Supplemental Files, Figures and Tables

Supplemental File 1. LncRNAs enriched in basal-like and TNBC patient tumors were assessed for detection in SUM149 cells and PDX 7482 by QPCR. TNBC versus non-TNBC or basal-like vs non-basal breast cancer RNA-seq expression of 48 lncRNAs previously identified as being enriched in TNBC and basal-like breast cancer by Zhang et al., *Nat. Struct. Mol. Biol.* 23, 522–530 (2016) were assessed for detection by QPCR in SUM149 cells and PDX 7482.

Supplemental File 2. Genes regulated by NRAD1 or ALDH1A3 in MDA-MB-468 cells, and NRAD1-regulated genes identified in Gene Ontology (GO) term enrichment analysis processes.

Supplemental File 3. NRAD1 ChIRP-seq peaks.

Supplemental Figures

**Supplemental Figure 1. KM plotter Regression Free Survival analysis of lncRNAs enriched in CSC populations.** Of the 10 IncRNAs identified in Figure 1E, six of the IncRNAs had probes in breast cancer gene chip array data and clinical data compiled by KM plotter. Regression Free Survival based on median expression in 360 basal-like breast cancer patient tumors assessed with KM Plotter (gene chip array). For IncRNAs where more than one probe set is available, the data for all available probe sets is shown. HR = hazard ratio.
Supplemental Figure 2. TCGA (breast invasive carcinoma, Cell 2015 dataset, cbioportal) overall survival analysis of lncRNAs enriched in CSC populations. Of the 10 lncRNAs identified in Figure 1E, five of the lncRNAs had RNAseq expression data that was extractable by cbioportal from the TCGA, breast cancer Cell 2015 dataset. The overall survival for the 107 patients with basal invasive ductal carcinoma was plotted based on expression of the lncRNAs and their associated clinical data that was also extracted from cbioportal. The patients were divided into high or low expression based on being in the top or bottom half and the survival plots were generated with Graphpad Prism software. HR = hazard ratio (log-rank). The p-value was calculated based on the log-rank (Mantel-Cox) test.
Supplemental Figure 3. TCGA (Breast Invasive Carcinoma, BRCA dataset) overall survival analysis of lncRNAs enriched in CSC populations. Using the TANRIC portal, expression of the 10 lncRNAs identified in Figure 1E were assessed for correlations with overall survival in 139 basal breast cancer patients that are part of the TCGA-BRCA dataset. The survival plots were generated in TANRIC and exported from TANRIC. The hazard ratio was not included in the analysis by TANRIC.
Supplemental Figure 4. NRAD1 is predominately expressed in basal-like breast cancer cell lines. NRAD1 expression in 21 cancerous, and two normal-like breast cell lines was determined by QPCR. PUM1 and ARF1 are used as reference genes in the panel due to target stability values across all 23 cell lines (n=4). Error bars represent standard deviation.
Supplemental Figure 5. NRAD1 is poorly expressed in most normal human tissues. The GTEx Portal (gtexportal.org) was used to assess NRAD1 expression levels across human tissues. The Genotype-Tissue Expression (GTEx) project is a resource database and associated tissue bank used to study the relationship between genetic variation and gene expression, and other molecular phenotypes, in multiple reference tissues.
Supplemental Figure 6. LINC00162 expression correlation with ALDH1A3 expression in breast cancer patient tumors and cell lines. (A) RNA-seq co-expression of LINC00162 (PICSAR) and ALDH1A3 in the TCGA Cell 2015 dataset was retrieved with cBioportal. (B) RNA-seq co-expression of LINC00162 and ALDH1A3 in the Cancer Cell Line Encyclopedia (only breast cancer cell lines) was retrieved using the CCLE portal (r = pearson correlation).
Supplemental Figure 7. GAPDH expression in cytoplasmic and nuclear compartments is nearly equal. Post cellular fractionization, RNA was isolated and cDNA synthesized. The GAPDH levels in the cytoplasmic and nuclear compartments were measured using QPCR in MDA-MB-468 cells. Transcript levels in each compartment is represented relative to the total levels of GAPDH in the cell, which is set to 100% for each n (i.e. levels of GAPDH are nearly equal in both compartments, n =3). Error bars represent standard deviation.
Supplemental Figure 8. QPCR validation of a representative sampling of the microarray-identified NRAD1-regulated genes in MDA-MB-468, SUM149, and MCF7 cells. Log2 fold change of transcript levels in cells treated with anti-NRAD1-specific GapmeR#3 or #4 versus control GapmeR in MDA-MB-468 cells (A), SUM149 (B), and MCF7 cells (C). Expression is normalized to reference genes PUM1 and ARF1 and represented as fold change over GapmeR control-treated cells (n=4). Error bars represent standard deviation (ND = not detected, i.e. expression levels below quantification threshold).
Supplemental Figure 9. Uncropped ALDH1A3 western blots (right) and total protein loading control blots (left).
### Supplemental Table 1. Primers used in QPCR, western antibody details, and GapmeR sequences.

