eLIFE
elifesciences.org

Figures and figure supplements

Musashi proteins are post-transcriptional regulators of the epithelial-luminal cell state

Yarden Katz, et al.

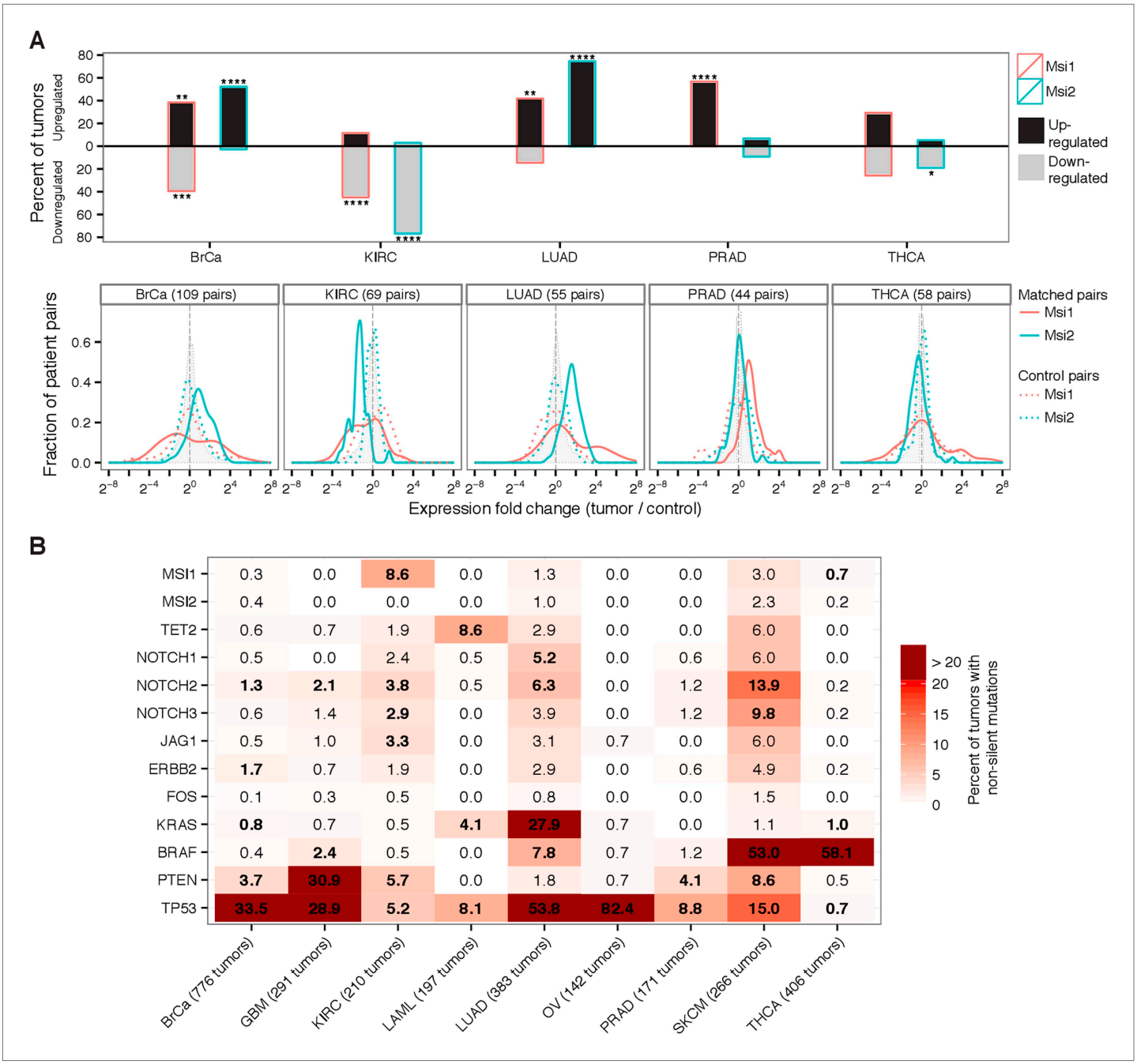

Figure 1. Msi genes are frequently overexpressed in breast, lung, and prostate cancer but downregulated in kidney cancer. (**A**) Top: percentage of matched tumor–control pairs with upregulated (black-fill bars) or downregulated (grey-fill bars) *Msi1* or *Msi2* in five cancer types with matched RNA-Seq data. Upregulated/downregulated defined as at least two-fold change in expression in tumor relative to matched control. Asterisks indicate one-tailed statistical significance levels relative to control pairs. Bottom: distribution of fold changes for *Msi1* and *Msi2* in matched tumor–control pairs (solid red and green lines, respectively) and in an equal number of control pairs (dotted red and green lines, respectively.) Shaded gray density shows the fold change across all genes. (**B**) Percentage of tumors with non-silent mutations in *Msi1*/*Msi2* and a select set of oncogenes and tumor suppressors across nine cancer types. Bold entries indicate genes whose mutation rate is at least two-fold above the cancer type average mutation rate.

DOI: 10.7554/eLife.03915.003

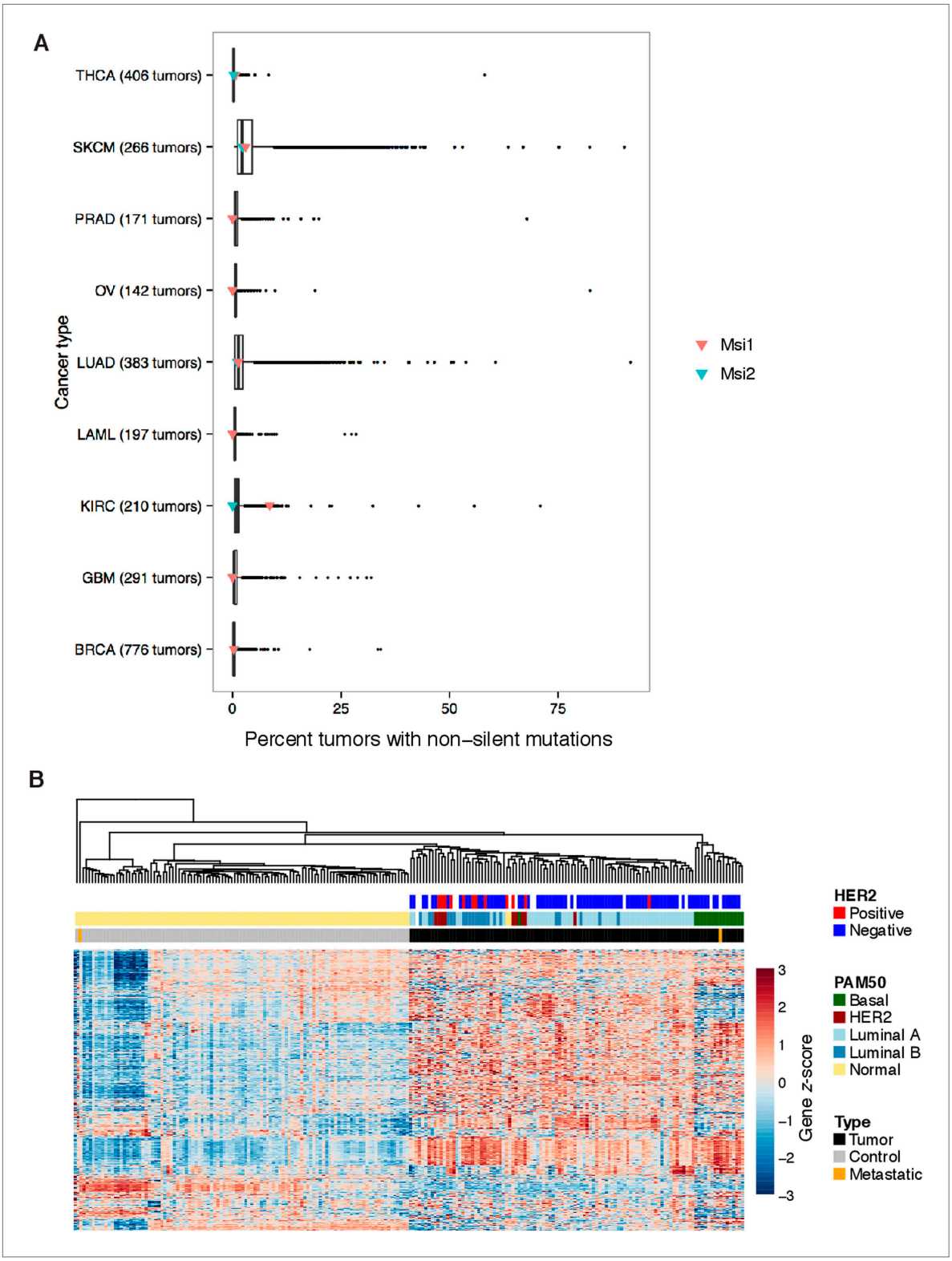

Figure 1—figure supplement 1. Analysis of *Msi1/Msi2* mutation and expression profiles in TCGA datasets. (**A**) Distributions of the percent of tumors with non-silent mutations across cancer types in TCGA DNA sequencing data. Red and green triangles indicate values for Msi1 and Msi2, respectively. (**B**) Unsupervised hierarchical clustering of breast cancer tumors and matched controls, with overlaid sample labels, clinical markers and PAM50 subtypes.

