LncATLAS database for subcellular localization of long noncoding RNAs

DAVID MAS-PONTE,[1,2,6] JOANA CARLEVARO-FITA,[3,4,6] EMILIO PALUMBO,[1] TONI HERMOSO PULIDO,[1] RODERIC GUIGO,[1,2,5] and RORY JOHNSON[3,4]

[1]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona 08003, Spain
[2]Universitat Pompeu Fabra (UPF), Barcelona 08003, Spain
[3]Department of Clinical Research, University of Bern, 3008 Bern, Switzerland
[4]Department of Medical Oncology, Inselspital, University Hospital and University of Bern, 3010 Bern, Switzerland
[5]Institut Hospital del Mar d'Investigacions Mèdiques (IMIM), 08003 Barcelona, Catalonia, Spain

ABSTRACT

The subcellular localization of long noncoding RNAs (lncRNAs) holds valuable clues to their molecular function. However, measuring localization of newly discovered lncRNAs involves time-consuming and costly experimental methods. We have created "lncATLAS," a comprehensive resource of lncRNA localization in human cells based on RNA-sequencing data sets. Altogether, 6768 GENCODE-annotated lncRNAs are represented across various compartments of 15 cell lines. We introduce relative concentration index (RCI) as a useful measure of localization derived from ensemble RNA-seq measurements. LncATLAS is accessible through an intuitive and informative webserver, from which lncRNAs of interest are accessed using identifiers or names. Localization is presented across cell types and organelles, and may be compared to the distribution of all other genes. Publication-quality figures and raw data tables are automatically generated with each query, and the entire data set is also available to download. LncATLAS makes lncRNA subcellular localization data available to the widest possible number of researchers. It is available at lncatlas.crg.eu.

Keywords: long noncoding RNA; lncRNA; subcellular localization; nucleus; cytoplasm; chromatin

INTRODUCTION

The functions of long noncoding RNAs (lncRNAs) are intimately linked to location in the cell. The first discovered lncRNAs tended to be located in the nucleus and chromatin, and epigenetically regulate gene expression (Hutchinson et al. 2007; Rinn et al. 2007; Zhao et al. 2008; Whitehead et al. 2009; Mondal et al. 2010; Tsai et al. 2010). However, we now appreciate lncRNAs' localization and molecular functions to be highly diverse. There exists a substantial population of lncRNAs in the cytoplasm (Ulitsky and Bartel 2013; van Heesch et al. 2014; Carlevaro-Fita et al. 2016), with evidence for roles such as translation regulation (Yoon et al. 2012; Schein et al. 2016; Zucchelli et al. 2016), miRNA decoys (Cesana et al. 2011), or protein trafficking (Willingham et al. 2005; Aoki et al. 2010; Kino et al. 2010). Consequently, ascertaining nuclear–cytoplasmic localization has become one of the primary sources of evidence when investigating the molecular role of newly discovered lncRNAs (Hutchinson et al. 2007; Ishizuka et al. 2014; Ounzain et al. 2015; Chen 2016; Chen et al. 2016; Hansji et al. 2016).

The various methods to map RNA molecules in the cell operate with trade-offs in throughput, convenience, and accuracy. Among the single-gene approaches, probably the most commonly used is qRTPCR on RNA extracts of purified cellular compartments (Wang et al. 2006). It yields information on relative RNA concentrations between compartments, but not of absolute molecule numbers per cell. Another method is fluorescence in situ hybridization (FISH), which can in principle yield absolute counts of molecules at subcellular resolution (Raj et al. 2008; Dunagin et al. 2015). However, FISH is time-consuming and low-throughput, and requires expensive reagents (Cabili et al. 2015). More recently, the ingenious in situ sequencing method, FISSEQ, has established high-throughput subcellular RNA counting (Lee et al. 2015). But at present just one data set is available, and it is restricted to several hundred highly expressed lncRNAs (Lee et al. 2014).

The only method currently capable of whole-genome localization mapping is subcellular RNA sequencing (subcRNAseq). In this process, cells are fractionated, extracted, and RNA sequenced (Djebali et al. 2012). SubcRNAseq yields high-throughput and quantitative data, although as with RT-PCR approaches mentioned above, the absolute counts of RNA molecules per cell are lost (Ulitsky and Bartel 2013). Recently, large amounts of raw subcRNAseq data have become available, most notably from the ENCODE Consortium (Djebali et al. 2012; Dunham et al. 2012). These data remain underutilized and have not been made readily accessible.

In light of the growing use of RNA localization to infer the function of newly discovered lncRNAs, and the availability of large amounts of unprocessed subcRNAseq data, we have created a resource to make lncRNA localization data available to the broader scientific community. This resource, "lncATLAS," enables nonexpert users to rapidly access a rich variety of easily interpreted data on their lncRNA of interest.

