Combined RNAi and localization for functionally dissecting long noncoding RNAs

Debojyoti Chakraborty1,2, Dennis Kappel1,2, Mirko Theis1–3, Anja Nitzsche1,2, Li Ding1,2, Maciej Paszkowski-Rogacz1,2, Vineeth Surendranath2, Nicolas Berger1,2, Herbert Schulz4, Kathrin Saar4, Norbert Hubner4 & Frank Buchholz1,2

Whereas methods to comprehensively study cellular roles of protein-coding genes are available, techniques to systematically investigate long noncoding RNAs (lncRNAs), which have been implicated in diverse biological pathways, are limited. Here we report combined knockdown and localization analysis of noncoding RNAs (c-KLAN) that merges functional characterization and localization approaches to study lncRNAs. Using this technique we identified transcripts that regulate mouse embryonic stem cell identity.

In eukaryotic genomes a wide array of both protein and non–protein coding transcripts are transcribed, with the latter receiving a great deal of scientific attention only in recent years. LncRNAs are non–protein coding transcripts ranging from a few hundred bases to several kilobases and are known to participate in various biological processes such as X-chromosome inactivation, modulation of epigenetic states at the chromatin level, maintenance of embryonic stem cell (ESC) identity, transcriptional control and the regulation of disease states1. Although tagging and imaging approaches in combination with loss-of-function studies have been very successful in large-scale studies of protein-coding genes2–3, similar methods to visualize and phenotypically characterize long noncoding transcripts have not been described. Endoribonuclease-prepared siRNAs (esiRNAs) have proven to be particularly suitable for RNAi screening4–9 because they efficiently deplete the target transcript without causing prominent off-target effects owing to their inherent complex pools of siRNAs10,11. The siRNAs that make up this pool exist in comparable amounts and have the same on-target competence. Therefore, silencing capacity for the intended target is additive, whereas off-target effects are diluted out. We reasoned that the same advantages should apply for silencing lncRNAs. Here we report the generation of an esiRNA library targeting 594 IncRNAs for loss-of-function screens in mouse cells. Simultaneously, we adapted the esiRNA production protocol to implement a strategy for seamless, renewable generation of specific riboprobes to determine the localization of IncRNAs in cells (Fig. 1a,b).

To design esiRNAs to IncRNAs, we used ‘design and quality control of (e)siRNAs’ (DEQOR), an algorithm that has been used to predict efficient and specific esiRNAs for protein-coding transcripts12. To investigate whether esiRNAs can be used to deplete lncRNAs, we transfected mouse ESCs with esiRNAs targeting 55 different IncRNAs and measured changes in IncRNA expression by quantitative real-time (RT)-PCR. We observed knockdowns exceeding 50% for 38 of the 55 (69%) transfected esiRNAs, indicating that many IncRNAs are susceptible to silencing with esiRNAs (Fig. 1c). To investigate whether DEQOR optimization improved silencing of IncRNAs, we compared the silencing efficacy of DEQOR-optimized versus non–DEQOR-optimized esiRNAs for ten randomly chosen IncRNAs. Notably, nine of these (90%) DEQOR-optimized esiRNAs had an increased knockdown efficacy (Fig. 1c). The DEQOR-optimized esiRNAs likewise performed better than most tested chemically synthesized siRNAs (Fig. 1c).

Difficulties in annotating IncRNA transcripts present an obstacle for generating high-quality siRNA and short hairpin (sh)RNA libraries. Furthermore, siRNA and shRNA libraries are usually based on chemical synthesis and hence validation of IncRNA expression is typically not included in the experimental design, increasing time and costs for library generation and screening. In contrast, esiRNAs are generated from cDNA and are thus only generated to authentic IncRNAs. To generate a lnc-esiRNA library we investigated the expression of 1,386 reported mouse lncRNAs experimentally (Supplementary Table 1 and Online Methods). We synthesized esiRNAs for 594 of these (Supplementary Table 2), providing a high-quality resource to study mouse IncRNAs.