| Primer sequences | Forward         | Reverse                  |
|------------------|----------------|--------------------------|
| PUM1             | GGCCTTAGCATGGTGGAGTA | CATCCCTTGGGCCAATCCT     |
| ARF1             | GTGTTCGCCCAACAAGCAGG | CAGTTCTCTTGCGGTAGTGA    |
| GAPDH            | GAGATCAACGGATTGGTCGA | TTCTCCAGTGGTGGAAGAC     |
| B2M              | AGGCTATCCAGCGTACTCCA | CGGATGGATGAACCACAGACA   |
| NRAD1 (LINC00284) | CCAGGGGATAAAACCCGCT | TAAGCAACAGTCAGCTGCTC    |
| ALDH1A1          | TGTTAGCTGTAGCCGACTTG | TTCTTAGCCCGCTCAACACT    |
| ALDH1A3          | TCTCGCACAAGGCCCTGAAGT | TATTCGGCCAAAGCGTATTC    |
| ZFAS1            | CAACTACTAGAGGCCTCGG | CCAAGATGGCTTTGCACC      |
| SNHG6            | CTGTCTTGGATGCTGCTC | CGGCATGACTACCGCTCTT     |
| LRCC75A-AS1      | TTCCCCGTTGTATGGAGG GC | TCCAGTTCTCTCGGTTTGC     |
| AC004542.2       | CAGAAAGGCGACCATACCA | ACCCCTGAGTACCCAGAGTAA   |
| AC0093001        | TCAGCCAAAGGAATTAACCTACCA | CGGCTGGGTTATATTTCA A   |
| UCA1             | CCAGAGAGCCGATCAGACAAA | GGGATGCGCATTGGGAAGGA    |
| TUG1             | AGCGTGGAATGTCAGTAAGG | GCAAGATTGGGGAACCTT     |
| NORAD            | CTAAAGGGGCCCTGGAAGGT | AGAATGAAGACCAACC G C C C |
| MALAT1           | GCAAAAACGTTGCGCTGCTT | GTGGCAAATGGCGGACTTT    |
| LINC-ROR         | GAATCAGAGTGCTGGGCAGT | TCACGAGCTCATGCCCTAAC   |
| HULC             | ACTCTGAAGTTAAGGC CGGAA | TGCCAGGAAACTTCTTGCTT  |
| HOTAIR           | GGGACAGAAGGGAAGCCCTC | GAGTCAGAGTTCCCCACTGC   |
| H19              | GAGAGCTTTGTTGGAGGCAAG | CCTACTCCACACTCCTCAGT    |
| CYTOR            | TTCCAACCTCCGTCCTGAC T | GGGGCTGAGTCTGATTTTT    |
| TUNAR            | ACCGGCAGCCTTATTGT TTC | CTAATCCCGAGCTTTCCCCC    |
| GAS1RR           | AAGGGGCAGAAGTCGTATAC GC | GGAAGTGCTTGATCAGTGT    |
| NEAT1            | CCTCCCTTTAAACCTTATCCATTCAC | TCTCTTCCACACCATTA CCA   |
| DANCR            | AGGACTGTCTGTCTTACCTGCT | TGAATACCAGCAACAGGACA    |
| PART1            | CAGGCTAGCCAAACTATAGGAC | TTCAGCTTTCCAGAGCAGT    |
| LINC00511        | TTCCCCAGGAAACCCACAC | CATCACCTGTCTCCTTGGCA    |
| LINC01198        | TGCCGAATAGCTCTGACCTG | GTTGAAGAAGGTTGCACTTG GA |
| LINC01354        | ATGCACACATCTCGAGGGGA | CTGTTGAGAGACCTTGAGAG    |
| LINC02159        | TCCCATCGGCTTTGGCGCTT | GACACCTGTCTCCTGCTTTAG  |
| AC022509.2       | AGTGGGAACTTGTGACACGCA | GCTTTCCAGCCTCACTTTGG    |
| LINC00518        | ACCAAAACATTTTGCTGCC | GCCTAAACATTTTGCTGCC    |
| Gene ID          | Sequence 1                      | Sequence 2                      |
|------------------|---------------------------------|---------------------------------|
| LINC00162        | ATTGCTTAGGTGGGAGCAC             | TCATGACTGAGCTTTCCGTG           |
| (PICSAR)         |                                 |                                 |
| RP11-120J1.1     | TCCCCAAACCCCATGACTCTG           | AGCCCCATCACAGTGGTCCCTT         |
