

AnnoLnc2: the one-stop portal to systematically annotate novel lncRNAs for human and mouse

Lan Ke^{1,†}, De-Chang Yang^{1,†}, Yu Wang¹, Yang Ding² and Ge Gao^{1,*}

¹School of Life Sciences, Biomedical Pioneering Innovation Center (BIOPIC) & Beijing Advanced Innovation Center for Genomics (ICG), Center for Bioinformatics (CBI) and State Key Laboratory of Protein and Plant Gene Research, Peking University, Beijing 100871, China and ²Beijing Institute of Radiation Medicine, Beijing 100850, China

Received March 12, 2020; Revised April 21, 2020; Editorial Decision April 28, 2020; Accepted April 29, 2020

ABSTRACT

With the abundant mammalian lncRNAs identified recently, a comprehensive annotation resource for these novel lncRNAs is an urgent need. Since its first release in November 2016, AnnoLnc has been the only online server for comprehensively annotating novel human lncRNAs on-the-fly. Here, with significant updates to multiple annotation modules, backend datasets and the code base, AnnoLnc2 continues the effort to provide the scientific community with a one-stop online portal for systematically annotating novel human and mouse lncRNAs with a comprehensive functional spectrum covering sequences, structure, expression, regulation, genetic association and evolution. In response to numerous requests from multiple users, a standalone package is also provided for large-scale offline analysis. We believe that updated AnnoLnc2 (<http://annolnc.gao-lab.org/>) will help both computational and bench biologists identify lncRNA functions and investigate underlying mechanisms.

INTRODUCTION

Long noncoding RNAs (lncRNAs) have been demonstrated to play many important regulatory roles via various mechanisms (1,2), and many characteristics have been used to infer their functions. For example, analyzing the natural selection pressures on lncRNAs can reveal whether they are functional (3), and their subcellular localization indicates where they perform functions (4). Functional enrichment of protein-coding genes whose expression profiles are highly correlated with lncRNAs provides important hints for potential lncRNA functionality (5). In addition, it has been reported that many lncRNAs perform their functions by interacting with other cellular factors; thus, studying interacting partners of lncRNAs, e.g. proteins and miRNAs, provides insights into their possible molecular mecha-

nisms (6,7). Multiple tools have been developed to annotate lncRNA functions via particular mechanisms (e.g., LncTar (8) for predicting interacting RNA partners, SFPEL-LPI (9) for predicting lncRNA–protein interactions, lncFunTK (10) for annotating GO functions and regulatory networks based on coexpressed genes along with transcription factors and miRNA binding profiles, and lncLocator (11) and iLoc-lncRNA (12) for predicting lncRNA subcellular localization), along with several databases available for known lncRNAs (e.g. LncBook (13) and NONCODE (14)) (see Supplementary Table S1 for a detailed comparison).

As part of our continual attempts to provide the community with intuitive online tools for annotating novel lncRNAs effectively and efficiently, we present AnnoLnc2, the new version of the AnnoLnc web server (15). AnnoLnc2 supports on-the-fly annotation for any human or mouse lncRNA. In addition to supporting new species and new annotation modules, AnnoLnc2 has been fully updated, with many new datasets incorporated (Table 1). To enable computational biologists to run high-throughput analysis efficiently, we also provide a fully functional standalone package for large-scale offline analysis. All these resources are freely available at <http://annolnc.gao-lab.org/> and open to all users without login requirements.

RESULTS

Effectively annotating novel lncRNAs in human and mouse

To systematically annotate novel lncRNAs, AnnoLnc2 is equipped with multiple annotation modules, covering sequence and structure, expression and regulation, function and interaction, as well as genetic association and evolution (Figure 1).

