
Number at Risk
Low 398 231 178 124 85 51 37 28 22 18 13
High 630 361 256 193 132 88 66 42 32 25 20

B


Overall Survival of TNBC Patients (Curtis)
Cumulative Proportion Surviving
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
HN1L-
HN1L+
p = 0.1024
0
12
24
36
48
60
72
84
96
108
120
Months
Number at Risk
HN1L- 230 217 189 167 147 134 120 104 99 92 88
HN1L+ 20 19 17 13 12 10 9 8 8 5 4

C

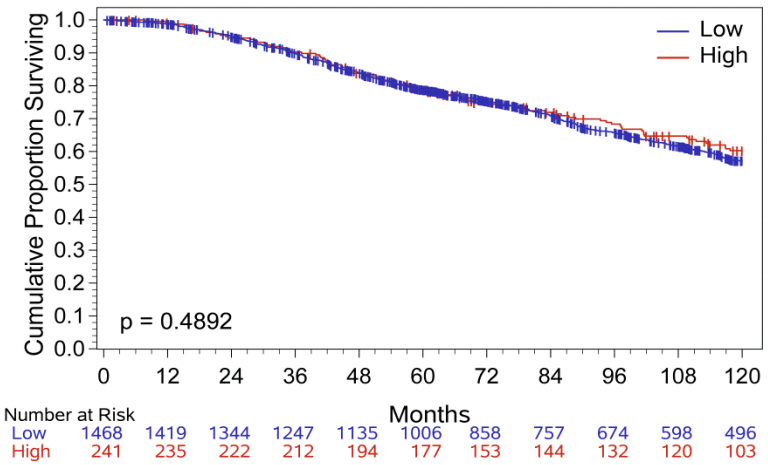

Overall Survival of non-TNBC Patients (Curtis)
Cumulative Proportion Surviving
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Low
High
p = 0.4892
0
12
24
36
48
60
72
84
96
108
120
Months
Number at Risk
Low 1468 1419 1344 1247 1135 1006 858 757 674 598 496
High 241 235 222 212 194 177 153 144 132 120 103