RESULTS

A database of lncRNA localization based on human RNA-seq data

In light of growing interest in lncRNA and their functions, we decided to create a resource for accessing and visualizing lncRNA localization within human cells. We collected data from the largest data set of subcRNAseq, produced by the ENCODE Consortium (Djebali et al. 2012; Dunham et al. 2012). Raw RNA-seq data from a panel of human cell lines were used to quantify the reference GENCODE gene annotation (Derrien et al. 2012; Harrow et al. 2012). RNA-seq experiments were obtained for a total of 15 cell lines comprising 48 individual experiments (see Supplemental Table S1). These cells originate from a wide diversity of adult and embryological organ sites and comprise both transformed and normal cells (Fig. 1A). For each cell, cytoplasmic and nuclear data are available, and for the majority of these, whole-cell data were also obtained. In addition, from a single cell line, K562, subnuclear, and subcytoplasmic compartment data are also available (Fig. 1B; Supplemental Table S1). Hereafter we refer to these as "compartments." Poly(A)$^+$ RNA samples were available for whole cell, cytoplasm, nucleus, and subcytoplasmic compartments, and total RNA for subnuclear and nuclear compartments.

Defining localization from RNA-seq data

Throughout the present study, for practical reasons, we adopt a relative scheme to define and quantify RNA localization: the "relative concentration index" (RCI). RCI is defined as the log$_2$-transformed ratio of FPKM (fragments per kilobase per million mapped) in two samples, for example, the cytoplasm and nucleus (Fig. 1C). A similar approach has been used previously (Derrien et al. 2012; Ulitsky and Bartel 2013). It is worth commenting on exactly how these values should be interpreted: RCI is the ratio of a transcript's concentration, per unit mass of sampled RNA, between two compartments. Sampled RNA populations may be poly(A)$^+$ RNA or total RNA, and we are careful to only compute RCI values within the same population.

The mass of RNA per compartment per cell is not equal, and typically not quantified prior to RNA-seq (Djebali et al. 2012). Therefore, without knowing the total mass of poly(A)$^+$ RNA in the nucleus and cytoplasm of a single cell, we cannot make statements about the relative "number" of RNA transcripts in cellular compartments of a single cell (Cabili et al. 2015).

We here briefly digress to contrast this approach with another possible way to define relative subcellular localization. Perhaps more obvious is a "molecular" definition, in terms of numbers of molecules of a given RNA transcript X in the compartments of a single cell. For example, if one cell has 10 and five molecules of X in the cytoplasm and nucleus, respectively, then its cytoplasmic/nuclear localization would be defined as 10/5 = 2. We define this measure as relative molecules index (RMI). Such information is, in principle, directly accessible from fluorescence-based techniques and has been calculated previously (Lee et al. 2014; Cabili et al. 2015). As mentioned above, our ignorance of the total poly(A)$^+$ RNA mass of the cell lines used here precludes the calculation of RMI in this study.

Computing localization across genes and cell types

RCI was calculated for various selected pairs of cellular compartments (Fig. 1D). In the majority of cases, we calculated the cytoplasmic/nuclear RCI—"CN-RCI" (Supplemental Table S2) (see Materials and Methods). This is a measure of the relative concentration of an RNA sequence in the cytoplasm, compared to the nucleus, in log$_2$ units. For one cell type, K562, total RNA data from subnuclear and poly(A)$^+$ RNA from subcytoplasmic compartments were also calculated, by reference to total RNA from the nucleus or poly(A)$^+$ RNA from the cytoplasm, as appropriate (Supplemental Table S3) (see Materials and Methods). Altogether this yielded localization estimates in 20 compartment/cell combinations (Fig. 1D).

