Materials and Methods

Cells and Tissues – Several independent isolates of primary HCASMC and human umbilical vein endothelial cells (HUVEC) were maintained in growth medium supplied by the manufacturer (Invitrogen). HUVEC were obtained from the Cell Culture Core in the Aab CVRI and plated onto gelatin-coated plates/chambers. HeLa, HEK293, SKLMS (a human leiomyosarcoma cell line of uterine SMC origin), LnCAP, and MCF7 cells were grown in medium as specified by the manufacturer (ATCC). Human tissue RNA samples were obtained from a commercial source (Zyagen). Dated human plasma was obtained through the University of Rochester Medical Center Blood Bank.

RNA-Sequencing Analysis – Total RNA was isolated from HCASMC using RNeasy extraction kit (Qiagen) under normal growth conditions or where *SENCR* was knocked down for 3 days with 25 nM of either a dicer substrate RNA to exon 2 (ds-*SENCR*-5, Table I in the online only Data supplement) or a control dicer substrate RNA. Following bioanalyzer quality control confirmation, RNA-seq was performed on the polyadenylated fraction using Illumina Genome Analyzer IIx platform at the University of Rochester Medical Center Genomics Research Center (http://www.urmc.rochester.edu/fgc/). Single-end sequencing was done at a depth of 20 million reads per replicate (n=3). Pre-processing of raw sequence reads included demultiplexing with CASAVA 1.8.2, transcript trimming of contaminating sequences with Sequence Cleaner (http://sourceforge.net/projects/seqclean/), removal of vector sequences with UniVec database (http://www.ncbi.nlm.nih.gov/VecScreen/UniVec.html), and FASTQ quality trimming using the FASTX Toolkit (http://cancan.cshl.edu/labmembers/gordon/fastx_toolkit/index.html). SHRiMP2.2.3 was used to align sequence reads to annotated transcripts on the UCSC Reference Genome (build GRCh37/hg19). Quantitative analysis, including the statistical analysis of differentially expressed genes was done with Cufflinks 2.0.2 and Cuffdiff2 (http://cufflinks.cbcb.umd.edu). For the *SENCR* knockdown RNA-seq experiment, the Benjamini-Hochberg method was applied for multiple test correction (FDR < 0.05). Data output files such as Volcano plots and scatterplots were generated with cummerbund (http://compbio.mit.edu/cummeRbund/). Gene ontology (GO) analysis was done using DAVID [1]. All RNA-seq data were deposited into NCBI's Gene Expression Omnibus (GSE51878).

Bioinformatics Methods for Identifying Novel LncRNAs – RNA-seq reads were aligned to the human genome (hg19) using TopHat 1.4 [2]. In this analysis, two iterations of TopHat alignment were performed in order to maximize the chance of identification of exon-exon junctions. The alignment data were used to define novel lncRNAs following the method described for lncRNA identification [3]. The aligned data for each sample were used independently by two complementary programs, Scripture [4] and Cufflinks [5], for assembling transcripts independent of gene annotation. We determined the threshold for the read coverage of each transcript across all samples by optimizing the sensitivity and specificity for identifying full length versus partial length transcripts of protein coding genes in RefSeq [3]. In the end, we kept assembled transcripts present in both Scripture and Cufflinks outputs, and with ≥2 exons, ≥200 bp and ≥ 2.7 read coverage as reads below this threshold were deemed unreliable in predicting exon structure. Next, we eliminated all transcripts that had an exon overlapping in the same strand with known transcripts from available databases. We then computed the coding potential of all remaining putative novel transcripts using PhyloCSF [6] and removed any transcripts containing an open reading frame with PhyloCSF score ≥ 100 across any of three reading frames. We further removed transcripts that were homologous to known protein coding domains in the Pfam database (release 26; both PfamA and PfamB) [7] using the program HMMER-3 (e-value = 10) [8]. Lastly, we computed expression values (in FPKM, fragments per kilobase of exon per million fragments mapped) of all remaining transcripts, together with all

coding and known non-coding genes. The final list of lncRNAs (Table II in the online only Data supplement) comprises transcripts with FPKM >0.7.