Fig. S3

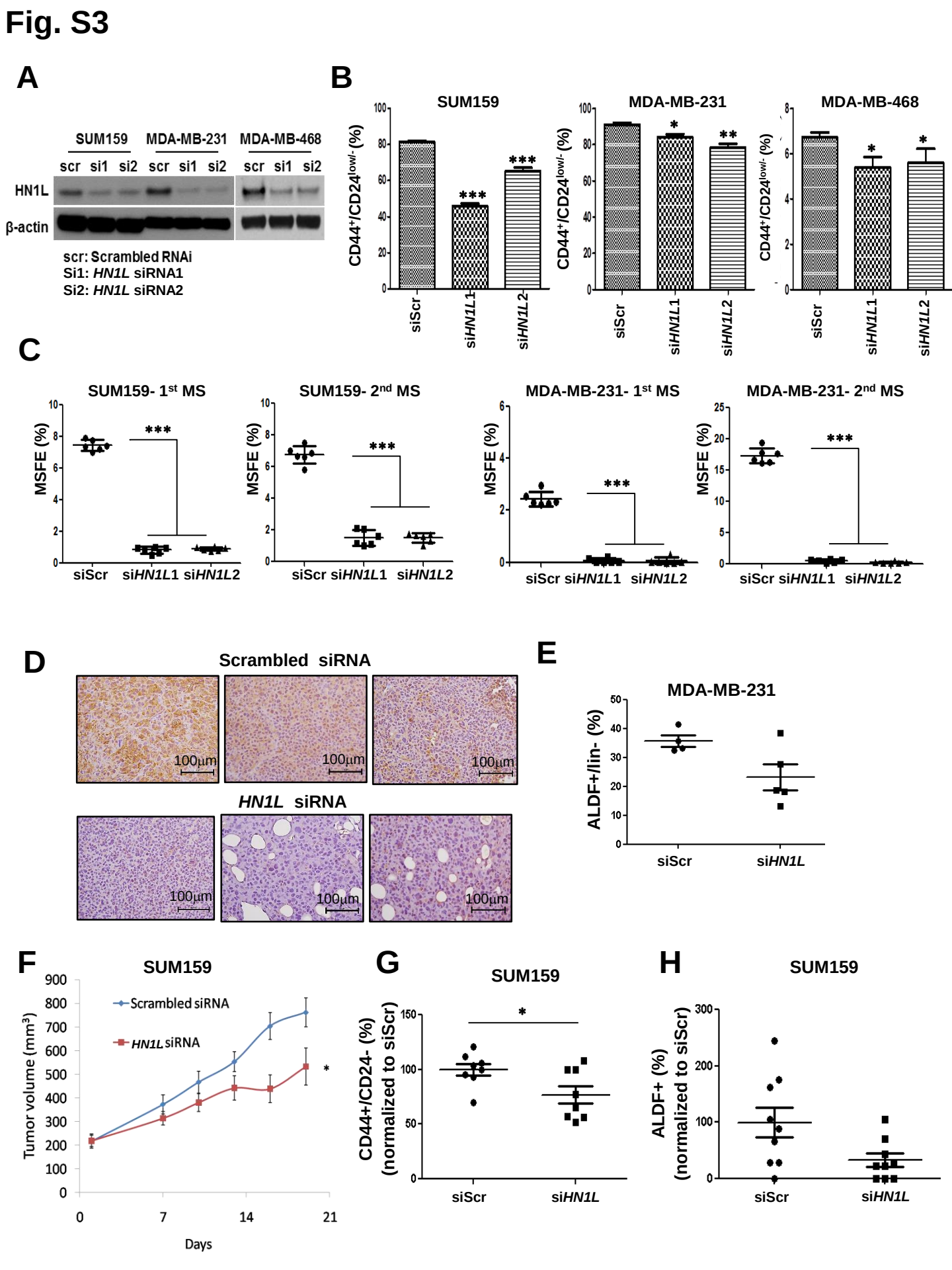

A
SUM159
MDA-MB-231
MDA-MB-468
scr si1 si2 scr si1 si2 scr si1 si2
HN1L
β-actin
scr: Scrambled RNAi
Si1: HN1L siRNA1
Si2: HN1L siRNA2
B
SUM159
MDA-MB-231
MDA-MB-468
CD44+/CD24low/- (%)
siScr
siHN1L1
siHN1L2
C
SUM159- 1st MS
SUM159- 2nd MS
MDA-MB-231- 1st MS
MDA-MB-231- 2nd MS
MSFE (%)
siScr siHN1L1 siHN1L2
D
Scrambled siRNA
HN1L siRNA
100µm
E
MDA-MB-231
ALDF+/lin- (%)
siScr
siHN1L
F
SUM159
Scrambled siRNA
HN1L siRNA
Tumor volume (mm3)
Days
G
SUM159
CD44+/CD24- (%) (normalized to siScr)
siScr
siHN1L
H
SUM159
ALDF+ (%) (normalized to siScr)
siScr
siHN1L