Where available, replicate data were used to assess reliability of RCI measurements (see Supplemental Table S1 for information about availability of replicates). Silent and unreliable genes were excluded from further consideration (see Materials and Methods for details in the filtering steps). Between 3114 (max) and 582 (min), lncRNAs' CN-RCI localization could be estimated per cell, after filtering (Figs. 1D, 2A). Note that H1.hESC has a greater number of detected genes (4923 genes), because biological replicates were not available for cytoplasm or nucleus for this cell line. A total of 24,538 genes (17,770 mRNAs and 6768 lncRNAs) were
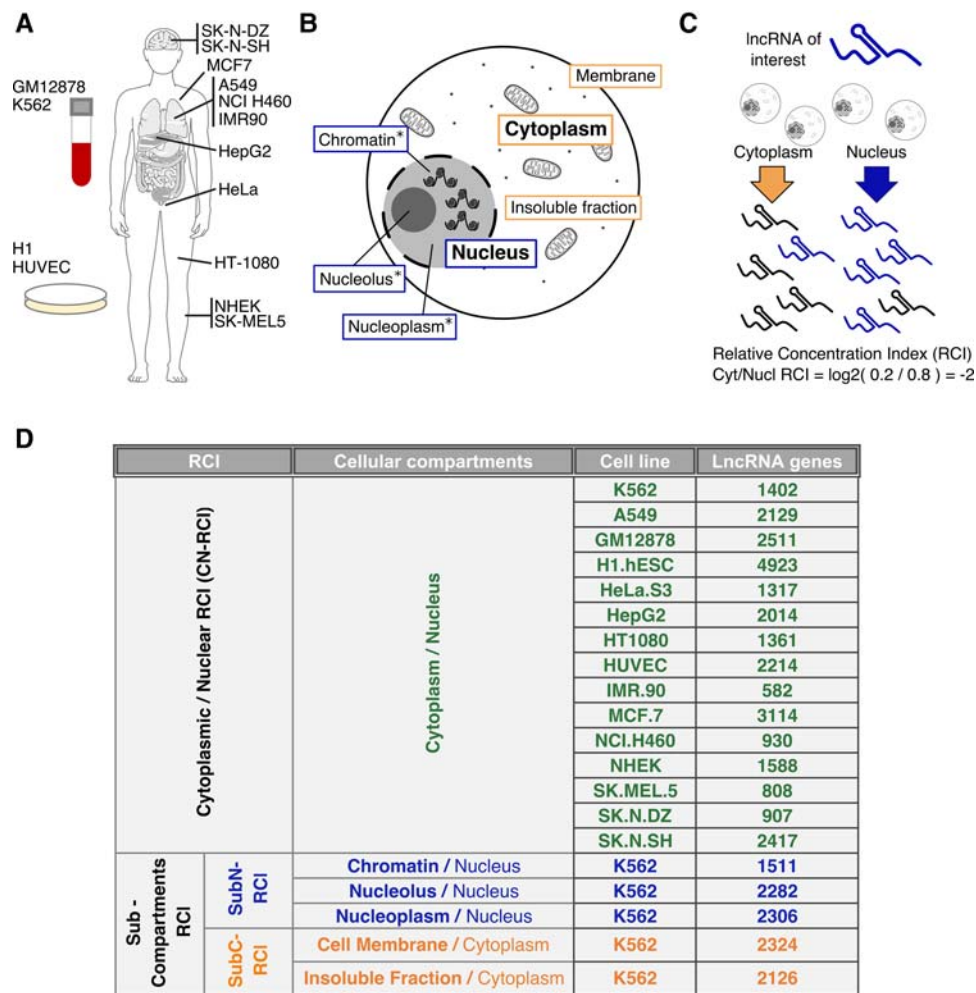
FIGURE 1. Overview of lncATLAS data. (*A*) Cell lines available in lncATLAS, indicating their approximate origin. (*B*) Cellular compartments available. (*) Compartments with only total RNA samples available. (*C*) The relative concentration index (RCI), in this case calculated for the cytoplasm and nucleus (CN-RCI). The RCI can be thought of as the log-ratio, between two compartments, of the concentration of a given RNA molecule per unit mass of RNA. (*D*) Overview of the cell lines and cellular compartments available for lncRNA RCI calculations. "LncRNA genes" column indicates the number of lncRNA for which the RCI could be calculated in the corresponding cell line (see Materials and Methods for details). Sub-C RCI and Sub-N RCI correspond to subcytoplasmic RCI and subnuclear RCI, respectively.

quantified in at least one cell type. Of these, 31 lncRNAs were detected in all samples tested (Fig. 2B; Supplemental Table S4). LncRNAs display a highly cell-type-specific detection pattern, in contrast to mRNAs, as observed previously (Fig. 2B; Derrien et al. 2012; Guttman and Rinn 2012; Cabili et al. 2015).

RCI data are consistent with known cytoplasmic–nuclear localization tendencies of lncRNAs and mRNAs. Among the top 15 most cytoplasmic measurements, 14 represent mRNAs (the remainder is a lncRNA overlapping a protein-coding gene's 3′UTR on the same strand, meaning that it is impossible to distinguish the RNA-seq of the two overlapping genes, and the cytoplasmic RCI reading may in fact arise from the protein-coding gene) (Supplemental Table S5). In contrast, 12 of the 15 most nuclear RCI values represent lncRNAs (Supplemental Table S6). The nuclear-enriched X-chromosome inactivating transcript *XIST* occupies the

top four positions (Brown et al. 1992; Clemson et al. 1996). Manual inspection of several well-known lncRNAs showed that localization reported here tended to be consistent with literature reports (see next section).

Detected lncRNA genes reported by lncATLAS cover a substantial fraction of the entries from manually curated and widely used databases, such as lncRNAdb (Quek et al. 2015) and LncRNADisease (Chen et al. 2013), and localization database RNALocate (Zhang et al. 2016). Note that these databases contain a mixture of GENCODE annotated and non-GENCODE annotated entries. 74, 128, and 150 genes from lncRNAdb, LncRNADisease, and RNALocate, respectively, are present in lncATLAS. These numbers represent 39%, 48%, and 47% of the total number of human lncRNAs and 82%, 75%, and 90% of the total number of human *GENCODE annotated* lncRNAs of each database, respectively (Fig. 2C). LncRNAs from these that are absent in