DOI: 10.7554/eLife.03915.004

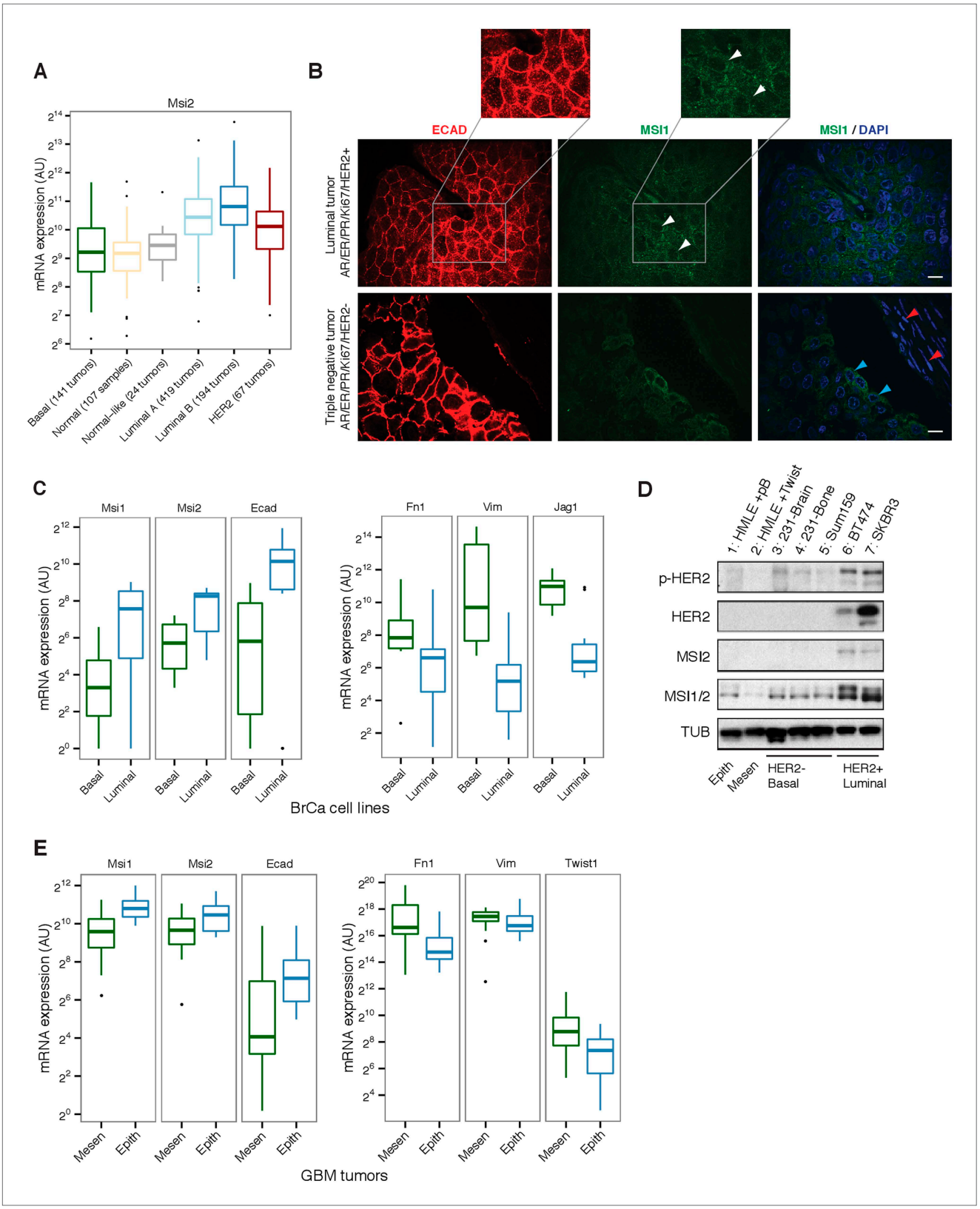

Figure 2. Msi is associated with the epithelial-luminal state in breast cancer. (**A**) mRNA expression of *Msi2* across different breast tumor types in TCGA RNA-Seq. (**B**) Immunofluorescence staining for Ecadherin (ECAD, red) and *Msi1* (MSI1, green). Top: luminal human breast tumor with high number of ECAD-positive cells. MSI1 shows primarily cytoplasmic localization (white arrowheads). Inset shows magnified version of ECAD and MSI staining. Bottom:

Figure 2. Continued on next page

Figure 2. Continued

triple negative, basal-like tumor. ECAD-positive cells showed strong cytoplasmic MSI1 stain (blue arrowheads) while ECAD-negative cells were MSI1-negative (red). Single confocal stacks shown, 10 μm scale. (**C**) mRNA expression of *Msi1*, *Msi2*, *Ecad*, *Fn1*, *Vim*, and *Jag1* in breast cancer cell lines by RNA-Seq (datasets are listed in ***Supplementary file 1***). (**D**) Western blot for MSI1/2 (MSI1/2 cross react. antibody), MSI2, phosphorylated HER2 (p-HER2) and HER2 in panel of breast cell lines. 'HMLE + pB' indicates HMLE cells infected with pB empty vector, 'HMLE + Twist' indicates HMLE cells infected with Twist transcription factor to induce EMT. MDAMB231-derived metastatic lines (231-Brain, 231-Bone) and Sum159 are basal, HER2-negative cancer cell lines. BT474 and SKBR3 are HER2-positive, epithelial-luminal cancer cell lines. Epithelial-luminal (HER2-positive) lines show increased expression of Msi proteins compared with basal lines, and Twist-induced EMT reduces Msi expression. (**E**) mRNA expression of *Msi1*, *Msi2*, *Ecad*, *Fn1*, *Vim*, and *Twist1* in GBM tumors classified as mesenchymal (n = 20) or epithelial (n = 20) using an EMT gene signature.

DOI: 10.7554/eLife.03915.005

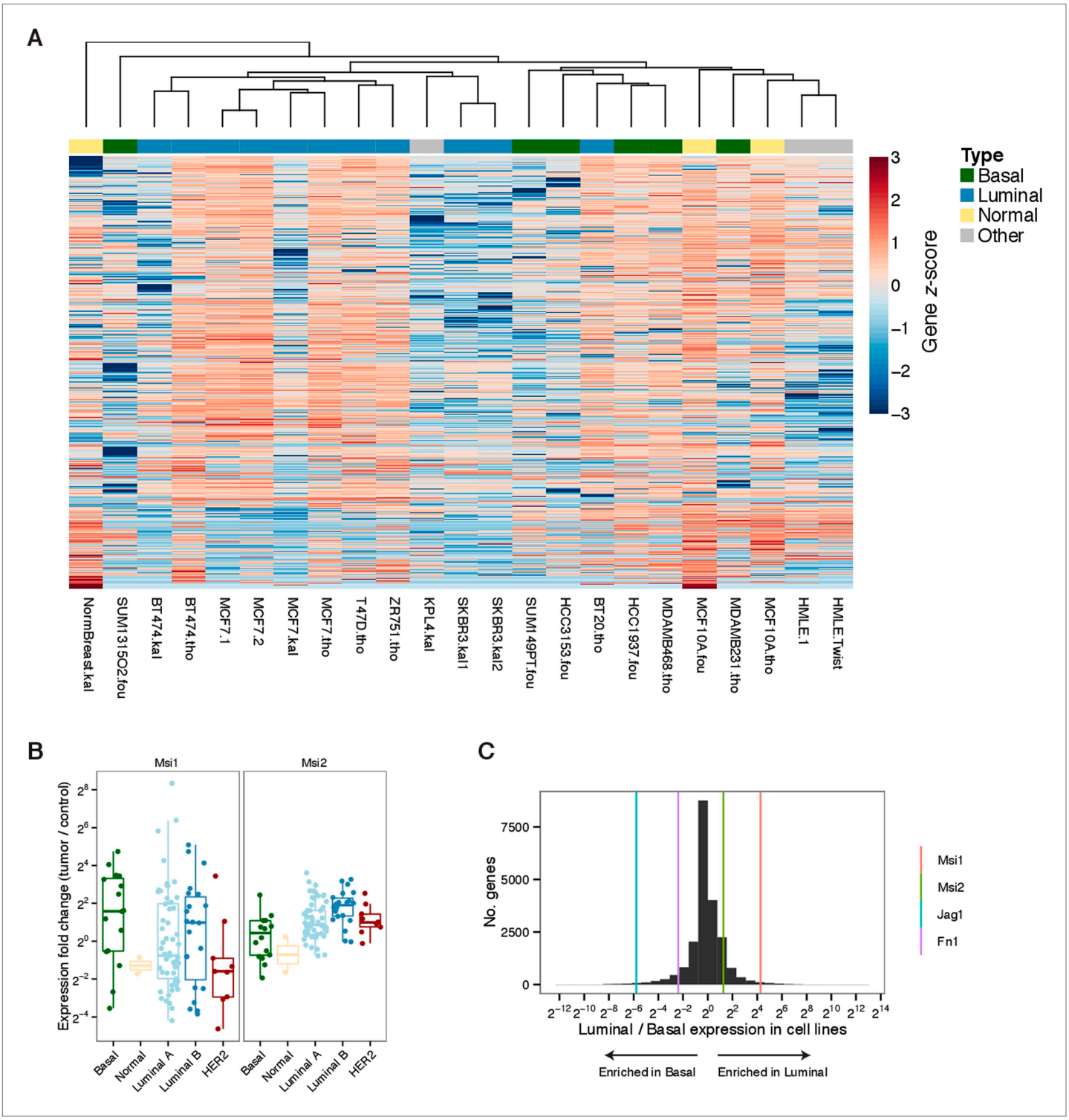

Figure 2—figure supplement 1. Expression of *Msi1/Msi2* in subtypes of breast cancer cell lines and breast cancer tumors. (**A**) Unsupervised hierarchical clustering of gene expression from RNA-seq of breast cancer cell lines. (**B**) Fold-change in tumor–control pairs of TCGA breast cancer tumors for Msi1 and Msi2 across tumor subtypes. Msi1 shows a variable bimodal distribution of fold changes, while Msi2 is enriched in Luminal B tumors relative to Basal tumors. (**C**) Ratio of luminal to basal cancer cell line fold changes for Msi1, Msi2, Jag1, and Fn1.