Fig. S4

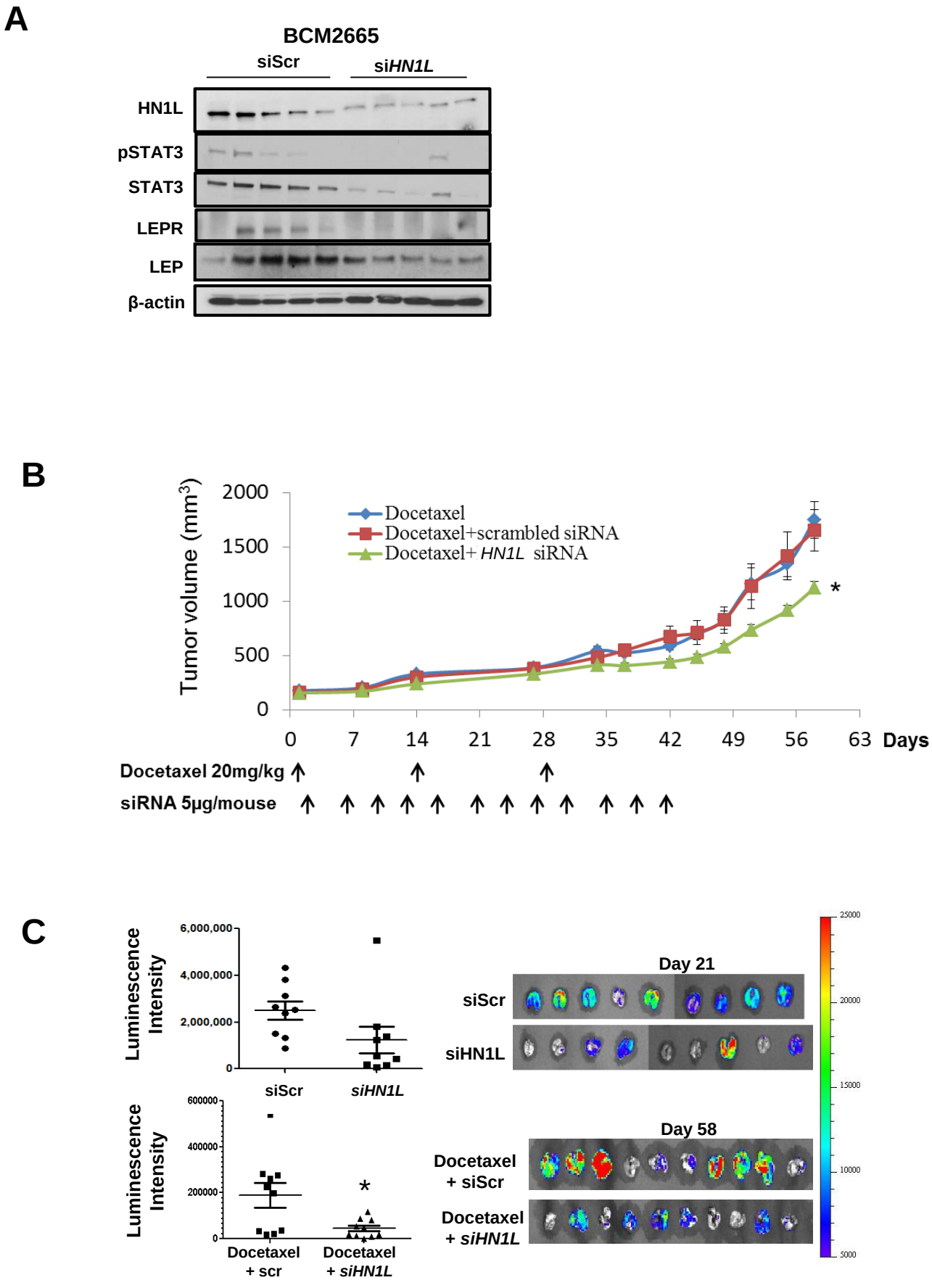

A
BCM2665
siScr
siHN1L
HN1L
pSTAT3
STAT3
LEPR
LEP
β-actin
B
Tumor volume (mm3)
2000
1500
1000
500
0
Docetaxel
Docetaxel+scrambled siRNA
Docetaxel+ HN1L siRNA
*
0
7
14
21
28
35
42
49
56
63
Days
Docetaxel 20mg/kg
siRNA 5μg/mouse
C
Luminescence Intensity
6,000,000
4,000,000
2,000,000
0
siScr
siHN1L
600000
400000
200000
0
*
Docetaxel + scr
Docetaxel + siHN1L
Day 21
siScr
siHN1L
Day 58
Docetaxel + siScr
Docetaxel + siHN1L
25000
20000
15000
10000
5000