A

B

C
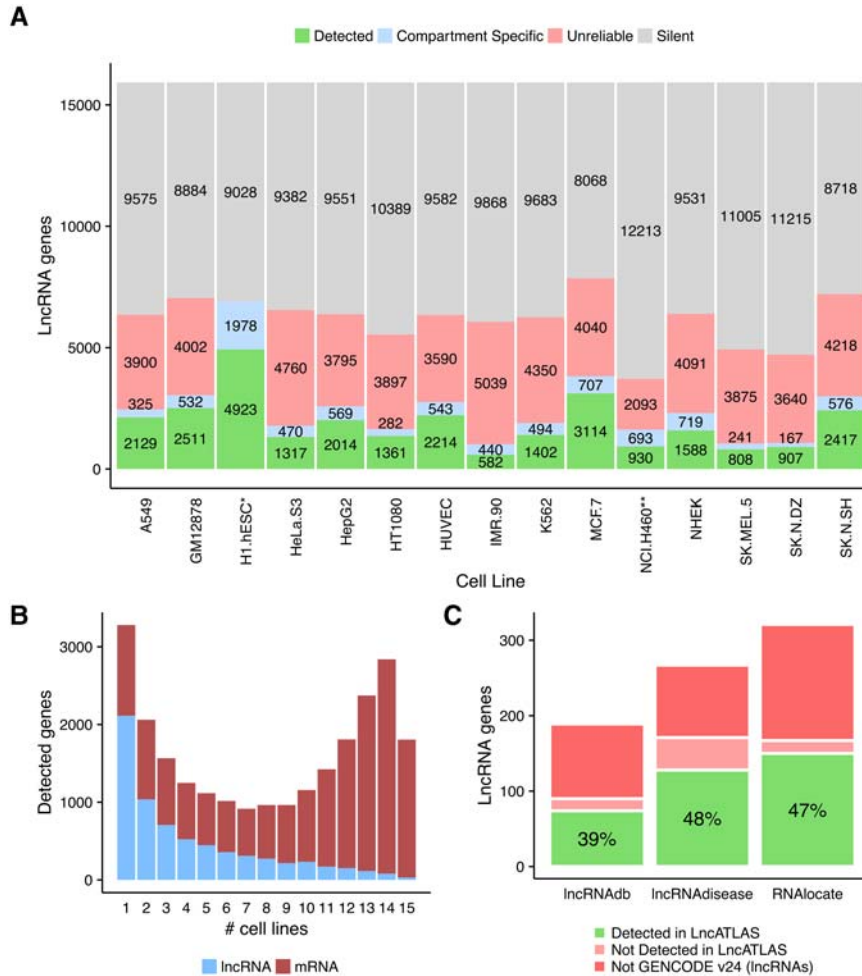
FIGURE 2. Summary of cytoplasmic/nuclear detected genes by lncATLAS. (*A*) Number of genes analyzed in lncATLAS. The *y*-axis maximum represents the total number of annotated lncRNA genes in GENCODE. (i) Detected genes: present a reliable expression value in compartments, cytoplasm, and nucleus. Other categories comprise (ii) compartment-specific genes: present a reliable non-zero value in one compartment and zero in the other; (iii) unreliable genes: genes that did not pass the twofold cutoff (see Materials and Methods, "Data Processing" section); and (iv) silent genes: not expressed in any compartment (see full category definitions in Materials and Methods, "Cytoplasmic–Nuclear Relative Concentration Index" section). (*) No biological replicates were available for cytoplasm or nucleus. (**) No biological replicates were available for cytoplasm. (*B*) The number of lncRNA genes defined as "detected" in the indicated numbers of cell types. LncRNA genes in blue and mRNAs in red. (*C*) Coverage by lncATLAS of widely used, manually curated lncRNA databases, LncRNAdb (Quek et al. 2015), and LncRNADisease (Chen et al. 2013). The *y*-axis maximum represents the total number of unique human gene loci in each database. The RNALocate database (Zhang et al. 2016) is also shown for comparison. Note that these databases hold a mixture of GENCODE annotated and non-GENCODE annotated lncRNAs. The bar plot displays the total number of human lncRNA genes in each database (whole bar). Bars are colored to represent, from the total number of human lncRNAs in a database, the number of genes that are present in lncATLAS (green); are part of GENCODE annotation but are not detected in any of the 15 cell lines in lncATLAS (pink); do not have GENCODE identifiers (red).

lncATLAS are either not detected in any of the cell lines, or else do not belong to the GENCODE annotation.

LncATLAS webserver for exploring localization data

The lncATLAS data set was compiled into a relational database that is searchable through a webserver at lncatlas.crg.eu.

LncRNAs of interest are accessed using official gene names or GENCODE gene identifiers. A maximum of three genes may be investigated simultaneously. Several well-known lncRNAs with known localization are also available for reference. Once a gene or genes has been selected, a series of data interpretations are presented and summarized below. As examples, results for *MALAT1* (nuclear localized) (Hutchinson et al. 2007) and *DANCR* (cytoplasm localized) (van Heesch et al. 2014; Cabili et al. 2015) lncRNAs data are shown in Figures 3 and 4.

The following sections summarize the data presented to the users for their gene of interest (GOI).

Inspect the cytoplasmic–nuclear localization of your gene of interest (GOI)

Data are only displayed for selected gene(s).