DOI: 10.7554/eLife.03915.006

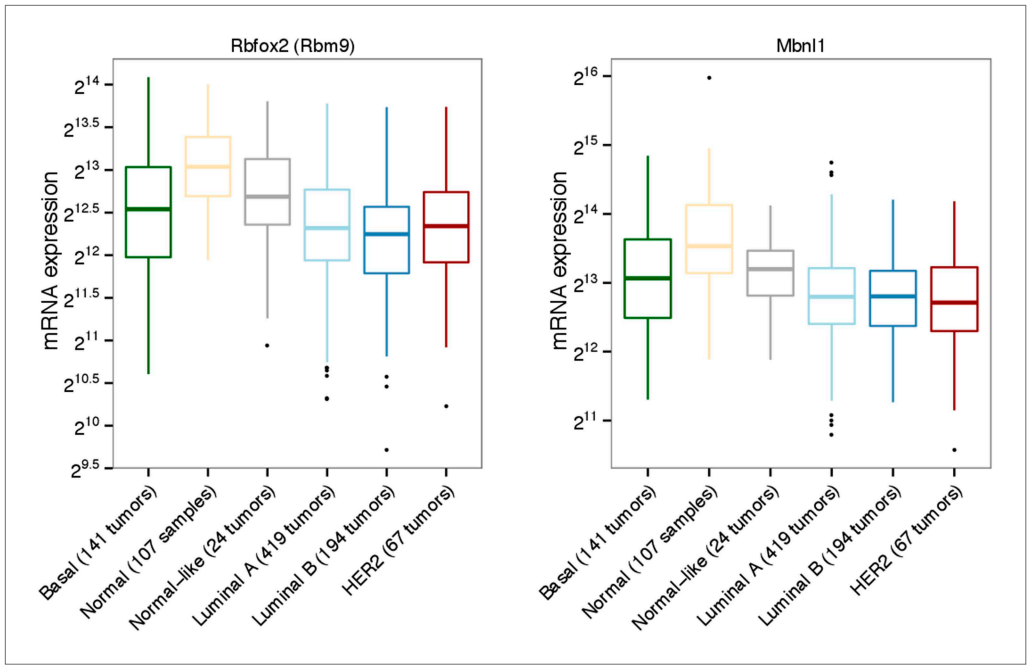

Figure 2—figure supplement 2. Expression of Rbfox2 (Rbm9) and Mbnl1 in subtypes of breast cancer tumors from TCGA. Expression values for Rbfox2/Mbnl1 plotted across PAM50 subtypes, after TMM normalization.
DOI: 10.7554/eLife.03915.007

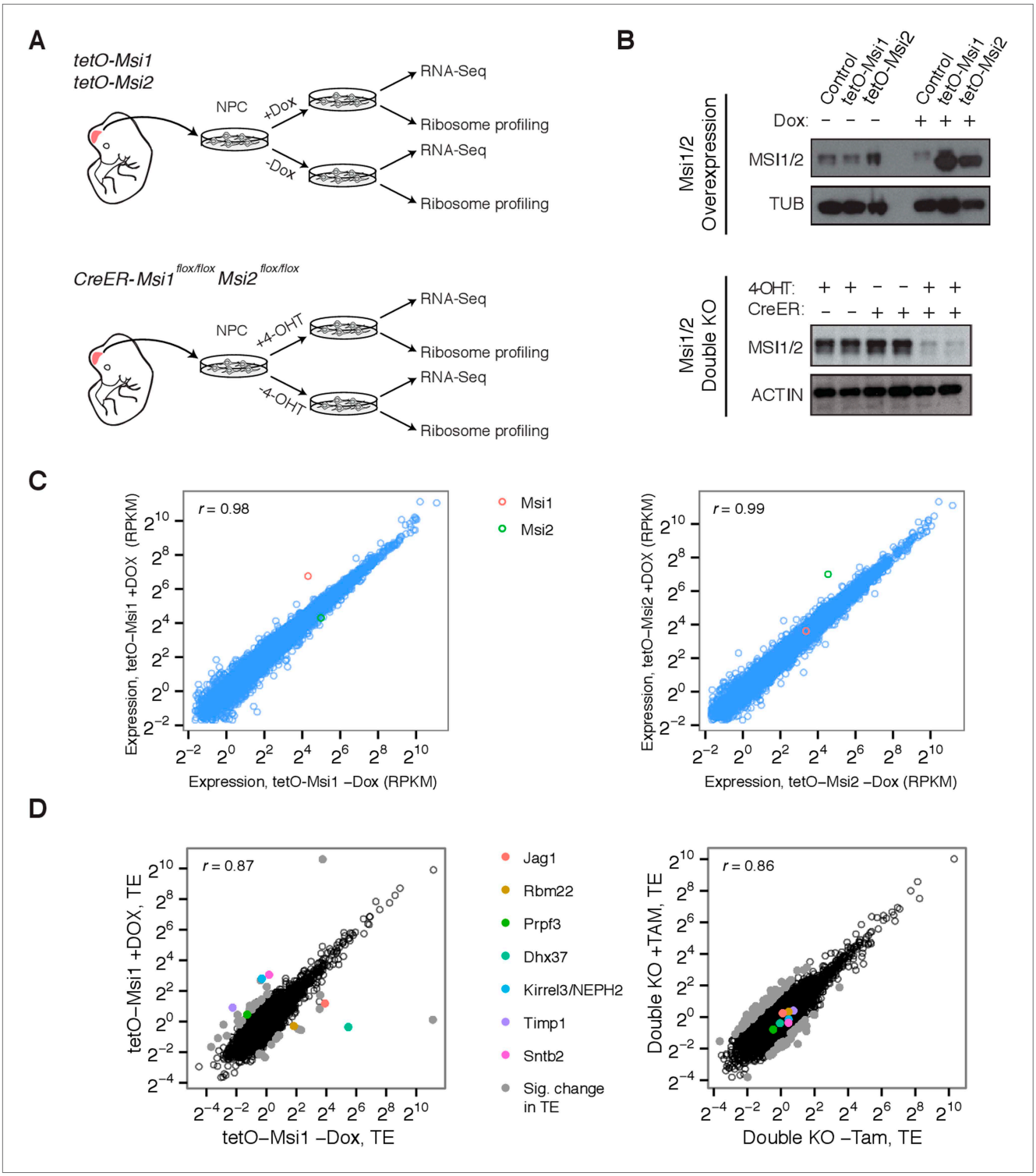

Figure 3. Genetic system for studying effects of Msi loss/gain of function on gene expression. (**A**) Experimental setup and use of *Msi1/2* inducible overexpression and conditional double knockout mice for derivation of neural stem cells, which were then used for ribosome profiling (Ribo-Seq) and mRNA sequencing (RNA-Seq). (**B**) Western blot analysis of Musashi overexpression and knockout in neural stem cells. Overexpression and conditional knockout cells were exposed to Dox and 4-OHT for 72 hr, respectively. (**C**) mRNA-Seq expression values (RPKM) scatters between *Msi1* overexpressing cells and controls (left), *Msi2* overexpressing cells and controls right (72 hr Dox). *Msi1/2* each robustly overexpressed with similar magnitude following Dox. (**D**) Comparison of translational efficiency (TE) values using Ribo-Seq on Msi1 overexpressing cells on Dox (72 hr) vs controls (left) and conditional knockout cells following 4-OHT for 48 hr (right). Colored points indicate select genes with large changes in TE.

DOI: 10.7554/eLife.03915.008

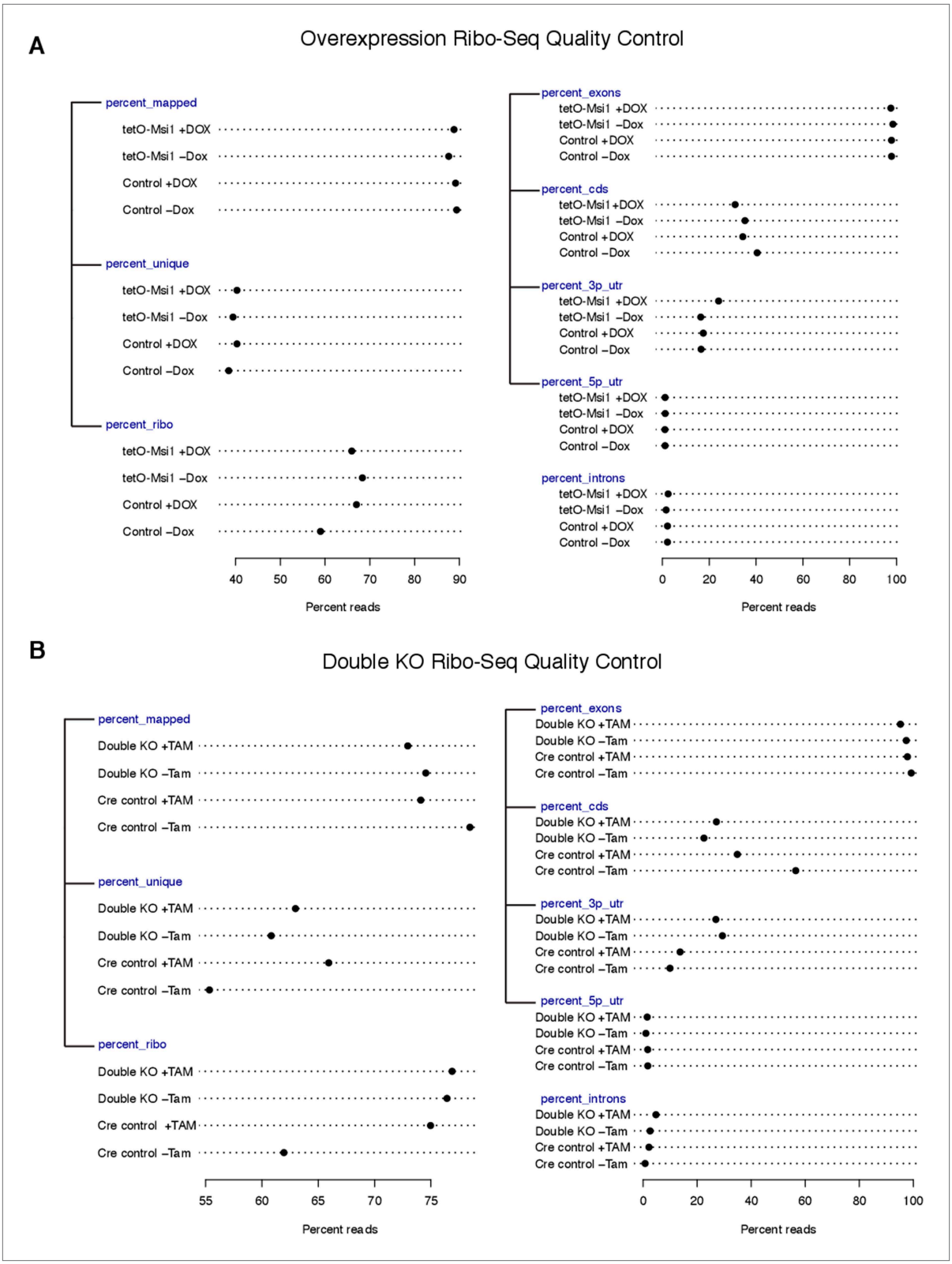

Figure 3—figure supplement 1. Quality control metrics for Ribo-Seq libraries. (**A**) Quality control metrics for overexpression Ribo-Seq libraries. Left panel: percentage of reads mapped to genome, and the percentages of reads that are unique ('percent_unique') and mapping to rRNA ('percent_ribo') out of those mapped. Right panel: percentage of reads mapping to exons ('percent_exons'), and out of those the percentage of reads in CDS regions ('percent_cds'), 3′ UTRs ('percent_3p_utr'), 5′ UTRs ('percent_5p_utr'). Percentage of reads mapping to introns ('percent_introns') also shown. (**B**) Quality control metrics for knockout Ribo-Seq libraries, same format as (**A**).