Fig. S5

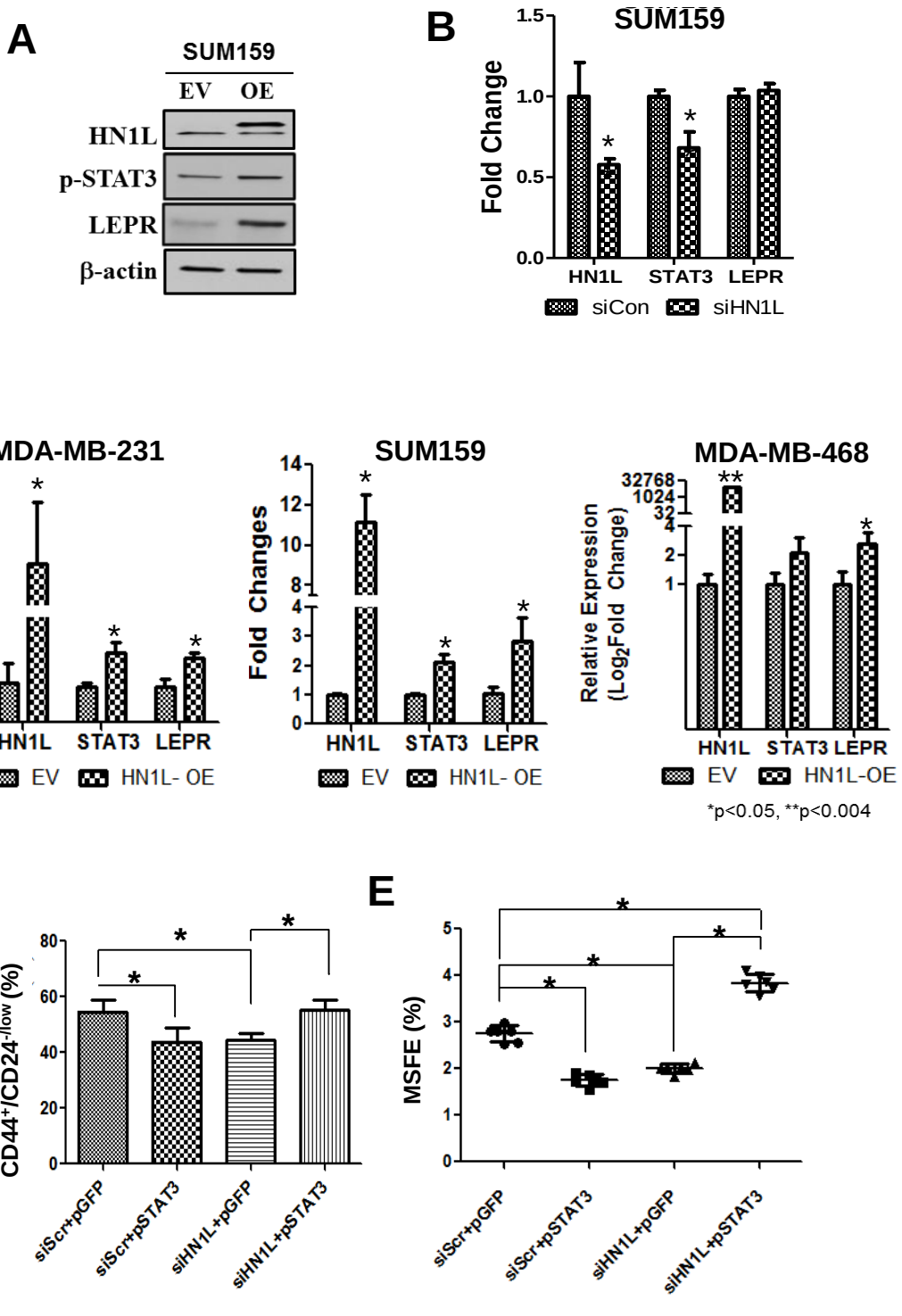