Plot 1: Cytoplasmic/nuclear localization: RCI and expression values (all cell types): In this basic summary, the cytoplasmic/nuclear RCI is shown as a bar plot across all available cell types. Bars are colored to reflect the expression level of the gene, as inferred from nuclear RNA-seq. The individual FPKM values, upon which RCI values are based, are displayed. When a gene is expressed only in one compartment, RCI cannot be computed; then, dashed bars with expression values are shown instead (Fig. 3A).
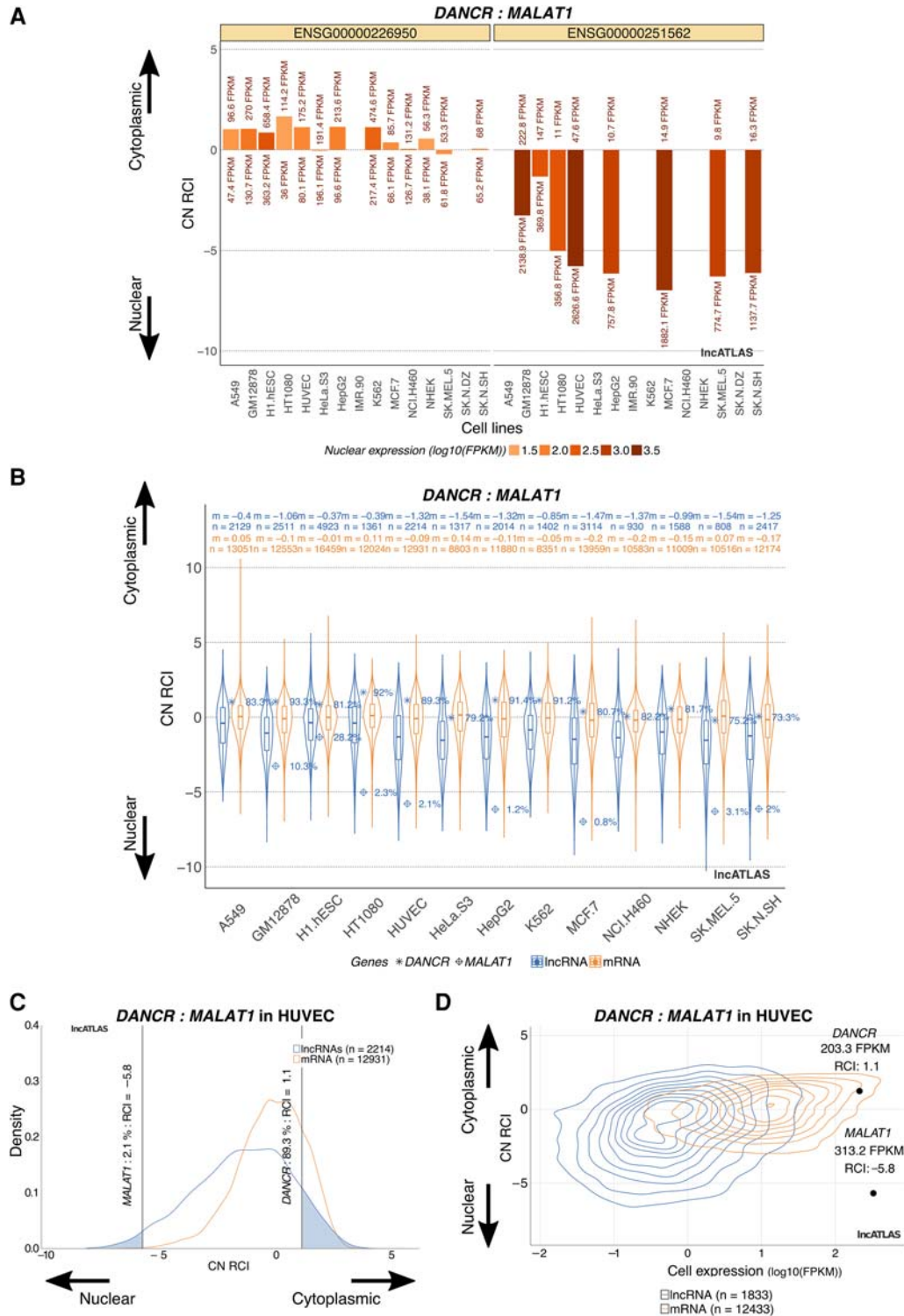
Inspect the cytoplasmic–nuclear localization of your GOI within the distribution of all genes

The aim of the second section is to understand, in terms of localization, how the genes of interest behave relative to all other genes. Three different plots show CN-RCI values distribution for all lncRNAs and mRNAs, within which the location of the GOI is indicated.

Plot 2: Cytoplasmic/nuclear localization: RCI distribution (all cell types): To put RCI values in context, their percentile rank within the distribution of all lncRNAs is indicated (ranks relative to lowest value). Data are shown for all cell types (Fig. 3B).

Plot 3: Cytoplasmic/nuclear localization: RCI distribution (individual cell type): The same data are shown as for Plot 2, but in the form of a density plot. The User must here