DOI: 10.7554/eLife.03915.009

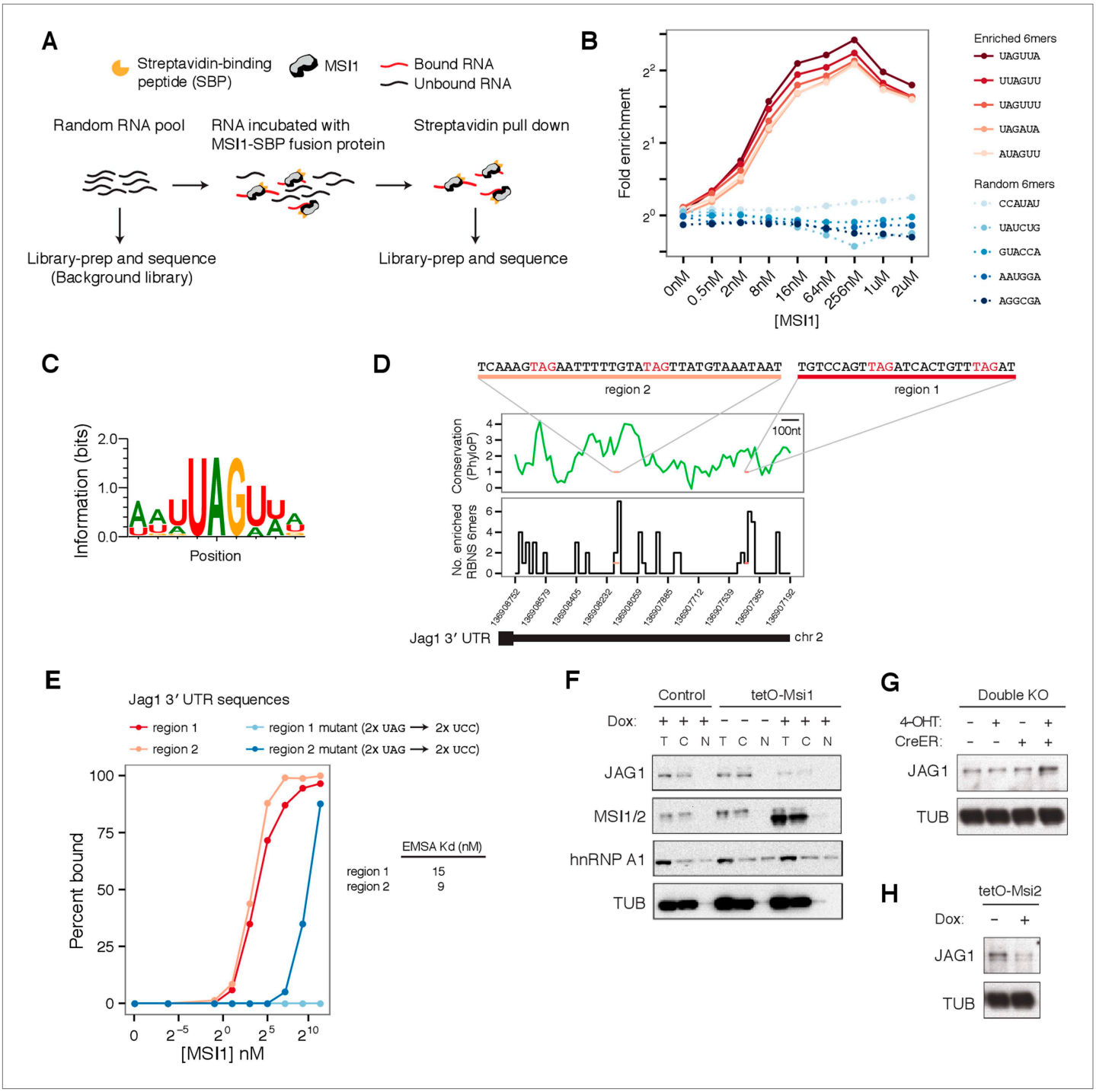

Figure 4. Profiling MSI1 binding preferences by RNA Bind-n-Seq. (**A**) Schemaic of Bind-n-Seq experiment for MSI1 protein. Increased concentrations of MSI1-SBP fusion protein incubated with random RNA pool, pulled by straptavidin pull-down, reverse-transcribed and sequenced. (**B**) Fold enrichment of top five enriched 6mers (red curves) and five randomly chosen 6mers (blue curves) across protein concentrations. (**C**) Binding motif for MSI1. Position-weight matrix generated by global alignment of top 20 enriched 6mers. (**D**) Two sites in *Jag1* 3' UTR, region 1 and region 2, containing a high density of enriched 6mers. Top: PhyloP conservation score for 3' UTR in 20 nt windows (based on UCSC vertebrates multiple alignment). Bottom: number of enriched 6mers from BNS in 20 nt windows of 3' UTR. (**E**) Percent binding of MSI1 protein to region 1 and region 2 (red curves) and mutants where UAG sites are disrupted (blue curves), measured by gel-shift (see ***Figure 4—figure supplement 1***). K_d estimates for region 1 and region 2 are shown (mean of 2 gel-shifts per sequence). (**F**) Western blot analysis of *Jag1* regulation by Msi: top left panel, *Jag1* expression in *Msi1* overexpression cells and controls in cellular fractions (T—total lysate, **C**—cytoplasmic and N—nuclear fractions). *Jag1* is translationally repressed upon induction of *Msi1* and detected

Figure 4. Continued on next page

Figure 4. Continued

only in total and cytoplasmic lysates. hnRNP A1, known to shuttle between the nucleus and the cytoplasm and alpha-Tubulin used as loading controls. (**G**) Increased JAG1 protein levels in double knockout cells. (**H**) Reduced JAG1 protein levels upon *Msi2* overexpression.
DOI: 10.7554/eLife.03915.010