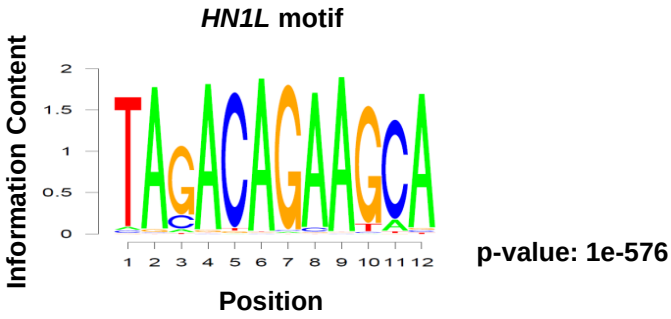
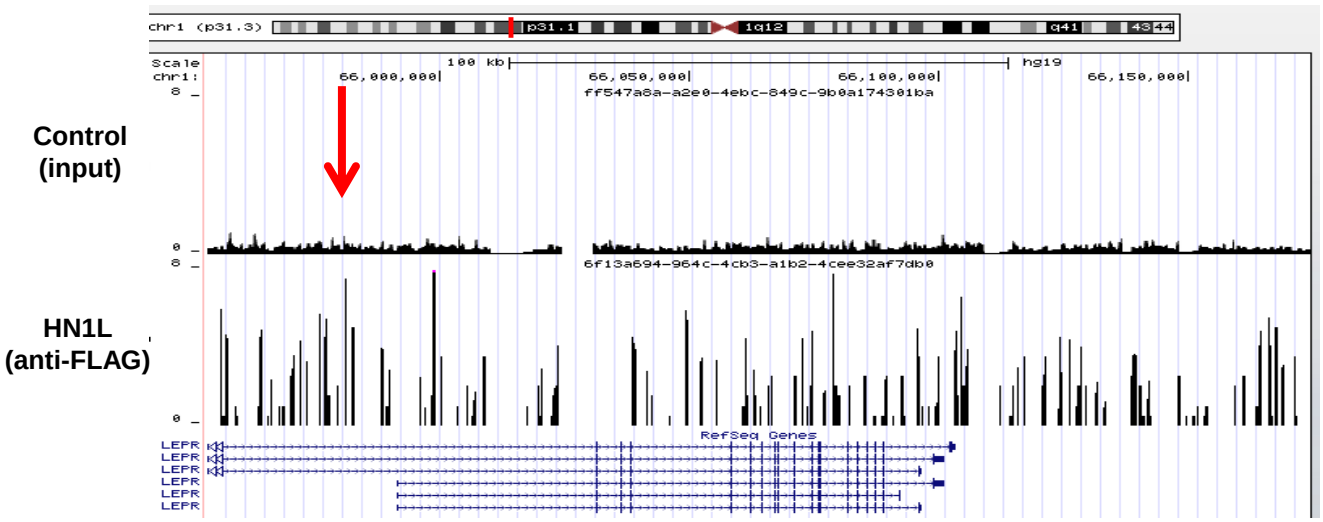