Dicer-Substrate RNA Knockdown – Several dicer-substrate RNA (dsRNA) molecules to different exons of *SENCR* or control dsRNA (ds-Ctrl) were synthesized (Integrated DNA Technologies) and pre-tested in HCASMC and HUVEC (see Table I in the online only Data supplement for list of all DNA molecules used in this study). The ds-Ctrl does not target known transcribed sequences in human, mouse, or rat genomes and thus serves as a negative control for dsRNA transfections. Briefly, cells were Lipofectamine-transfected with each dsRNA (20-30 nM) for three days and then total RNA or protein was isolated for further analysis. Results were confirmed using at least two independent dsRNA constructs to *SENCR*, in up to five independent isolates of HCASMC and HUVEC, often times by multiple investigators.

Gene Expression Assays – Total RNA was isolated using the RNeasy kit (Qiagen). RNA integrity was assessed by spectrophotometery (NanoDrop, Thermo Scientific) and agarose gel electrophoresis. cDNA was synthesized from 1μg of total RNA using iScript (Bio-Rad) plus random decamers and/or an oligodT primer. RT-PCR was performed using Platinum PCR Supermix (Invitrogen) with a MyCycler thermocycler (BioRad) and PCR products were resolved in a 1% agarose gel. Some gels shown were adjusted uniformly in Photoshop using the "invert" function. For lncRNA validations, we included a no RT step that revealed little to no products indicating authentic polyA+ RNAs were amplified as opposed to contaminating genomic DNA (data not shown). Quantitative RT-PCR was performed using IQ SYBR Green Supermix with a MyiQ single color real-time PCR detection system (BioRad). Experiments shown are representative of multiple independent experiments using different lots of HCASMC and HUVEC, performed by separate investigators to ensure quality control and accurate interpretation of observed changes in gene expression.

Western Blotting – Total protein was isolated from HCASMC following ds-*SENCR* or ds-Ctrl knockdown using RIPA buffer (50 mM Tris, pH 8.0, 150 mM NaCl, 0.1% SDS, 1.0% NP-40, 0.5% sodium deoxycholate and Roche protease inhibitor cocktail) and resolved in acrylamide gels for Western blotting as previously done [9]. Antibodies were LMOD1 (ProteinTech, 1:2000), TAGLN (Abcam, 1:4000), CNN1 (DAKO, 1:2000), ANPEP (R&D, 1:1000), FLI1 (Santa Cruz, 1:250) and PPIA (Santa Cruz, 1:2000), used as an internal control.

RNA Fluorescence In Situ Hybridization (FISH) – RNA FISH with single-molecule sensitivity was performed using QuantiGene® (QG) ViewRNA ISH Cell Assay reagents (Affymetrix) based on branched DNA technology [10]. Custom probe oligonucleotide pair pools specific for *SENCR* long and short isoforms were designed and synthesized by Affymetrix as "Type 6" (50 pairs targeting all 3 exons; product ID, VA6-14704) and "Type 4" (19 pairs excluding exon 2; product ID, VA4-14958), respectively. A probe pair pool specific for human *PP1B* housekeeping mRNA (VA1-10148, "Type 1") and the lncRNA, *NEAT1* (VA1-12621, "Type 1") were used as cytoplasmic and nuclear controls, respectively, to assist in interpreting spatial localization of *SENCR* RNA. HUVEC (± *SENCR* knockdown) were grown on acid-washed #1.5 glass cover slips (Thermo Scientific) to 70%-80% confluence, washed with PBS, and fixed for 30 min in fresh 0.45μm-filtered 4% paraformaldehyde (Electron Microscopy Sciences) dissolved in Dulbecco's PBS without $CaCl_2$ and $MgCl_2$ (Invitrogen). RNA-FISH was performed with minor modifications from the manufacturer's protocol as follows: After permeabilization using QG Detergent Solution, cells were treated with 0.5% Triton X-100/PBS for 5 min at room temperature. Partial protease digestion was carried out with a 1:6,000 dilution of QG Protease K for 10 min at room temperature. Coverslips were incubated with primary probe pair sets (3-color multiplexing) or QG Probe Set Diluent as negative control at 40°C for 3 hr. Pre-amplifiers were incubated for an extended period of 1 hr. Between probe set incubations, cells were washed 4 times each in QG Wash Buffer for a total of 10 min. After counter-staining with DAPI, coverslips were mounted in home-made anti-fade mounting medium (www.spectorlab.cshl.edu/protocols) and sealed with nail polish.