The standard technology to localize IncRNAs is fluorescence in situ hybridization (FISH), using a labeled antisense probe directed against the IncRNA of interest13. We modified the standard esiRNA production protocol to include different RNA polymerase promoter sequences at the 5′ and 3′ ends of the fragments, making it possible to generate either sense or antisense labeled riboprobes for RNA FISH or dsRNA for esiRNA synthesis (Fig. 1a,b and Online Methods). Because the pipeline allows the synthesis of both sense and antisense probes, it also gives an opportunity to investigate cis-natural antisense transcripts. As a proof of principle, we randomly selected 26 amplicons from the 594 IncRNA-specific PCR products, generated labeled probes and hybridized them to mouse ESCs (Online Methods). We detected distinct signals for 20 of the 26 probes (Fig. 1d–i, Supplementary Fig. 1).

1University of Technology Dresden, University Hospital and Medical Faculty Carl Gustav Carus, Department of Medical Systems Biology, Dresden, Germany.
2Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany. 3Eupheria Biotech GmbH, Dresden, Germany. 4Max Delbruck Center of Molecular Medicine, Berlin, Germany. Correspondence should be addressed to F.B. (buchholz@mpi-cbg.de).

RECEIVED 21 JUNE 2011; ACCEPTED 11 JANUARY 2012; PUBLISHED ONLINE 12 FEBRUARY 2012; DOI:10.1038/NMETH.1894
Supplementary Table 3 and Supplementary Notes) and observed diverse localization patterns at varying expression levels for different IncRNAs. These ranged from dispersed spots throughout the nucleus to spots in certain nuclear structures (for example, nucleolus, nuclear periphery and nucleolus), to distinct subcellular regions in the nucleus and cytoplasm (Fig. 1d–f and Supplementary Fig. 2). The wide variety of different localizations of IncRNAs suggests a diverse role of these transcripts in different biological processes. We conclude that c-KLAN is a versatile technology for functional and localization studies on IncRNAs.

To test the utility of c-KLAN in a pilot screen, we assayed the Inc-esiRNA library for potential modulators of ESC identity, using the Oct4-GFP assay (Fig. 2a). We transfected Oct4-GFP cells, which express GFP driven by the Oct4 promoter, with the Inc-esiRNA library in 96-well plate format and assayed them for loss of GFP expression 96 h after transfection using a high-throughput FACS-based readout (Fig. 2a, Supplementary Fig. 3a and Online Methods). We selected for verification IncRNAs that scored above or below the threshold in at least two replicates and did not strongly affect cell viability (Supplementary Tables 4 and 5). Six IncRNAs qualified for validation with secondary, independent esiRNAs (Supplementary Fig. 3b). Of these, we identified three IncRNAs, here referred to as pluripotency-associated noncoding transcripts 1–3 or Panct1–3, for 2900057E15Rik, Gm5101 and AK081885, respectively, that satisfied stringent selection criteria based on average scores from six replicates (Online Methods, Supplementary Fig. 4 and Supplementary Table 6). We characterized the role of Panct1 in ESC pluripotency because it showed the strongest and most robust phenotype in the Oct4-GFP assay.

Panct1 is a 4-kilobase transcript encoded on chromosome X and is transcribed within the first intron but independently (data not shown) of the RefSeq gene A830080D01Rik. It is expressed in mouse ESCs at low levels, is polyadenylated and is not an alternative exon of the host protein-coding gene (Supplementary Figs. 5 and 6).

First, we investigated the localization of Panct1 in ESCs using the c-KLAN protocol. We simultaneously generated RNA and DNA probes and compared RNA FISH signals between the two different probe sets. Staining of mouse ESCs with both probes uncovered a punctate localization pattern of the Panct1 transcript, predominantly in the nucleus, with the signals vanishing when we treated the cells with RNase A or transfected them with esiRNAs targeting Panct1 (Fig. 2b,c and Supplementary Fig. 7). This suggests that Panct1 exerts its function primarily in the nucleus. Next, we monitored expression of Panct1 upon differentiation of mouse ESCs by embryoid body formation. As is seen for many genes implicated in the maintenance of pluripotency, the amount of Panct1 steadily declined to <40% of initial amounts over an 8-d differentiation (Fig. 2d).