FIGURE 3. Subcellular localization plots displayed by lncATLAS. *MALAT1* and *DANCR* genes are selected as examples of nuclear and cytoplasmic lncRNAs, respectively. (*A*) Bars representing CN-RCI values for the selected genes across all cell lines. Expression values (FPKMs) for the genes of interest are shown for both compartments (cytoplasm on *top* of the bar, nucleus on the *bottom*). Bars are colored by their absolute nuclear expression. (*B*) Boxplot showing CN-RCI values distribution of all lncRNAs (blue) and mRNAs (orange) for each cell line ("*n*" indicates total number of genes, "*m*" median of CN-RCI values for lncRNAs and mRNAs separately). LncRNAs of interest are located in the distribution and a percentage indicates their percentile rank within the distribution of all lncRNAs (ranks relative to lowest value). (*C*) Same as in the previous plot but in this case distribution is shown as a density plot and only for a particular cell type, HUVEC. Again, genes of interest are located in the distribution and their percentile rank (relative to lowest value) and RCI are indicated. (*D*) Contour plot showing lncRNA and mRNA populations as a function of CN-RCI values and whole cell expression [$\log_{10}$(FPKMs)]. LncRNAs of interest are specifically displayed together with their whole-cell expression and RCI.
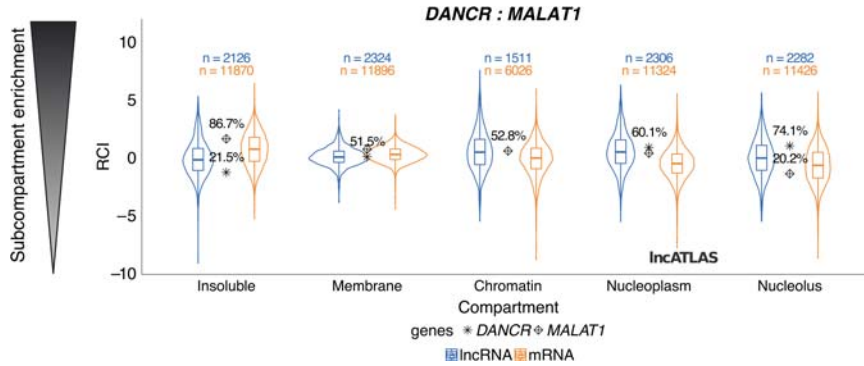
FIGURE 4. Subcompartment data as displayed in lncATLAS. For K562 cells, subnuclear and subcytoplasmic fractions were available and RCI for all lncRNAs and mRNAs was computed (see Materials and Methods). Boxplot shows the distribution of these values for each subnuclear and subcytoplasmic compartment ("*n*" indicates total number of genes in a distribution). Percentile rank (relative to lowest value) of each gene of interest is displayed to contextualize the relative enrichment of these genes in a subcompartment compared to the rest of lncRNAs and mRNAs.

specify a single cell type. When genes are not classed as "Detected," RCI cannot be computed and no data are shown (Fig. 3C).

Plot 4: Cytoplasmic/nuclear localization: comparison with expression (individual cell type): As for Plot 3, gene values are shown in the context of all other genes in the same cell, but here also indicating whole-cell expression values. As before, the data are shown for a single-cell type chosen by the User, and plots are only generated for cells where RCI values are "Detected" (Fig. 3D).

Inspect the localization of your GOI at subcompartment level

The final section gives information about enrichment in the cytoplasmic and nuclear subcompartments of K562 cells. As in the previous section, RCI values for the genes of interest are indicated in the context of full lncRNA and mRNA distributions.

Plot 5: Subcytoplasmic and subnuclear localization in K562 cells: Here data are shown for subnuclear and subcytoplasmic compartments K562 cell line. As in Plots 2 and 3, distributions across all detected genes are shown (Fig. 4).

In the examples shown, the differences in localization of *MALAT1* and *DANCR* are clear. Their cytoplasmic–nuclear localizations are highly divergent (Figs. 3, 4) and broadly consistent across all the cell lines observed. The difference in localization is observed even in cells where their overall expression level is similar (e.g., HUVEC, Fig. 3D).

All figures may be downloaded as publication-quality files in pdf format. Similarly, in the *Get Raw Data* tab, the underlying RCI and raw expression values for selected genes may be accessed as a batch query. Furthermore, the entire set of data tables for lncATLAS may be downloaded in the same tab using the *Download All raw data* button.

DISCUSSION

Subcellular localization provides important clues to the molecular function of novel lncRNAs. LncATLAS is designed to make such data available to the largest number of researchers. To our knowledge, only one other database of lncRNA localization exists: RNALocate (Zhang et al. 2016). RNALocate contains manually curated localization classifications across multiple species. Despite focusing on a single species (human), due to the limited availability of subcellular RNA-seq data, lncATLAS has two key advantages: it is quantitative, and it is based on standard GENCODE annotations, the de facto official annotation for both protein-coding and lncRNA genes (Derrien et al. 2012). These features boost the usefulness of lncATLAS data for other research groups and ensure its integration with diverse other genomics data sets. Future subcellular RNA-seq data from other cell types, or other species, will be integrated as they become available.

MATERIALS AND METHODS

Data source

Cytoplasmic and nuclear poly(A)$^+$ RNA-seq data from 15 different cell lines were obtained from ENCODE (Djebali et al. 2012). (ENCODE RNA-seq data in BAM format were obtained from the ENCODE Data Coordination Centre [DCC] in September 2016; https://www.encodeproject.org/matrix/?type=Experiment.) For most cell lines, whole-cell data were also obtained (exceptions being HT1080, NCI.H460, SK.MEL.5, and SK.N.DZ). A full list of processed RNA-seq libraries is available in Supplemental Table S1.