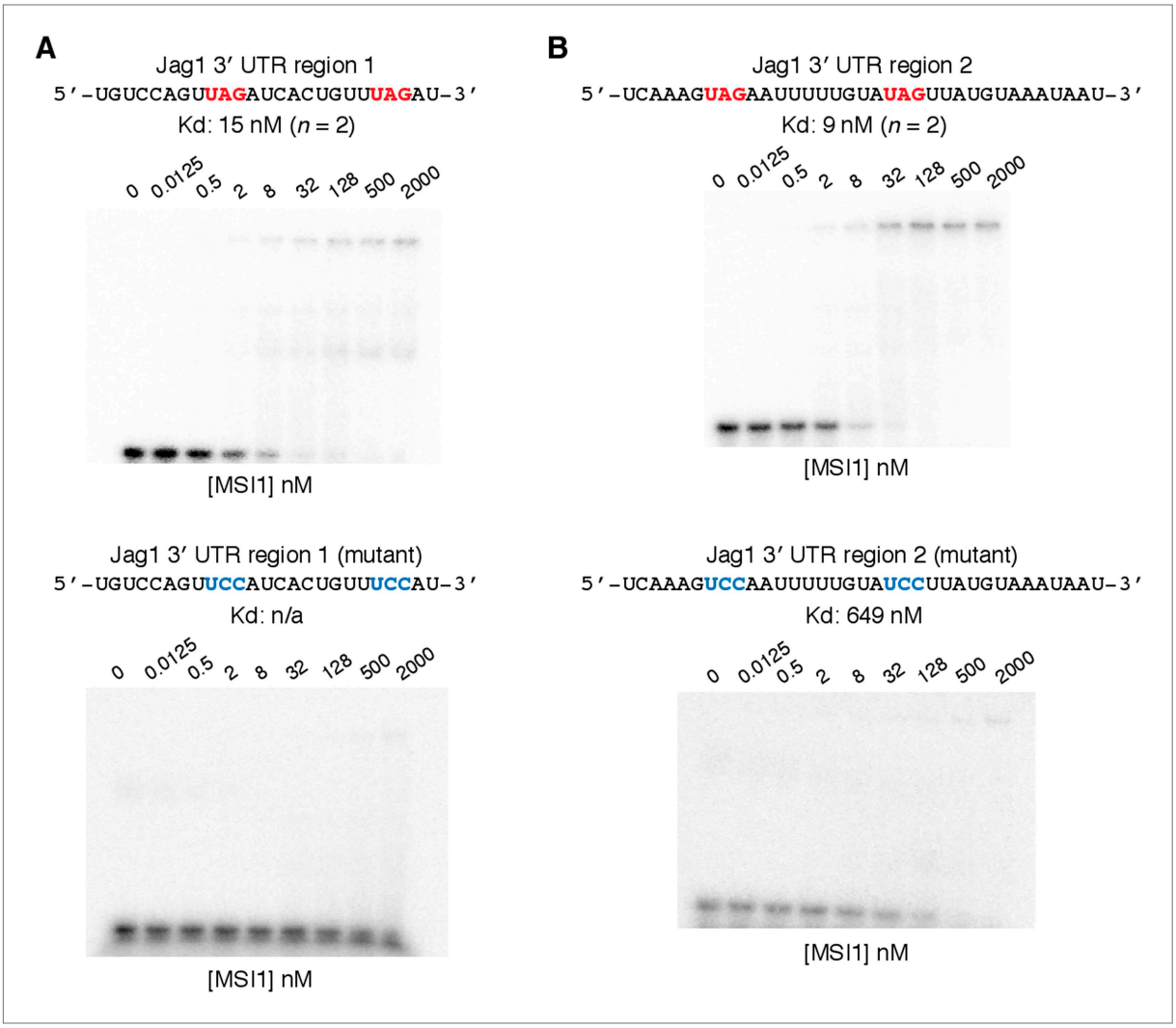

Figure 4—figure supplement 1. Validation by gel-shift of MSI1 binding to Jag1 3' UTR sequences. (**A**) Top: gel-shift MSI1 binding assay for Jag1 3' UTR sequence 1. Kd estimate shown (15 nM) is average of two gel shifts. Bottom: gel-shift for Jag1 3′ UTR sequence 1 mutant, where UAG sites mutated to UCC. Kd cannot be estimated (no binding to mutant could be detected.) (**B**) Top: gel-shift MSI1 binding assay for Jag1 3′ UTR sequence 2. Kd estimate shown (9 nM) is average of two gel shifts. Bottom: gel-shift for *Jag1* 3′ UTR sequence 2 mutant, where UAG sites are also mutated to UCC. Kd for mutant sequence was 649 nM.
DOI: 10.7554/eLife.03915.011

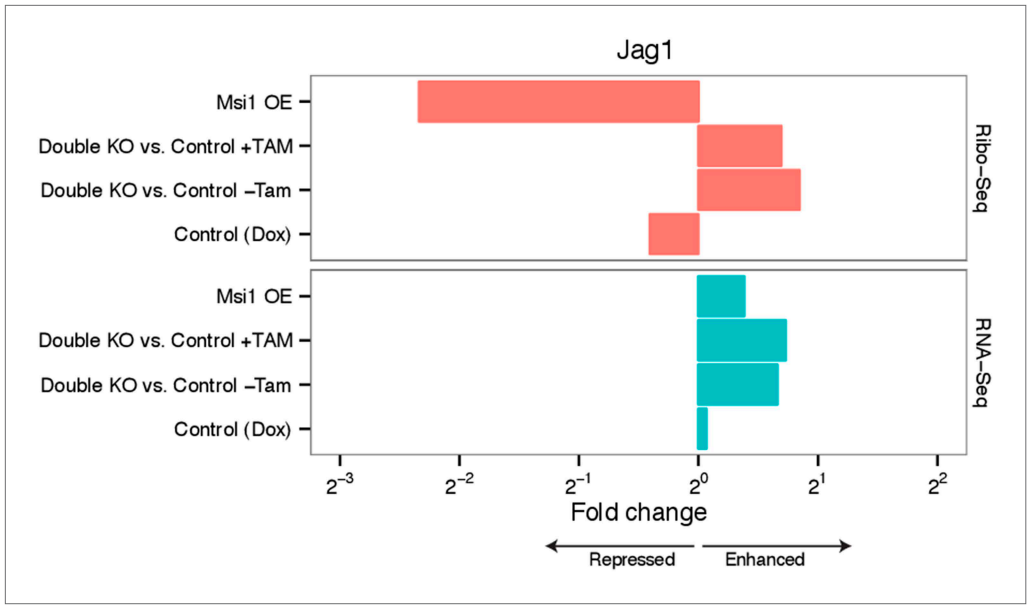

Figure 4—figure supplement 2. Effect of *Msi1* gain and loss of function on *Jag1* mRNA levels and protein expression. Fold-change in *Jag1* expression in Msi1 overexpression and double knockout samples for Ribo-Seq and RNA-Seq experiments.
DOI: 10.7554/eLife.03915.012

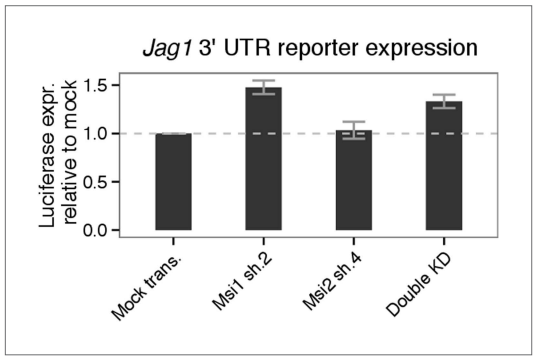

Figure 4—figure supplement 3. Validation of Msi-dependent regulation of *Jag1* protein levels using luciferase reporters containing *Jag1* 3' UTR. Luciferase expression for *Jag1* 3′ UTR reporter transfected into 293T cells. Mean values shown for three biological replicates (±standard deviation). For knockdown lines, *Jag1* 3′ UTR reporter expression was normalized relative to reporter expression in mock transfected 293T cells (represented by dashed horizontal line.) Note that Msi2 sh.4 was effective in knocking down Msi2, but consistently increased Msi1 mRNA levels, and therefore did not reduce total Msi mRNA levels. This likely explains why Msi2 sh.4 293T cells did not show increased *Jag1* 3′ UTR reporter expression.
DOI: 10.7554/eLife.03915.013

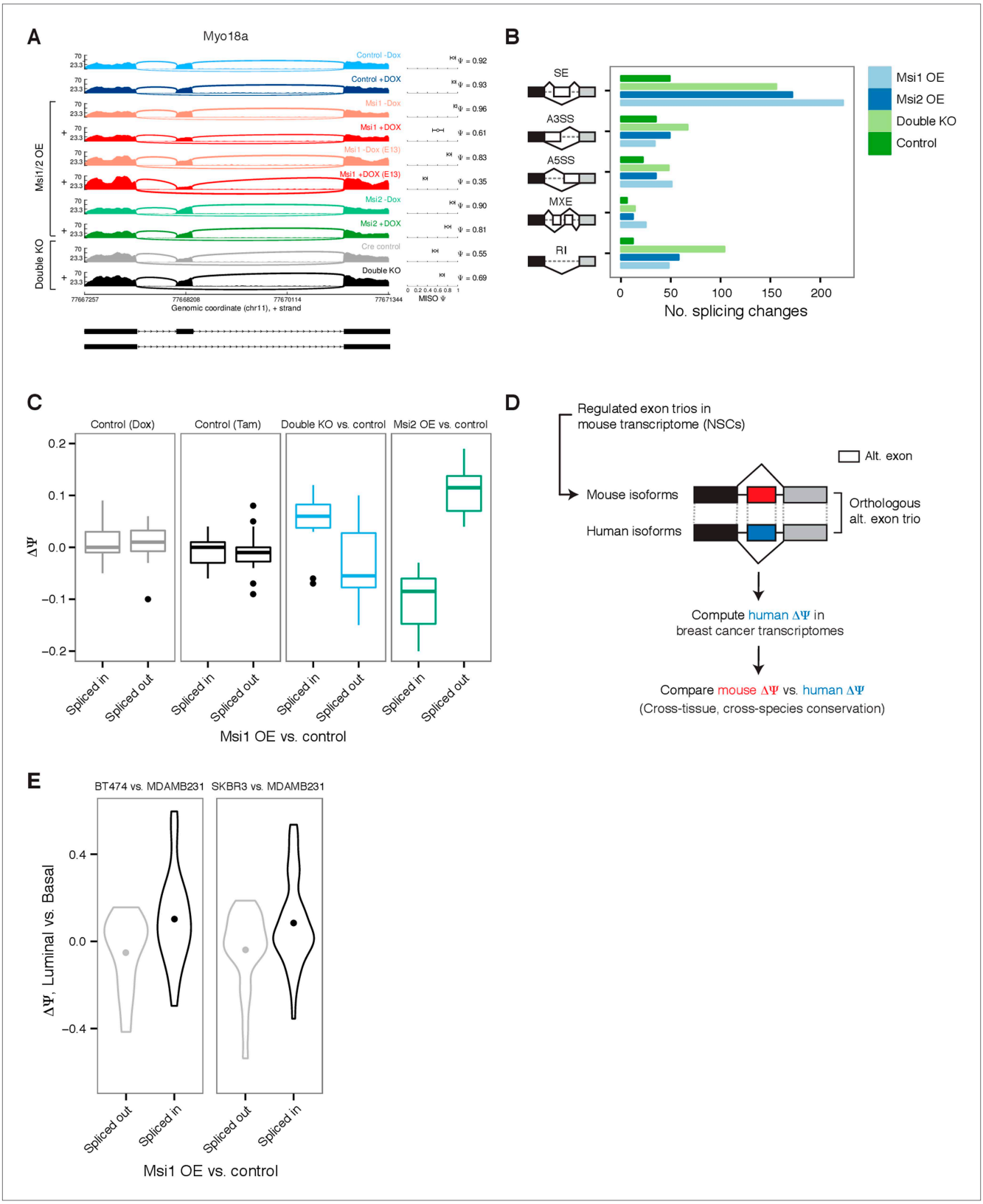

Figure 5. Global impact of Msi proteins on alternative splicing. (**A**) Sashimi plot for *Myo18a* alternative exon 38 with Percent Spliced In (Ψ) estimates by MISO (values with 95% confidence intervals, right panel.) Exon splicing is repressed by *Msi1* overexpression and slightly increased in knockout *Msi1/2* cells. '+' indicates samples treated with Dox/Tam for overexpression/knockout cells, respectively. E12.5 neural stem cells were used for all samples