A
SUM159
EV
OE
HN1L
p-STAT3
LEPR
β-actin
B
SUM159
Fold Change
HN1L
STAT3
LEPR
siCon
siHN1L
C
MDA-MB-231
Fold Changes
HN1L
STAT3
LEPR
EV
HN1L- OE
SUM159
Fold Changes
HN1L
STAT3
LEPR
EV
HN1L- OE
MDA-MB-468
Relative Expression (Log2Fold Change)
HN1L
STAT3
LEPR
EV
HN1L-OE
*p<0.05, **p<0.004
D
CD44+/CD24-/low (%)
siScr+pGFP
siScr+pSTAT3
siHN1L+pGFP
siHN1L+pSTAT3
E
MSFE (%)
siScr+pGFP
siScr+pSTAT3
siHN1L+pGFP
siHN1L+pSTAT3

Fig. S6

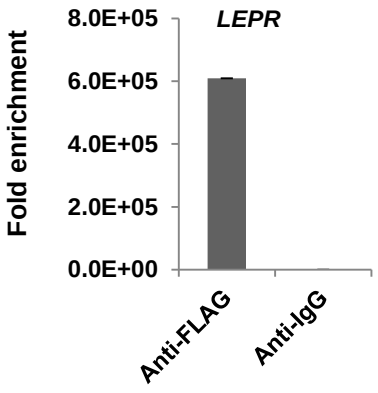
A



B



C

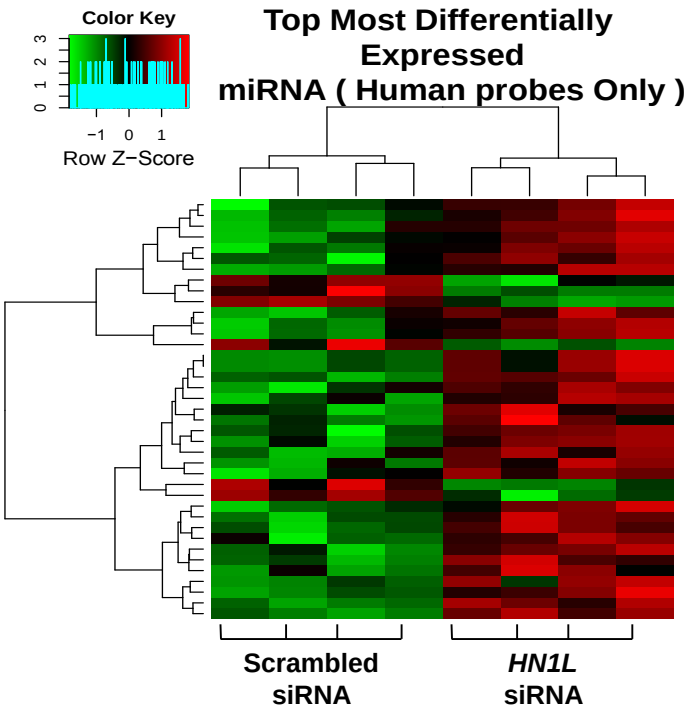


D

Pathway ID (GO)	Pathway Description	False Discovery Rate	Matching Proteins in the Network
0045893	Positive regulation of transcription, DNA-templated	5.75E-05	AR,ARID1B,EDN1,FGFR2,FOXP1, HOXA11,MYC,MYO6,PAX1,PBX1, PPP3CA,RERE, STAT3,TBX5,VDR
0045944	Positive regulation of transcription from RNA polymerase II promoter	5.75E-05	AR,EDN1,FGFR2,FOXP1,MYC,MYO6, PAX1,PBX1,PPP3CA,RERE,STAT3, TBX5,VDR
0009891	positive regulation of biosynthetic process	6.27E-05	AR,ARID1B,EDN1,FGFR2,FOXP1, HOXA11,MYC,MYO6,NT5E,PAX1,PBX1, PPP3CA, RERE, STAT3,TBX5,VDR
0022603	Regulation of anatomical structure morphogenesis	6.27E-05	EDN1,EPHB3,FGD6,FGFR2,FOXP1, HOXA11,MYC,NEDD4L,PPP3CA, SH3KBP1,TBX5,VDR
0001763	Morphogenesis of a branching structure	0.000104	EDN1,FGFR2,HOXA11,MYC,PBX1, RERE, VDR

Fig. S7

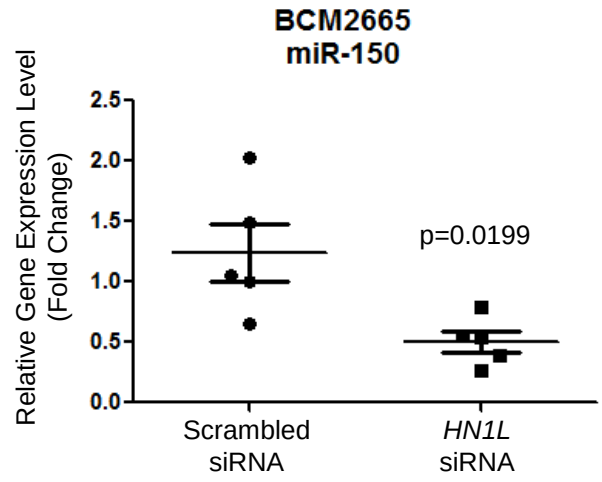
A



B

Transcript.ID. Array Design	logFC	t	p Value
hsa-miR-150-5p	-2.183121	7.984466	1.59E-05
hsa-miR-3921	-1.606919	5.161487	0.000493
hsa-miR-5195-3p	-1.173754	4.828139	0.000792
hsa-miR-3188	-1.150942	4.696959	0.000959
hsa-miR-6746-3p	1.7503	-5.45189	0.00033
hsa-miR-6504-5p	1.76714	-4.73776	0.000904
hsa-miR-2113	2.16653	-4.11002	0.002328
hsa-miR-6868-3p	2.32171	-4.73751	0.000904

C



D