Sequence and structure. For inputted lncRNAs, AnnoLnc2 will first try to map them against the latest human (hg38) and mouse (mm10) genome builds. During mapping, AnnoLnc2 tries to associate the inputted lncRNAs with known lncRNAs by searching GENCODE gene models based on genomic location and intron/exon structure, and

*To whom correspondence should be addressed. Tel: +86 010 62755206; Email: gaog@mail.cbi.pku.edu.cn

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Table 1. Summary of annotation data sources for each module

Module	Annotation data
Expression	Human: 52 RNA-seq samples for 26 tissues (58), 39 CCLE cancer cell lines (59), and 4 RNA-seq samples for H1 and GM12878 from ENCODE (60); Mouse: 56 RNA-seq samples for 28 tissues and 13 RNA-seq samples for 7 cell lines from ENCODE (60).
Transcriptional regulation	Human: 13 515 ChIP-seq samples for 1339 TFs (25); Mouse: 10 728 ChIP-seq samples for 738 TFs (25).
Subcellular localization	Human: 40 RNA-seq samples covering 10 cell lines (29).
miRNA regulation	Human: 170 AGO CLIP-seq samples in various human tissues (e.g. brain cortex) and cell lines (e.g. H1, HK-2 and osteoblast); Mouse: 153 AGO CLIP-seq samples in various mouse tissues (e.g., cortex and liver) and cell lines (e.g. mESC, CD4+ T cell and keratinocytes).
Protein interaction	Human: 385 CLIP-seq samples for 188 RBPs in various human tissues (e.g. brain, adrenal gland and hippocampus) and cell lines (e.g. HEK293T, HeLa and K562); Mouse: 238 CLIP-seq samples for 62 RBPs in various mouse tissues (e.g. brain, testis, and cortex) and cell lines (e.g. B cell, mESC, 220-8 cell and N2A cell).
Genetic association	Human: 96 455 trait-associated SNPs from the GWAS catalog (42), 298 590 SNPs in linkage disequilibrium with GWAS SNPs from dbSNP (61), and 4 365 369 eQTLs from GTEx (41); Mouse: 4013 phenotypic alleles and 997 QTL alleles from MGI (43).
Evolution	Human: phyloP and phastCons scores for the primate, mammal, and vertebrate clades were computed by a local implemented UCSC MultiZ-Phast pipeline based on 100-way multiple alignments from UCSC Genome Browser (62); derived allele frequency (DAF) is based on the 1000 Genomes Project (47); Mouse: phyloP and phastCons scores for the glire, euarchontoglires, placental and vertebrate clades were based on 60-way multiple alignments from UCSC Genome Browser (62).

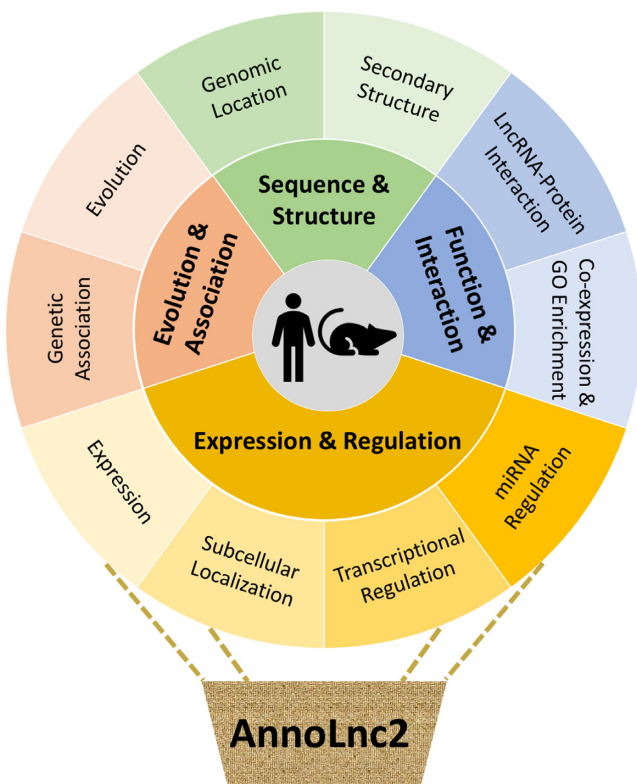


Figure 1. Framework of AnnoLnc2. The AnnoLnc2 web server provides on-the-fly services for annotating novel lncRNAs from 10 perspectives, covering from sequence and structure, expression and regulation, function and interaction, to evolution and association, for human and mouse.