We then performed functional studies of Panct1 in mouse ESCs. Knockdown of Panct1 with two independent esiRNAs (Fig. 2e)
Figure 2 | Implementation of c-KLAN to study the role of Panct1 in mouse ESCs. (a) Screening strategy for lncRNAs modulating Oct4 expression. (b) RNA FISH using Panct1 riboprobes on cells transfected with control (Luc) and Panct1 esiRNAs. FISH signals are in cyan, DAPI staining is in red. Scale bars, 5 µm. AU, arbitrary units.

(c) Average number of fluorescent spots per cell upon transfection with esiRNA against Luc (41 cells) and Panct1 (39 cells). Error bars indicate positive or negative deviation from average values (represented as black dots) from two replicate experiments. (d) Levels of the indicated transcripts (relative to day 0) during embryoid body formation.

(e) Panct1 levels in cells transfected with two independent esiRNAs relative to control-transfected cells (Luc). (f,g) Relative transcript levels of indicated pluripotency markers (f) or the indicated lineage markers (g) upon Panct1 knockdown relative to control-transfected cells (Luc). Error bars in d–g, s.e.m. (n = 3). *P < 0.01–0.05, **P < 0.05–0.005, ***P < 0.005 (two-tailed Student’s t-test).

(h) The micrographs show alkaline phosphatase staining of ESCs after treatment with Panct1 esiRNAs. Scale bars, 0.4 mm.

resulted in loss of expression of pluripotency markers. Both Oct4-driven GFP signals (Supplementary Fig. 8) and amounts of endogenous Oct4 and Nanog mRNA were reduced (Fig. 2f). In contrast, expression of lineage markers such as Gata6 (endoderm) and Fgf5 (ectoderm) increased (Fig. 2g). A global expression analysis of Panct1-depleted cells showed that many genes in the Oct4 circuit including Nrob1, Zscan4f, Sox1, Pax6, Foxd3, Ash2l and Fgf4 were downregulated, whereas lineage marker genes such as Fgf5 and Brachyury (T) were upregulated (Supplementary Fig. 9a). Gene ontology analyses of differentially regulated genes showed strong enrichments for terms associated with cell proliferation, cell differentiation and multicellular organism development (Supplementary Fig. 9b). Knockdown of Panct1 resulted in strongly reduced expression of alkaline phosphatase, also a marker for pluripotency (Fig. 2h).

Finally, we measured up to 27% reduction of cells in DNA synthesis (S) phase upon Panct1 knockdown, consistent with an exit of the cells from the pluripotency program (Supplementary Fig. 10).

Based on these observations, we conclude that expression of Panct1 is required to maintain ESC identity. While our manuscript was under review, an shRNA-based screen was published revealing additional pluripotency-associated lncRNAs10.

Rapid advances in lncRNA biology require robust investigation of these transcripts, but they are often difficult to probe owing to the complexities of the transcriptome and the limitations of proper annotation. We introduced a combinatorial approach for localization and functional analysis of lncRNAs at large scale. c-KLAN provides a rapid, easy and reliable way to investigate lncRNA-mediated control of a variety of cellular processes.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturemethods/.


Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We thank S. Rose, A. Weise and R. Zschau for help in esiRNA library synthesis, I. De Vries and I. Nusslein for technical assistance, D. White and J. Schindelin for help with Delaunay imaging and postprocessing, A. Smith (University of Cambridge) for the Oct4-Gfp cell line and all members of the Buchholz laboratory for helpful discussions and suggestions. This work was supported by the...
ONLINE METHODS