Figure 5. Continued on next page

Figure 5. Continued

except *Msi1* overexpression for which an additional E13.5 NSC time point was sequenced. (**B**) Number of differential events (MISO Bayes factor ≥10, ΔΨ ≥ 0.12) in each alternative RNA processing category (SE—skipped exons, A5SS—alternative 5′ splice site, A3SS—alternative 3′ splice site, MXE—mutually exclusive exons, RI—retained introns) for *Msi1* overexpression ('Msi1 OE'), *Msi2* overexpression ('Msi2 OE'), double knockouts ('Double KO'), and a Dox control pair ('Control'). (**C**) Comparison of ΔΨ in *Msi1* overexpression vs control binned by direction ('Spliced in' or 'Spliced out', x-axis) to ΔΨ in *Msi2* overexpression cells and in double knockout cells (along with respective Tam and Dox controls, y-axis). (**D**) Computational strategy for identifying human orthologs of alternative exon trios regulated in mouse neural stem cells. Orthologous exon trios were identified by synteny using multiple genome alignments. (**E**) Comparison of ΔΨ mouse alternative exons by *Msi1* (comparing overexpression to control, x-axis) and ΔΨ of their orthologous exon trios in human (comparing luminal and basal cell lines, y-axis). Two pairs of luminal and basal cells compared: BT474 vs MDAMB231 and SKBR3 vs MDAMB231. ΔΨ value distributions summarized by violin plots with a dot indicating the mean ΔΨ value.
DOI: 10.7554/eLife.03915.014

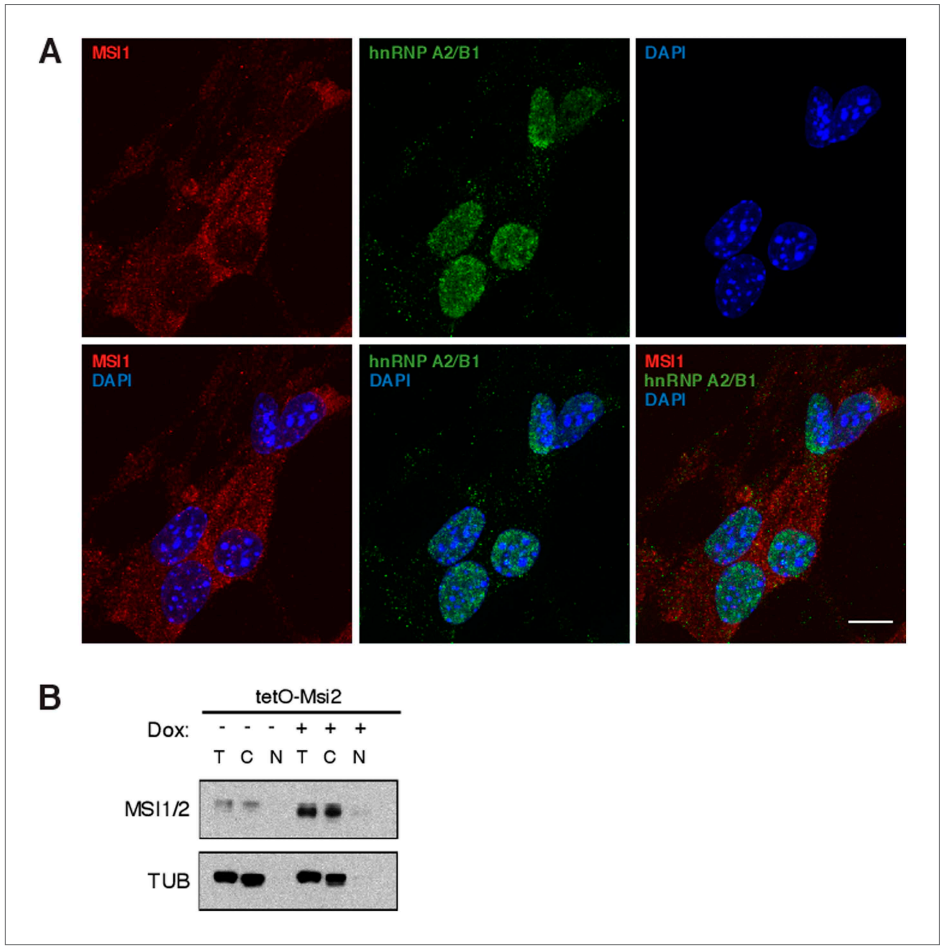

Figure 5—figure supplement 1. Subcellular localization of MSI1 protein in murine NSCs. (**A**) Immunofluorescence staining in mouse neural stem cells for MSI1 (red) and hnRNP A2/B1 (green). MSI1 shows predominantly cytoplasmic localization, while hnRNP A2/B1, a splicing factor, is predominantly nuclear. Confocal maximum Z intensity projections shown, 10 μm scale. (**B**) Western blot analysis for MSI1/2 and alpha-Tubulin (TUB) in total protein lysate (T), cytoplasmic protein lysate (**C**) and nuclear protein lysate (N) in control and Msi2 overexpressing cells.
DOI: 10.7554/eLife.03915.015

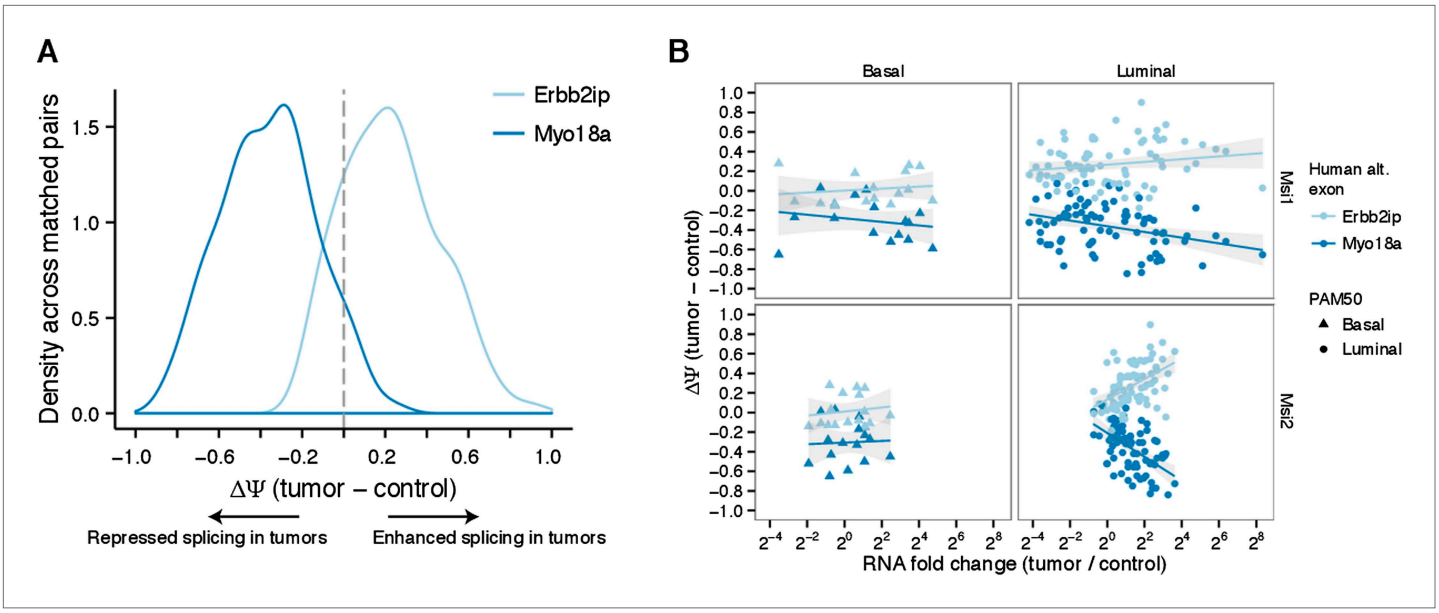

Figure 5—figure supplement 2. Analysis of two conserved Msi-induced splicing changes in breast cancer tumors. (**A**) Distribution of MISO ΔΨ values in matched tumor–control pairs for Erbin (Erbb2ip) exon in light blue and Myo18a in dark blue. Right and left shifts from center (marked by dotted grey line at ΔΨ = 0) indicate tumor-enhanced and tumor-repressed splicing patterns, respectively. (**B**) Comparison of RNA fold changes in matched tumor–control pairs for Msi1 and Msi2 in Basal (left) and Luminal (right) tumors with ΔΨ values for Erbin and Myo18a exons. Points/triangles indicate luminal/basal tumor types determined by PAM50.