matched hits will be reported as part of the genomic mapping results. In addition, AnnoLnc2 also reports known repeats (16) in the inputted lncRNAs given their potential importance for lncRNAs' functionality (17,18).

lncRNAs tend to acquire complex structures that are strongly linked to their biological functions (19,20). AnnoLnc2 performs *de novo* secondary structure prediction with the ViennaRNA/RNAfold program (21). To help identify potential structural/functional domains, the result is further visualized as an interactive graph, allowing flexible base coloring on the basis of various scores, including conservation and energy entropy.

Expression and regulation. lncRNAs exhibit temporal-spatial expression with sophisticated regulation (22,23). Based on the LocExpress (24) algorithm, AnnoLnc2 implements on-the-fly expression profiling for hundreds of samples. Specifically, the AnnoLnc2 web server supports simultaneous expression calling in 95 human samples (52 normal tissue samples, 39 cancer cell lines from Cancer Cell Line Encyclopedia (CCLE) and 4 ENCODE cell line samples) or 69 mouse samples (56 normal tissue samples and 13 ENCODE cell line samples) for any lncRNA (Supplementary Table S2) in nearly real time. Compared to AnnoLnc, AnnoLnc2 offers a more abundant and reliable representation of lncRNA expression with a novel species (mouse) and a ~50% increased human sample pool. Likewise, AnnoLnc2 also presents an eight-fold increase in its transcriptional regulation module, with 91.7 million ChIP-seq-based binding sites for 1339 human transcription factors (TFs) as well as 80.8 million sites for 738 mouse TFs (25). For miRNA-based regulation, AnnoLnc2 implements a dedicated AGO-CLIP-based module to identify putative functional miRNA-binding events for inputted lncRNAs based on 170 human datasets and 153 mouse datasets and allows users to run TargetScan (26) to predict lncRNA-miRNA interactions. Furthermore, for each predicted miRNA binding site, AnnoLnc2 provides a structural view, with this site highlighted for linking the binding site and structural context.

Several recent studies have well demonstrated that, in addition to expression level, the subcellular localization of lncRNAs is also under intricate regulation (27,28). In an attempt to map the lncRNA subcellular localization landscape at fine-scale, we designed and implemented a novel module to empirically characterize lncRNA cytoplasmic/nuclear localization preference in ten human cell lines (29). Briefly, the inputted lncRNAs are compared with a precompiled localization catalog (Supplementary Figure S1). AnnoLnc2 then returns the localization preference, measured by the fold change in nuclear/cytosolic expression in all ten cell lines, along with the corresponding FDR-adjusted *P*-value (Figure 2C, see Supplementary Figure S1 for the detailed pipeline). Moreover, many studies have shown that multiple sequence motifs contribute to lncRNA localization (30). AnnoLnc2 will report known localization-related motifs in the query sequence based on a curated list compiled from published researches (Supplementary Table S3) (31–35). Meanwhile, we also implemented a dedicated analysis module for running motif identification (64) over a set of selected lncRNAs in the standalone version.

Function and interaction. lncRNAs play functional roles via various interactions (36). AnnoLnc2 reports lncRNA–protein interactions based on both experimentally validated data and sequence-oriented prediction (37). Briefly, we first manually curated and processed recently published human and mouse CLIP-seq datasets for RNA-binding proteins (RBPs) from GEO (see Supplementary Table S4 for a detailed list as well as Supplementary Figure S2 for details on data processing) and then merged these newly identified binding sites with published sites (38) to obtain the final catalog. By integrating 385 human CLIP-seq datasets (for 188 RBPs) and 238 mouse CLIP-seq datasets (for 62 RBPs), AnnoLnc2 nearly quadruples the number of experimentally validated RNA-binding proteins compared with that of AnnoLnc and provides the most comprehensive online resource for CLIP-seq-based lncRNA–protein interaction annotation so far. Similar to that of miRNA binding sites, each protein binding site can be viewed in a structural context. For RNA-binding proteins out of the list, we utilized lncPro (37) to predict lncRNA–protein interaction *ab initio*. Moreover, benefiting from its own effective expression profiling module, AnnoLnc2 also systematically computes all coexpressed genes of the inputted lncRNA and annotates the lncRNA with enriched GO terms for these genes at the biological process and molecular function levels using the ‘Guilt-by-Association’ strategy (5,22,39).