Data processing

Data were mapped to human genome assembly GRCh38 using STAR software (Dobin et al. 2013) and quantified with RSEM (Li and Dewey 2011) for all GENCODE v24 genes, within the GRAPE analysis pipeline (Harrow et al. 2012; Knowles et al. 2013). Data consisted of two independent biological replicates per cell line and fraction (exceptions being H1.hESC cytoplasm and nucleus and NCI. H460 cytoplasm for which only one replicate was available) (see Supplemental Table S1 for a full list of source data sets). For subcytoplasmic RCI, instead of using poly (A)$^+$ cytoplasmic samples from the Gingeras laboratory (used for CNRCI), we used the corresponding sample from the laboratory where the subcytoplasmic fractionation was done (Lécuyer laboratory). This is not considered as an additional biological replicate.

Throughout, RNA-seq data are processed at the level of genes, rather than transcripts. From the whole GENCODE v24 annotation, genes contained in the "Long noncoding RNA gene annotation"

data set define our lncRNA gene set. The protein-coding gene set is defined by gene biotype "protein_coding" (see Supplemental Table S7).

In order to remove genes with high variability between replicates, genes with a greater than twofold difference between replicates are labeled "unreliable" and excluded from further analysis. This cutoff was not possible for the samples mentioned above, for which replicate experiments were not available.

Cytoplasmic–nuclear relative concentration index (CN-RCI)

For cytoplasmic/nuclear localization, poly(A)$^+$ RNA data were used. At this stage, all genes are defined in one of four categories in each cell line: (i) Detected: genes with non-zero and "reliable" values in both cellular compartments; (ii) Compartment-specific: genes considered "reliable" in both compartments but expressed >0 FPKM only in one; (iii) Silent: both compartments have FPKM = 0; (iv) Unreliable: genes that are unreliable in at least one compartment.

Genes classed as "Detected" were retained, and their localization was computed as the C/N relative concentration index (CN-RCI) thus:

$$\text{RCI}_{C/N}^{pA+} = \log2\left(\frac{\text{Cytoplasmic expression (FPKM)}}{\text{Nuclear expression (FPKM)}}\right).$$

Only detected genes are shown in plots of lncATLAS, with the exception of compartment-specific genes in Plot 1. For this group of genes, CN-RCI value is not available in the plot and the bar only indicates the tendency of the gene toward nucleus or cytoplasm. Color and FPKM are shown normally, to indicate the level of expression.

Subcytoplasmic and subnuclear relative concentration index for K562 (sub-N RCI, sub-C RCI)

RNA-seq data from subnuclear fractions (chromatin, nucleolus, and nucleoplasm) and from subcytoplasmic fractions (membrane and insoluble fraction) were retrieved and processed as above. These data are only available for K562 cells, and in the case of subnuclear samples correspond to total RNA [not poly(A)$^+$-selected RNA]. Data were processed and RCI calculated as above, with the only differences being (i) the RCI was calculated with reference to the nuclear fraction for subnuclear compartments, and cytoplasmic fraction for subcytoplasmic compartments; (ii) for subnuclear compartments total RNA samples were used, instead of poly(A)$^+$.

Database design

LncATLAS is a relational database implemented in MySQL (http://www.mysql.com) and designed through the official MySQL WorkBench tool for Linux. The entity-relationship (ER) diagram extracted summarizes its structure (Supplemental Fig. S1). The tables are hierarchically organized from general information of the samples to the expression value per gene that is stored in the expression table.

Web-tool implementation

LncATLAS is constructed using the Shiny R package (version 0.13.2) (http://www.rstudio.com/shiny/). The database is connected to the application itself via the R package RMySQL (version 0.10.9). Other packages used in the implementation are the *ggplot2* package (version 2.2.1) and *dplyr* (version 0.5.0) used to build custom plots and manipulate the data.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

REFERENCES

Aoki K, Harashima A, Sano M, Yokoi T, Nakamura S, Kibata M, Hirose T. 2010. A thymus-specific noncoding RNA, Thy-ncR1, is a cytoplasmic riboregulator of MFAP4 mRNA in immature T-cell lines. *BMC Mol Biol* **11:** 99.

Brown CJ, Hendrich BD, Rupert JL, Lafrenière RG, Xing Y, Lawrence J, Willard HF. 1992. The human XIST gene: analysis of a 17 Kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* **71:** 527–542.

Cabili MN, Dunagin MC, McClanahan PD, Biaesch A, Padovan-Merhar O, Regev A, Rinn JL, Raj A. 2015. Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol* **16:** 20.

Carlevaro-Fita J, Rahim A, Guigó R, Vardy LA, Johnson R. 2016. Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. *RNA* **22:** 867–882.

Cesana M, Cacchiarelli D, Legnini I, Santini T, Sthandier O, Chinappi M, Tramontano A, Bozzoni I. 2011. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* **147:** 358–369.

Chen LL. 2016. Linking long noncoding RNA localization and function. *Trends Biochem Sci* **41:** 761–772.

Chen G, Wang Z, Wang D, Qiu C, Liu M, Chen X, Zhang Q, Yan G, Cui Q. 2013. LncRNADisease: a database for long-non-coding RNA-associated diseases. *Nucleic Acids Res* **41:** D983–D86.