DOI: 10.7554/eLife.03915.016

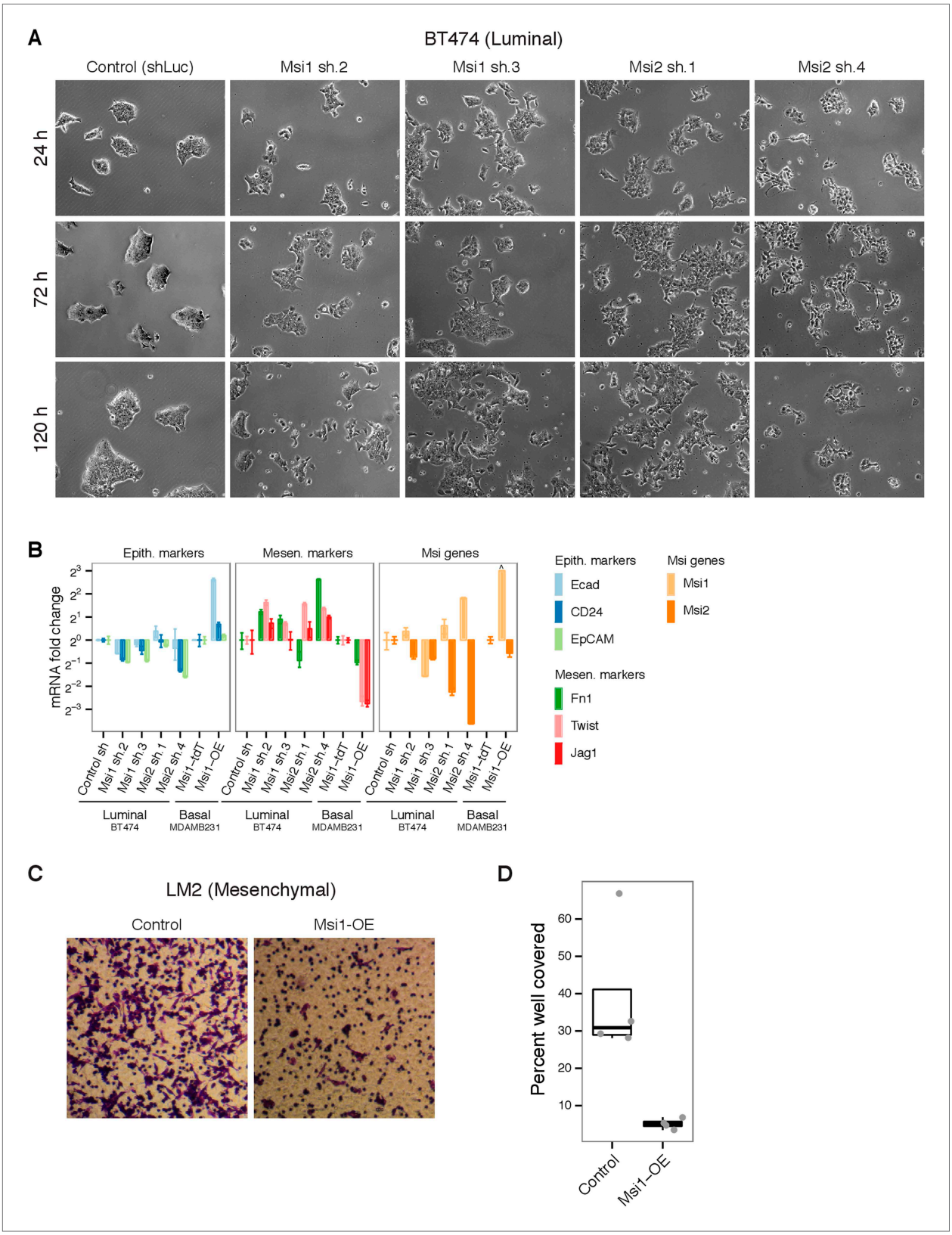

Figure 6. Msi levels alter EMT processes breast cancer cell lines. (**A**) Knockdown of *Msi1/Msi2* in BT474 breast cancer cell line using lentiviruses carrying short hairpins (shRNAs). Brightfield images (10x magnification) shown at 24, 72, and 120 hr after Puromycin-selection. (**B**) mRNA expression of epithelial and mesenchymal markers upon knockdown of *Msi1/Msi2* in epithelial-luminal breast cancer cell line (BT474) and overexpression of *Msi1* in

Figure 6. Continued on next page

Figure 6. Continued

mesenchymal-basal line (MDAMB231). Values plotted are fold changes normalized to GAPDH. For BT474 knockdown, cells infected with hairpin against luciferase were used as control ('Control sh'). For MDAMB231 overexpression, cells infected with tdTomato were used as controls ('Msi1-tdT'). *Msi1* levels were below detection limit in control MDAMB231 cells, therefore *Msi1* fold change in MDAMB231 *Msi1*-overexpression cells (relative to controls) was truncated arbitrarily in plot, indicated by '^'. (**C**) Representative transwell assay image for LM2 control and Msi1-OE breast cancer cells. (**D**) Quantification of percent of well covered in transwell assay for LM2 control and Msi1-OE cells (4 wells per condition, individual well values plotted as dots.).
DOI: 10.7554/eLife.03915.017

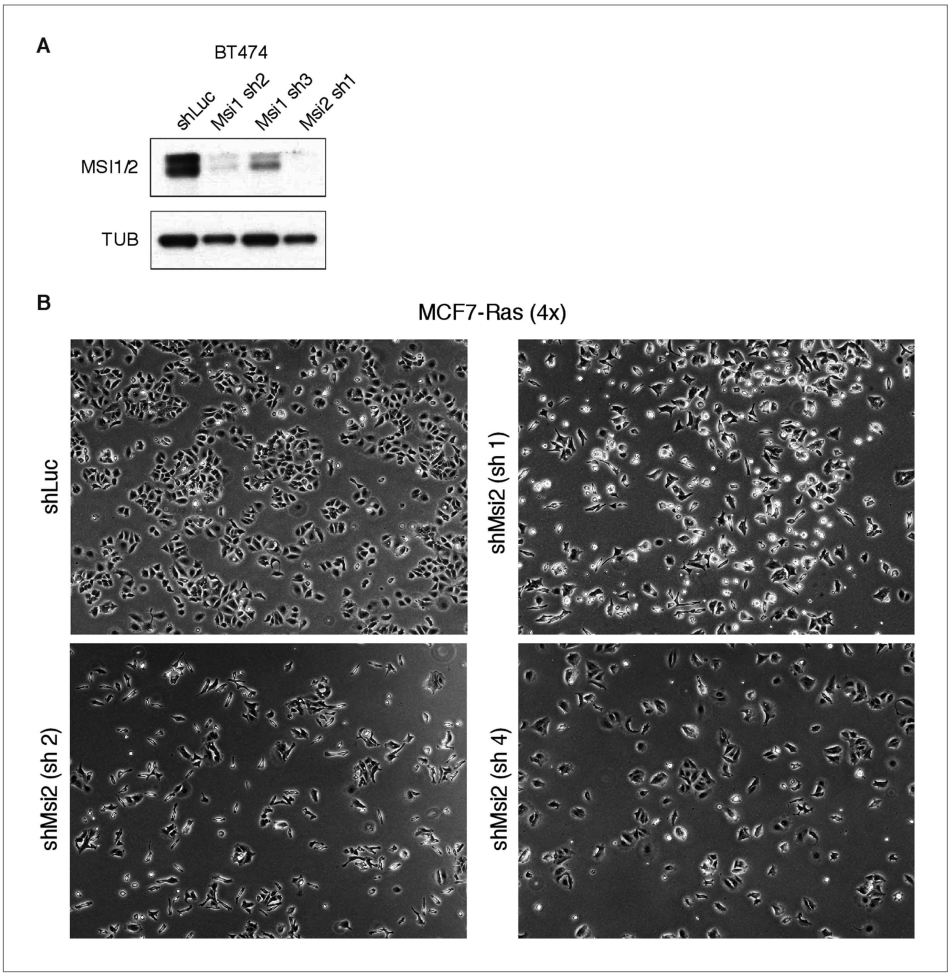

Figure 6—figure supplement 1. Knockdown of *Msi1/2* in breast cancer cell lines. (**A**) Western blot for BT474 cells with control (shLuc) or Msi1/2 targeting hairpins. (**B**) Morphology of MCF7-Ras cells upon Musashi knockdown.
DOI: 10.7554/eLife.03915.018

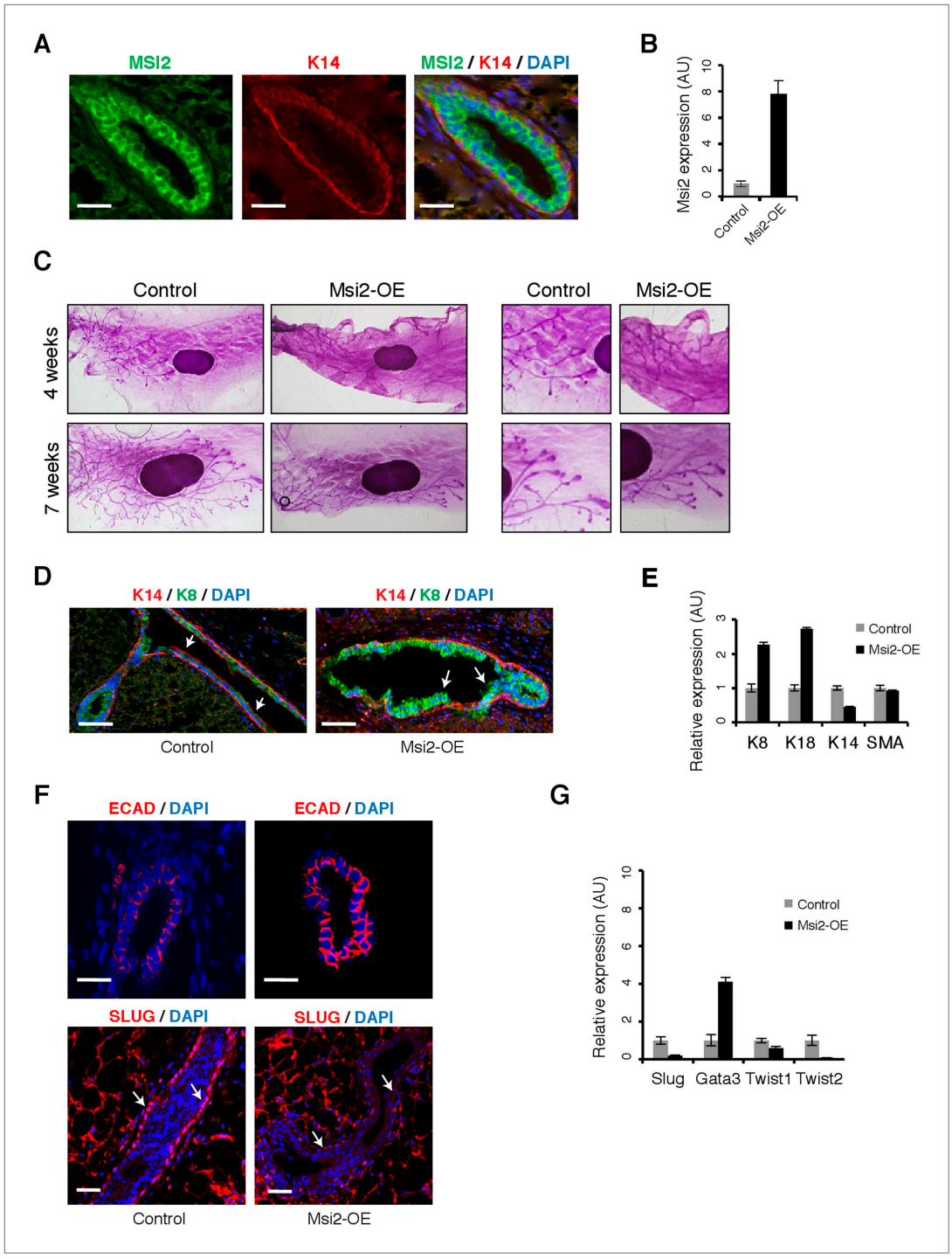

Figure 7. Msi2 activation represses EMT and expands mammary luminal cell layer in vivo. (**A**) Immunostaining for MSI2, K14, and DAPI in control sections of mammary gland. Scale bar: 50 μm (**B**) qRT-PCR for *Msi2* in mammary epithelial cells from control and *Msi2* overexpressing mice ('Msi2-OE'). (**C**) Whole mount stain for mammary glands from control and *Msi2* overexpressing mice (left: low magnification, right: high magnification.) (**D**) Immunostaining for K14, K8, and DAPI in mammary gland sections from control and *Msi2* overexpressing mice. Scale bar: 100 μm (**E**) qRT-PCR for luminal markers (K8, K18), basal markers (K14), and smooth-muscle Actin (SMA) in mammary epithelial cells from control and *Msi2* overexpressing mice. (**F**) Staining for E-cadherin (ECAD) (top) and EMT-marker SLUG (bottom) in mammary glands from control and *Msi2* overexpressing mice. Luminal cell layer is expanded upon Dox (arrows). Scale bar: 100 μm. (**G**) qRT-PCR for Slug, Gata3, Twist1, Twist2 in mammary epithelial cells from control and *Msi2* overexpressing mice. Slug expression in basal cell layer is reduced upon Dox (arrows). Scale bar: 50 μm.