Mouse esiRNA library for long noncoding RNAs. All esiRNAs for IncRNAs are available through Eupheria Biotech. The library for mouse esiRNAs was generated from a template consisting of cDNA obtained from NIH3T3 cells and R1/E ESCs (originally established from a 3.5 d blastocyst produced by crossing two 129 substrains (129S1/SvImJ and 129X1/SvJ)), reverse-transcribed using the SuperScript III first-strand synthesis system for reverse-transcription (RT)-PCR (Invitrogen) using both oligo(dT) and random hexamers according to the manufacturer's instructions. cDNA synthesized with oligo(dT) and random hexamers each from NIH3T3 cells and ESCs (0.25 parts each of the final cDNA mix) was used in the subsequent PCRs. DNase I (Qiagen) treatment was done before cDNA synthesis to eliminate genomic DNA. esiRNAs of 300–600 base pairs (bp) were designed to allow efficient knockdown while minimizing off-target silencing. The DEQOR algorithm was used to calculate the optimal region for esiRNAs synthesis, and Primer3 (ref. 19) was used to design primer sequences tagged with a partial SP6 sequence (5′-TGACACTATAGAAGTG-3′) at the 5′ end of the forward primer and a partial T7 sequence (5′-CTCAGTATAGGGAGA-3′) at the 3′ end of the reverse primer. For non–DEQOR-optimized esiRNAs, regions outside the DEQOR-proposed regions were used to design primer sequences. PCR was done in two steps to prepare the amplicons for esiRNA synthesis. The products of the first PCR with the above primers were amplified in a second round with a forward primer tagged with a combined full T7 and partial SP6 sequence (5′-GCTAAATAGGCAGATGACATGAAAGTG-3′) and a reverse primer tagged with a full T7 sequence (5′-GCTATACAGCTATAGGGAGA-3′). The products of the second PCR round were sequenced and analyzed on a LabChip GX Electrophoresis system (Caliper). esiRNA synthesis was performed as described previously (Eupheria Biotech). The esiRNAs were arrayed in 96-well format at a concentration of 20 ng µl−1 for the pluripotency screen. Primer sequences for all the esiRNAs synthesized are available in Supplementary Tables 1 and 7.

RNA FISH on cultured embryonic stem cells. The products from the first PCR round used in the esiRNA synthesis were then amplified using a forward primer with full SP6 sequence (5′-GAATTAGGTGACACTATAGGGAGA-3′) and a reverse primer with full T7 sequence (5′-GCTATACAGCTATAGGGAGA-3′). The amplicons were in vitro–transcribed using Chromatide Alexa Fluor 546–tagged UTPs (Invitrogen) according to the manufacturer's instructions with either T7 polymerase for antisense riboprobe or SP6 polymerase for sense riboprobe. The RNA probes were purified using RNAeasy mini kit (Qiagen) and diluted in hybridization buffer containing 1 part 20× SSC, 2 parts 50% dextran sulfate (Sigma), 2 parts 10 mg ml−1 BSA (Fluka) and 5 parts formamide (Merck). For generating DNA probes to Punct1, asymmetric PCR was performed on the products of the first round of PCR used in the esiRNA synthesis using the Atto 550 PCR Labeling Kit (MoBiTec) to produce Atto550-dUTP–labeled antisense DNA probes.

R1/E ESCs were grown on gelatin-coated chambered slides (Labtek) until they reached 70% confluency. They were then washed with PBS and incubated for 30 s in cytoskeletal (CSK) buffer containing 100 mM NaCl, 300 mM sucrose, 3 mM MgCl2 and 10 mM PIPES pH 6.8 at room temperature (20–25° C). They were then incubated for 10 min in CSK buffer supplemented with 0.5% Triton X-100 and 10 mM vanadium ribonucleoside. Cells were then incubated in CSK buffer for 30 s and fixed in 4% paraformaldehyde for 10 min at room temperature. Slides were stored in 70% ethanol before hybridization.

For hybridization, a mix containing 50–100 ng of probe and 1 µl of 10 ng ml−1 salmon sperm DNA was resuspended in 5 µl of 100% formamide at 37 °C for 10 min followed by denaturation at 74 °C for 7 min and then resuspended in 5 µl hybridization buffer.

Slides were dehydrated through 80%, 95% and 100% ethanol for 3 min each and air-dried at 42 °C on a heating plate; 20 µl of probe was then added to the cells and placed in a humidified chamber overnight.

On the next day slides were washed with 4× saline–sodium citrate (SSC) (600 mM NaCl and 75 mM Na2C6H3O7) followed by washes (three times each) with 2× SSC, 50% formamide at 39 °C for 5 min and 2× SSC at 39 °C for 5 min. They were then washed with 1× SSC at room temperature for 10 min and stained with Prolong Gold antifade reagent containing DAPI (Invitrogen).