Image Acquisition and Analysis – Cells were imaged on a DeltaVision Core system (Applied Precision) based on an inverted IX-71 microscope stand (Olympus) equipped with a 60x U-PlanApo 1.40 NA oil immersion lens (Olympus). Images were captured using a CoolSNAP HQ CCD camera (Photometric) as 10μm image stacks with a z-spacing of 0.2μm at a 1x1 binning. Stage, shutter and exposure were controlled through SoftWorx (Applied Precision). Image deconvolution was performed using SoftWorx. Parameters for acquisition and post-acquisition processing were identical for all coverslips. Analysis was done on individual image stacks in 3D space by counting the number of *SENCR* hybridization signals divided by the number of cells in each field of view (≥50 cells in ≥10 randomly chosen fields per experiment). In some experiments, we employed two fluorescently-tagged probe sets (above) in the absence or presence of *SENCR* knockdown to further confirm spatial localization. Only signals that showed overlap of QG "Type 4" and "Type 6" probe sets were considered, thus minimizing potential false-positive signal counts when using single color analysis.

Luciferase Assay – The putative promoter of *SENCR* was defined through 5' RACE (Ambion). Several constructs of varying 5' and 3' length were PCR amplified from HCASMC genomic DNA, cloned into the pGL3 Basic Vector (Promega), and sequence confirmed (URMC Genomics Research Center). HUVEC were plated in 12-well dishes and grown to ~60%-70% confluency and transfected with various *SENCR* promoter constructs or a *DLL4* reporter gene as a positive control. Lipofectamine was used in transfections and the normalized average luciferase activity calculated for each reporter plasmid.

Migration Assays – HCASMC were plated onto coverslips and transfected with either ds-Ctrl or dsRNAs targeting non-overlapping sequences in *SENCR*. Sixty hr after transfection, cells were "scratched" with a sterile P200 pipette tip and the culture medium immediately changed to DMEM containing 10% FBS. 12 hr after scratch wounding, the cells were fixed and stained with Alexa Fluor® 660 Phalloidin and DAPI (Invitrogen) according to manufacturer's instructions. The cells were then imaged by confocal microscopy (Olympus FV1000) and the migratory index (percentage of cells that migrated into the time 0 wound area) was calculated using NIH Image J software. An independent assay for migration was done using a modified Boyden chamber (Corning). Briefly, HCASMC were transfected for three days with either ds-Ctrl or ds-*SENCR* (25 nM) and then seeded into a 24-well Boyden chamber plate. Cells were then serum-deprived overnight and subsequently treated either with PDGF-BB (25 ng/ml) or vehicle for 6 hr. Cells were then fixed, stained with hematoxylin, and imaged with an inverted phase contrast microscope. Migration assays are representative of multiple experiments performed independently by two authors (RDB and XL).

Statistical Analysis – Student's t-test or one way ANOVA followed by Tukey's post-hoc test were used to determine statistical significance of the means (± standard deviation) and graphs were plotted (Graph-Pad Prism 5.0). Statistical significance was assumed at $p < 0.05$.

1. Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. David: Database for annotation, visualization, and integrated discovery. *Genome Biology*. 2003;4:3
2. Trapnell C, Pachter L, Salzberg SL. Tophat: Discovering splice junctions with rna-seq. *Bioinformatics*. 2009;25:1105-1111
3. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. Integrative annotation of human large intergenic noncoding rnas reveals global properties and specific subclasses. *Genes Dev*. 2011;25:1915-1927
4. Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, Fan L, Koziol MJ, Gnirke A, Nusbaum C, Rinn JL, Lander ES, Regev A. Ab initio reconstruction of cell

type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincrnas. *Nature Biotechnology*. 2010;28:503-510
5. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by rna-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*. 2010;28:511-515
6. Lin MF, Jungreis I, Kellis M. Phylocsf: A comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics*. 2011;27:i275-282
7. Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer EL, Bateman A. The pfam protein families database. *Nucleic acids research*. 2008;36:D281-288
8. Finn RD, Clements J, Eddy SR. Hmmer web server: Interactive sequence similarity searching. *Nucleic acids research*. 2011;39:W29-37
9. Nanda V, Miano JM. Leiomodin 1: A new serum response factor-dependent target gene expressed preferentially in differentiated smooth muscle cells. *The Journal of biological chemistry*. 2012;287:2459-2467
10. Battich N, Stoeger T, Pelkmans L. Image-based transcriptomics in thousands of single human cells at single-molecule resolution. *Nature methods*. 2013;10:1127-1133