| LINC01315        | CCAATTTCCCAGGTTTTCC            | GGCATCCACTTCATCGCTCA           |
| AP000851.1       | TGCTTTGCCACATAGCTCACA          | GGGTCCATGTTTTCCAAGACG          |
| LNC-DPMI-1:1:2   | CCCCTGCACTGTCATACTGTCTG        | GTGGAAAAATCTCTGACATGTTGTA     |
| LINC01833        | CTTTGTTGGGTTGGGAGGCG          | AGGTTGAGCTGGGCGAATACTG         |
| LINC00839        | GGGCCAGATTGTTCACAGGAT          | TGGTAAAGGCGAGATCCCCCA          |
| LINC00880        | CCGGAAAGGGTGACTCGTG            | CCAGGGGCTTTGATACCTCT          |
| LINP1            | GACCAGGGCCTCTGTGTAAGG          | GCAGTGGAGTCTGAGATCCCC          |
| AP001626.1       | CCCCGTGGAGAACCTCTTCT            | GATAGCCAGGCGGCAATTT            |
| AC016995.3       | ACTCTGCTCTCTGTGACTGC           | CTACTCAGCTCCCTTTGGTC           |
| LINC01956        | AGTTGTTGACTTTGGGGACTGC         | CAAAAGTCCGCACTACCTG            |
| LINC00707        | CCCAGACATGAGCCGATGAC           | ATTTGGTTTGCTGCGGCT            |
| LNC-NKX-1-2-1:1  | GACCCCTCTGGTTTACCAGA           | GGATTTGTTGGGAAAGGCACT          |
| LINC02188        | TTTGAGGACCCCTGGAATTGC          | TCCACGCGTTTGCTGCTG             |
| AC091053.1       | GAAGCCAGAAACACATCTGAGAC        | GAGAGGCGCTTCCCGCAATCT          |
| FOXP4-AS1        | CTCTGTTTCCGGTGCAACCT           | GACCTGGAGCTGTCATCGAG           |
| DGC5             | CCATGCTTCAGCCGACATTC           | CCAGGGGCCTTCTTTTCT            |
| SOX9-AS1         | CAACACTGCATTTGGGGGAG           | TGCTAGGCGCTTACACTG             |
| LINC01819        | CTAGGCACACTTGCCCTACC           | GTGGGTCTACCTGTATGCC            |
| LINC00092        | CCCATTTCCTTTAGGCCCGT          | AGAAACATGCTTTGCGCTG           |
| LNC-DSC2-1:1     | AGCTGCCTCTTTATTCTCCTGTAG       | TTGCTAGGAGAATGGCAGT            |
| AC027031.2       | GTGCAGGATCCGGAACACAGGA         | AACATGCGAACGCTGGATGGA          |
| VIM-AS1          | GCCCCAGGCAATGAGCTACAT           | CGACGCTTTGTCCTGATGGA           |
| LINC02487        | GTCTGTGTCTTCCCTCAAGAGGC       | AGAAACCAGGTGGCTGGATGTA         |
| AC025154.2       | ACCACAGGTGCGCCACATAG           | CAGCTGCCCCTGTGAGTAA           |
| AC009041.2       | GTCCAACCCGGGCTCC              | GGTGAAAGGATCGCGAGCA           |
| AC015712.4       | AAGAACCAAGGTTGCAACAGGA        | GCACCTTGGACTTCCCGTC           |
| CYP4Z1           | GAATCCTGGTGGTGGTGCTGAG         | AGGTTTCAACATCTGGGCT            |
| CTNNAL1          | CCATGATGGGCTCCTTCTAGGCA        | ACCCATCCTGTTTTTCCATCTGA        |
| EIF5A2           | AGAACCGGCTTCCGTGGTACTG         | CGTGCTTTCCGGTCTGTCGA           |
| FOSL1            | CTGGTGCCAAAGCATCACAC           | ACTGAGGCTAGGTAGCTAGG           |
| FSCN1            | GCAAGAATGCCAGCTGTAC            | ACAAACCTTGCCATTTGGGAGCA        |
| IL7R             | TTCTCTGTGGCTCTGTGTTGTC         | ACTGGGCCATACGATAGGCT           |
| SERPINE2         | ATTGAACTGCGCTACCACGG          | GTGCGGAGATACGGCAGCA            |
### Western blot antibodies