Image acquisition and postprocessing. RNA FISH images were acquired using a Deltavision Core widefield deconvolution fluorescence microscope (Applied Precision) using an Olympus UPlanSApo 100×, 1.4 oil-immersion lens. Fluorescence images were taken with a Photometrics Cool Snap HQ2 charge-coupled device (CCD) camera using emission filters for DAPI 435/48 (batch 34–100613–000) and TRITC 594/45 (batch 34–100615–000). Images were measured at a resolution of 15,528 pixels µm−1.

Z-stacks for both channels (DAPI and TRITC) comprised of single scans taken at 0.2 µm were acquired and deconvolved using Resolve3D SoftWoRx-Acquire version 4.1.2 release 1 (Applied Precision) with fast acquire settings. The images were then processed using Fiji as follows. The display range of the Deltavision images was adjusted to the point where the background was invisible followed by a maximum projection of z-stacks in focus. For the representative figures, a fire look-up table was used for the Alexa Fluor 546 channel with a calibration bar representing arbitrary signal intensity. For analysis of fluorescent spots, 8-bit images were threshold adjusted to make background invisible and a particle count was performed with settings optimized for detection of circular particles using the analyze particles tool in Fiji.

Cell culture and high-throughput screen. ESCs (E14TG2a, R1/E, Oct4-Gip21) were cultured in Dulbecco’s Modified Eagle Medium (Invitrogen) supplemented with 4.5 g l−1 D-glucose and pyruvate, 10% fetal bovine serum (Pan Sera), 30 µM 2-mercaptoethanol, 0.6x NEAA (Invitrogen), 300 units of penicillin-streptomycin (Invitrogen) and 8 ng LIF (prepared in-house). Medium was changed every day and ESCs were detached by treatment with trypsin-EDTA (Invitrogen) and split every 2 d. For comparing esiRNA mediated IncRNA knockdowns with other RNAi agents, ESCs (30,000 R1/E cells in 1 ml ESC medium) were reverse-transfected in 12-well plates with 800 ng esiRNA, 1 µl Lipofectamine 2000 (Invitrogen) and 100 µl Opti-MEM (Invitrogen). Renilla luciferase esiRNA transfections served as negative control. RT–quantitative PCR experiments were performed after 48 h with technical replicates (2 replicates ). siRNA (Invitrogen) and non–DEQOR-optimized esiRNA-mediated knockdowns were
performed in the same manner. For the esiRNA screen, Oct4-Gip cells were reverse-transfected in 96-well plates with 100 ng esiRNA, 0.4 μl Lipofectamine 2000 and 40.5 μl Opti-MEM. All edge wells were kept empty and each plate contained 50 esiRNAs for IncRNAs, 6 esiRNAs for Renilla luciferase (negative control), 3 esiRNAs for Sox2 (positive control) and 3 esiRNAs for GFP (assay control). The controls were spread on four corners and the center of the plate to eliminate position bias as far as possible. Cells were seeded at a density of 3,600 per well in 100 μl ESC medium and incubated for 96 h after which GFP fluorescence and cell numbers were measured using a FACS Calibur (BD Biosciences) equipped with a high-throughput sampler loader. The primary screen was performed in three replicates and the secondary validation analysis was performed in six replicates.

**Hit evaluation.** For the primary screen, Z scores were calculated for percentage of GFP-negative cells and mean GFP intensity values for each replicate as follows: $Z = \frac{x - \mu}{\sigma}$, where $x$ is raw score of the replicate, $\mu$ = mean score for all esiRNAs in the plate, $\sigma$ = s.d. for all esiRNAs in the plate. Candidates in the primary screen were classified as hits if they had an average Z score of ±2, had at least two replicates scoring above or below this score for either of the parameters assayed and did not show a strong viability defect monitored from cell numbers relative to other transfections in the same row in the plate. For the validation with secondary esiRNAs, Z scores were calculated as follows: $Z = \frac{x - \mu}{\sigma}$, where $\mu$ and $\sigma$ are the mean and s.d. for the Renilla luciferase knockdowns (negative control), respectively. For the secondary validation, the Z-score threshold was raised to ±3.5, and IncRNAs were classified as ‘hits’ if their average score across six replicates were above or below the Z-score threshold for both parameters.