DOI: 10.7554/eLife.03915.019

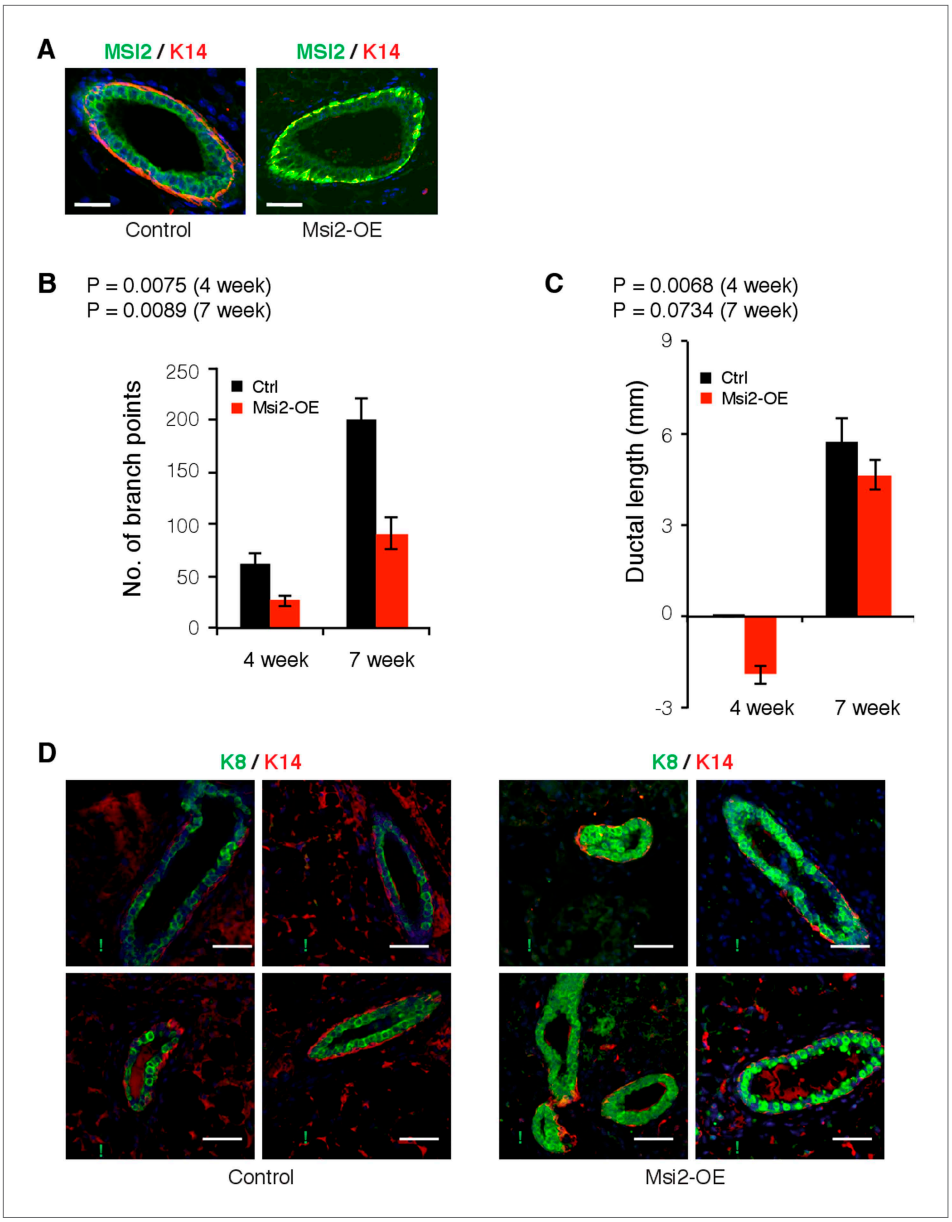

Figure 7—figure supplement 1. *Msi2* overexpression in mouse mammary gland alters mammary duct morphology. (**A**) Msi2 expression in mammary glands co-stained with basal cell marker K14 in control and Msi2 overexpressing mice. (**B**) Quantification of number of branch points in control and Msi2 overexpression mice. Student's *t*-test was used to compute p-values. (**C**) Lengths of longest mammary ductal branches (measured from Center of Lymph Node, CLN) for control and Msi2 overexpression mice. CLN defined as '0': negative length values indicate that longest ductal branch ends prior to start of CLN, positive length values indicate that longest ductal branch grew past center of CLN. Student's *t*-test was used to compute p-values. (**D**) Co-staining for luminal cell marker K8 and basal cell marker K14 in control (left) and Msi2 overexpressing (right) mice.
DOI: 10.7554/eLife.03915.020

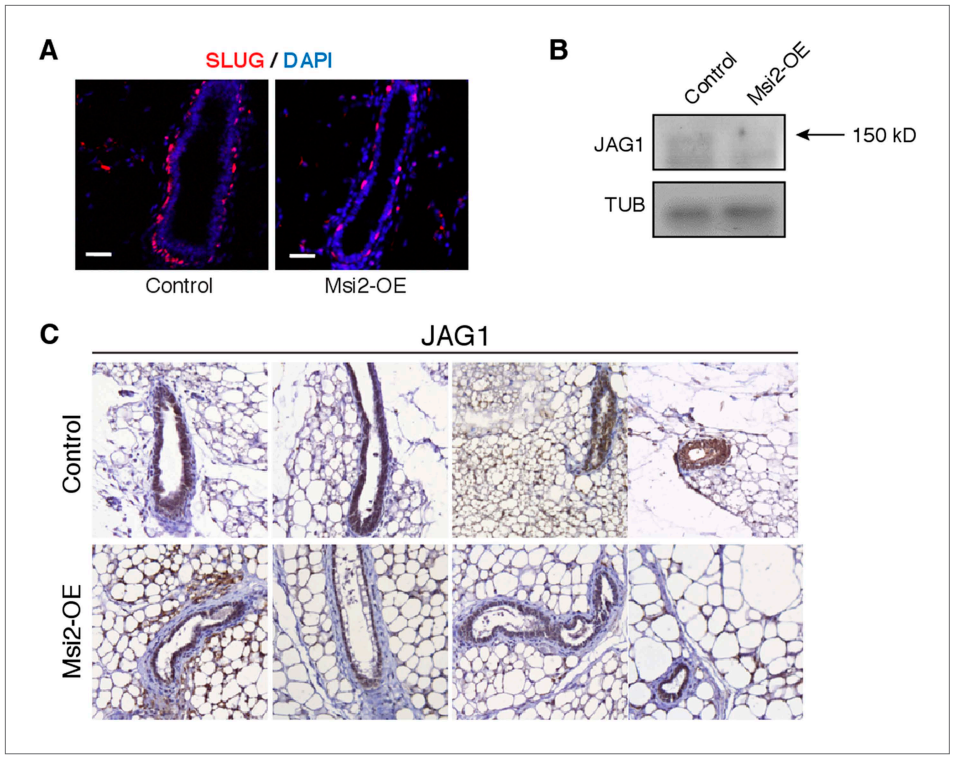

Figure 7—figure supplement 2. *Msi2* overexpression in mouse mammary gland represses Slug and Jag1. (**A**) Staining for EMT marker Slug in control and Msi2 overexpressing mice. Scale bar: 50 μm. (**B**) Western blot for JAG1 protein in mammary epithelial cells of control and Msi2 overexpressing mice 7 weeks after induction with Dox. Arrow indicates expected JAG1 band (150 kD). (**C**) Immunohistochemistry for JAG1 protein in mammary gland from control and Msi2 overexpressing mice 7 weeks after induction with Dox.
DOI: 10.7554/eLife.03915.021

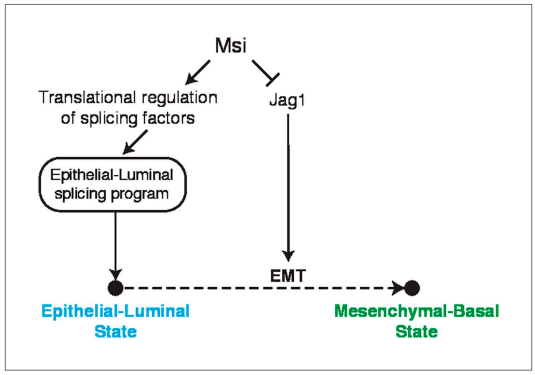

Figure 8. Model for Msi roles in regulation of cell state. Model for Msi role in the control of the epithelial state. We show that Msi represses translation of *Jag1*, a positive regulator of Notch and EMT. We also show that Msi promotes expression of an epithelial-luminal splicing program, which we hypothesize occurs through translational regulation of splicing factors. In the model, both the direct regulation of *Jag1* and indirect regulation of splicing contribute to maintenance of an epithelial-luminal cell state and inhibition of EMT.
DOI: 10.7554/eLife.03915.022