| ALDH1A3   | Details, catalogue number | Concentrations |
|-----------|---------------------------|-----------------|
| Mouse monoclonal antibody, clone OTI4E8. Origene catalogue number: TA502841 | WB: 1/1000 | |

### GapmeR Sequences

| Control GapmeR (negative control A) | 5’-AACACGTCTATACGC-3’ |
|-------------------------------------|-----------------------|
| **NRAD1 GapmeR#3**                  | 5’-GCTGAACGCTGCCTTT-3’ |
| **NRAD1 GapmeR#4**                  | 5’-CTTTGCTGAACTGATG-3’ |

### Supplemental Table 2. Summary of survival analyses completed on lncRNAs from Supplemental Figures 1, 2, and 3.

| lncRNA                                                                 | KM Plotter | TCGA (Cell 2015, cBioportal) | TCGA (BRCA, TANRIC) |
|-----------------------------------------------------------------------|------------|------------------------------|---------------------|
| RP11-1201.1 (ENSG00000225472.1)                                       | no data    | no data                      | p = 0.176 (trend, high expression better survival) |
| AC022509.2 (ENSG00000256234.1)                                        | no data    | no data                      | p = 0.401          |
| LINC00162 (PICSAR)                                                    | HR = 0.9, p = 0.54 | HR = 1.53, p = 0.444 | p = 0.817          |
| LINC00284 (NRAD1)                                                     | HR = 1.66, p = 0.0025 | HR = 2.36, p = 0.157 | p = 0.404          |
| LINC00511                                                             | no data    | no data                      | p = 0.92           |
| LINC00518                                                             | HR = 1.22, p = 0.22 (probe 1) | HR = 2.19, p = 0.172 | p = 0.737          |
|                                                                     | HR = 1.05, p = 0.79 (probe 2) |                   |                    |
| LINC01198 (ENSG00000231817.3)                                        | HR = 0.57, p = 0.00069 | no data | p = 0.925          |
| LINC01354 (ENSG00000231768.1)                                        | HR = 0.87, p = 0.41 | no data | p = 0.786          |
| LINC02159 (ENSG00000253417.1)                                        | no data    | HR = 1.178, p = 0.762 | p = 0.117 (trend, high expression better survival) |
| PART1                                                                | HR = 1.49, p = 0.0021 (probe 1) | HR =2.78, p = 0.071 | p = 0.532          |
|                                                                     | HR = 0.77, p = 0.043 (probe 2) |                   |                    |
Supplemental Table 3. NRAD1 is non-coding based on five metrics of protein-coding potential. Using online software Incipedia.org, the coding potential of NRAD1 was assessed by five metrics. The PRIDE reprocessing score analyzes the predicted open reading frames of a sequence against over 100 human proteomics mass spectra; score of 0 indicates no hits. The Lee translation initiation sites are mapped using lactimidomycin, an initiating ribosome inhibitor (compared to a no treatment control); a score of 0 means no difference (i.e. no translation). The PhyloCSF algorithm generates probabilistic models of coding potential based on Codon Substitution Frequencies; a score under +60.00 is likely to be non-coding (lower number increases odds that the sequence is non-coding). CPAT assesses ORF lengths (a long non-putative ORF is unlikely to be observed by random chance in a non-coding sequence) and ORF coverage (the ratio of ORFs to transcript lengths). Bazzini Small ORFs tests if small ORFs in the sequence are translated through ribosomal profiling. A score of 0 indicates no translation detected.

| Metric                  | Raw result | Interpretation |
|-------------------------|------------|----------------|
| PRIDE reprocessing 2.0  | 0          | non-coding     |
| Lee translation initiation sites | 0          | non-coding     |
| PhyloCSF score          | -78.2965   | non-coding     |
| CPAT coding probability | 0.83%      | non-coding     |
| Bazzini small ORFs      | 0          | non-coding     |