**RT–quantitative PCR.** Total RNA from ES cells was extracted using RNEasy Mini kit (Qiagen) with DNase I treatment. 1.5 μg of total RNA was reverse-transcribed using Superscript III reverse transcriptase (Invitrogen) and a 1:1 mix of oligo(dT)12–18 primer (Invitrogen) and random hexamers (Applied Biosystems) according to manufacturer’s instructions. cDNA was diluted 1:5 in water and quantitative (q)PCR was performed with the SYBR Green qPCR kit (Abgene) on an MX P3000 qPCR machine (Stratagene). PCR products were run on gel for the IncRNAs obtained as screen hits and those tested for comparison and knockdown efficiency. For all qPCR measurements, beta-2-microglobulin or Gapdh mRNA were used as internal controls and all measurements were normalized to these controls. A no-reverse-transcriptase control, which produced no detectable PCR product, was used to eliminate the possibility of DNA contamination and a water control was used to eliminate the possibility of contamination of qPCR reagents. All qPCR primers used in the study are listed in Supplementary Table 7.

**3’ rapid amplification of cDNA ends.** 3’ rapid amplification of cDNA ends (RACE) was performed using First Choice RLM-RACE kit (Applied Biosciences) according to manufacturer’s instructions. Primers used are listed in Supplementary Table 7.

**Embryoid body formation.** R1/E ES cells were differentiated into embryoid bodies by removal of leukemia inhibitory factor (LIF) for 8 d using the hanging-drop method. Aliquots for expression analysis were taken after 0 d, 2 d, 5 d and 8 d. RNA was isolated, and quantitative RT-PCR was performed as above.

**Alkaline phosphatase staining.** ESCs (E14TG2a) were reverse-transfected in 96-well format as described above. Forty-eight hours after incubation, the cells were split and transferred to 12-well plates. Five days after transfection, the cells were fixed in 4% paraformaldehyde (USB) and then washed with PBS. Cells were treated with alkaline phosphatase Blue Microwell Substrate (Sigma). Images were taken with an Olympus Tissue Culture Microscope fitted with a Canon G11 camera and processed using Fiji image processing software.

**Cell-cycle profiling.** ESCs (30,000 E14TG2a cells in 1 ml ESC medium) were reverse-transfected in 12-well plates with 400 ng esiRNA, 1 μl lipofectamine 2000 and 100 μl Opti-MEM. Five days after transfection, cells were trypsinized, washed with PBS and fixed with ice-cold 70% ethanol. After fixation, cells were washed with PBS, treated with 5 μg RNase A (Qiagen) in PBS and stained with 200 μl of 50 μg ml-1 propidium iodide solution (Invitrogen). Cell-cycle profiles were acquired on a FACS Calibur and analyzed using Cell Quest Pro (BD Biosciences).

**Microarray analysis of Paxnt1 knockdown.** Oct4-Gip ESCs (60,000 cells in 2 ml medium) were reverse-transfected in 6-well plates with 800 ng esiRNAs, 2 μl lipofectamine and 200 μl Opti-MEM. Seventy-two hours after transfection, RNA was extracted as described above. In total, six arrays including three replicates of control (Luc) and three replicates of experiment (Paxnt1 knockdown) were processed and hybridized on the mouse 430 version 2 Array (Affymetrix). Biotinylated cRNA was synthesized with PerkinElmer’s nucleotide analogs using the MEGAScript T7 kit (Ambion). After fragmenting of the cRNA for target preparation using the standard Affymetrix protocol, 15 μg fragmented cRNA was hybridized for 16 h at 45 °C to Mouse Genome 430 v2 array. Arrays were washed and stained with streptavidin-phycocerythrin in the Affymetrix Fluidics Station 400 and scanned using the Affymetrix GeneChip Scanner 3000 7G. The image data were analyzed with GeneChip operating software 1.4 using Affymetrix default analysis settings. Significance of changes in expression was assayed with Student’s t-test. False discovery rate estimates (q values) were calculated with an R package “qvalue,” using a method described previously.

Microarray probes showing an absolute expression change above twofold and having q values below 0.05 were considered hits. Gene ontology enrichment analysis was performed with the GeneCodis online tool.


doi:10.1038/nmeth.1894

© 2012 Nature America, Inc. All rights reserved.