Chen J, Liu Y, Lu S, Yin L, Zong C, Cui S, Qin D, Yang Y, Guan Q, Li X, et al. 2016. The role and possible mechanism of lncRNA U90926 in modulating 3T3-L1 preadipocyte differentiation. *Int J Obes (Lond)* **41:** 299–308.

Clemson CM, McNeil JA, Willard HF, Lawrence JB. 1996. XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *J Cell Biol* **132:** 259–275.

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, et al. 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* **22:** 1775–1789.

Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, et al. 2012. Landscape of transcription in human cells. *Nature* **489:** 101–108.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29:** 15–21.

Dunagin M, Cabili MN, Rinn J, Raj A. 2015. Visualization of lncRNA by single-molecule fluorescence in situ hybridization. *Methods Mol Biol* **1262:** 3–19.

Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R, et al. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489:** 57–74.

Guttman M, Rinn JL. 2012. Modular regulatory principles of large non-coding RNAs. *Nature* **482:** 339–346.

Hansji H, Leung EY, Baguley BC, Finlay GJ, Cameron-Smith D, Figueiredo VC, Askarian-Amiri ME. 2016. ZFAS1: a long noncoding RNA associated with ribosomes in breast cancer cells. *Biol Direct* **11:** 62.

Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, et al. 2012. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* **22:** 1760–1774.

Hutchinson JN, Ensminger AW, Clemson CM, Lynch CR, Lawrence JB, Chess A. 2007. A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics* **8:** 39.

Ishizuka A, Hasegawa Y, Ishida K, Yanaka K, Nakagawa S. 2014. Formation of nuclear bodies by the lncRNA Gomafu-associating proteins Celf3 and SF1. *Genes Cells* **19:** 704–721.

Kino T, Hurt DE, Ichijo T, Nader N, Chrousos GP. 2010. Noncoding RNA Gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci Signal* **3:** ra8.

Knowles DG, Röder M, Merkel A, Guigó R. 2013. Grape RNA-Seq analysis pipeline environment. *Bioinformatics* **29:** 614–621.

Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Yang JL, Ferrante TC, Terry R, Jeanty SS, Li C, Amamoto R, et al. 2014. Highly multiplexed subcellular RNA sequencing in situ. *Science* **343:** 1360–1363.

Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Ferrante TC, Terry R, Turczyk BM, Yang JL, Lee HS, Aach J, et al. 2015. Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat Protoc* **10:** 442–458.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* **12:** 323.

Mondal T, Rasmussen M, Pandey GK, Isaksson A, Kanduri C. 2010. Characterization of the RNA content of chromatin. *Genome Res* **20:** 899–907.

Ounzain S, Micheletti R, Arnan C, Plaisance I, Cecchi D, Schroen B, Reverter F, Alexanian M, Gonzales C, Ng SY, et al. 2015. CARMEN, a human super enhancer-associated long noncoding RNA controlling cardiac specification, differentiation and homeo-stasis. *J Mol Cell Cardiol* **89:** 98–112.

Quek XC, Thomson DW, Maag JL, Bartonicek N, Signal B, Clark MB, Gloss BS, Dinger ME. 2015. lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Res* **43:** D168–D173.

Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. 2008. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods* **5:** 877–879.

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, et al. 2007. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129:** 1311–1323.

Schein A, Zucchelli S, Kauppinen S, Gustincich S, Carninci P. 2016. Identification of antisense long noncoding RNAs that function as SINEUPs in human cells. *Sci Rep* **6:** 33605.

Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, Shi Y, Segal E, Chang HY. 2010. Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329:** 689–693.

Ulitsky I, Bartel DP. 2013. lincRNAs: genomics, evolution, and mechanisms. *Cell* **154:** 26–46.

van Heesch S, van Iterson M, Jacobi J, Boymans S, Essers PB, de Bruijn E, Hao W, MacInnes AW, Cuppen E, Simonis M. 2014. Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome Biol* **15:** R6.

Wang Y, Zhu W, Levy DE. 2006. Nuclear and cytoplasmic mRNA quantification by SYBR green based real-time RT-PCR. *Methods* **39:** 356–362.

Whitehead J, Pandey GK, Kanduri C. 2009. Regulation of the mammalian epigenome by long noncoding RNAs. *Biochim Biophys Acta* **1790:** 936–947.

Willingham AT, Orth AP, Batalov S, Peters EC, Wen BG, Aza-Blanc P, Hogenesch JB, Schultz PG. 2005. A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. *Science* **309:** 1570–1573.

Yoon JH, Abdelmohsen K, Srikantan S, Yang X, Martindale JL, De S, Huarte M, Zhan M, Becker KG, Gorospe M. 2012. LincRNA-p21 suppresses target mRNA translation. *Mol Cell* **47:** 648–655.

Zhang T, Tan P, Wang L, Jin N, Li Y, Zhang L, Yang H, Hu Z, Zhang L, Hu C, et al. 2016. RNALocate: a resource for RNA subcellular localizations. *Nucleic Acids Res* **35:** D810–D814.

Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. 2008. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322:** 750–756.

Zucchelli S, Patrucco L, Persichetti F, Gustincich S, Cotella D. 2016. Engineering translation in mammalian cell factories to increase protein yield: the unexpected use of long non-coding SINEUP RNAs. *Comput Struct Biotechnol J* **14:** 404–410.