Evolution and association. Association studies such as genome-wide association studies (GWASs) and expression quantitative trait loci (eQTLs) offer unique opportunities to characterize lncRNAs associated with diseases and other traits (40,41). In addition to annotating the inputted lncRNA with variants from the NHGRI GWAS Catalog (42) (with ~25% more candidate variants compared to that in AnnoLnc), AnnoLnc2 also integrates all human eQTLs from GTEx (41) and mouse phenotypic and QTL alleles from MGI (43). When reporting human GWAS SNPs, all SNPs linked with the tag SNP are also reported, along with

their position relative to the aligned lncRNA (e.g. ‘promoter’, ‘exon’ or ‘intron’) and all relevant information of the tag SNP (i.e. the associated trait, the *P*-value and the PubMed ID of the paper reporting this SNP). When reporting mouse alleles, the allele’s type, ID, symbol and name, its associated marker’s ID and name as well as the PubMed ID of the paper reporting the allele are listed.

Selection is a direct indicator of biological functions (3,44). To effectively identify selection signals across various evolutionary clades, we reimplemented the UCSC MultiZ-Phast pipeline (45,46) and recomputed conservation scores for the primate, mammal and vertebrate clades. Briefly, we downloaded the 100-way human multiple alignment data from UCSC Genome Browser, constructed a phylogeny for each clade with phyloFit (45), and computed phastCons and phyloP scores by PHAST and phyloP, respectively. In addition, the server will also list the derived allele frequency (DAF) of all the SNPs from the 1000 Genomes Project (47) that fall within the lncRNA locus to analyze potential population-scale selection.

Powerful and user-friendly interfaces

AnnoLnc2 takes RNA sequences in fasta format as input, and users can paste or upload sequences on the AnnoLnc2 homepage and choose corresponding species from a dropdown option list. After that, users simply need to click the ‘GO’ button, and AnnoLnc2 will analyze all inputted sequences (up to 100 sequences) simultaneously (Figure 2A, B). After a job is submitted, AnnoLnc2 will return a link to users for retrieving results in the future. AnnoLnc2 provides an item for each best alignment of all sequences, and users can view detailed annotation results from the link in the ‘Status’ column. All of the results are shown in interactive figures and tables and allow users to view detailed numeric information and search for items of interest, such as TFs, RBPs, traits, or functions. Of note, AnnoLnc2 offers a ‘Summary’ text section for briefly describing the annotation results for each module. To make using AnnoLnc2 more convenient, AnnoLnc2 supports batch uploading and downloading to avoid laboring operations for fetching results one by one. The AnnoLnc2 web server is now hosted on an elastic cloud server from the Huawei Cloud running an Ubuntu Linux system (18.04 LTS) with 16 CPU and 64 GB memory.

Moreover, in response to user requests, we also provide a standalone version of AnnoLnc2 for large-scale offline analysis (available from <http://annolnc.gao-lab.org/download.php>). In addition to providing all functionalities of the online server, the standalone AnnoLnc2 package is fully customized, and users are able to not only specify their own datasets for modules such as expression profiling and lncRNA–protein interaction but also fine tune/change the code for their specific requirements (or even port it to an entirely new species!). To facilitate customization, we also provide the online AnnoLnc2 Developer Kit as a set of scripts and guidelines for particular conversion and maintenance tasks. Currently, the standalone package requires a reasonable Linux server with a minimal requirement of 8 GB memory (see <http://annolnc.gao-lab.org/download.php> for the installation guide).

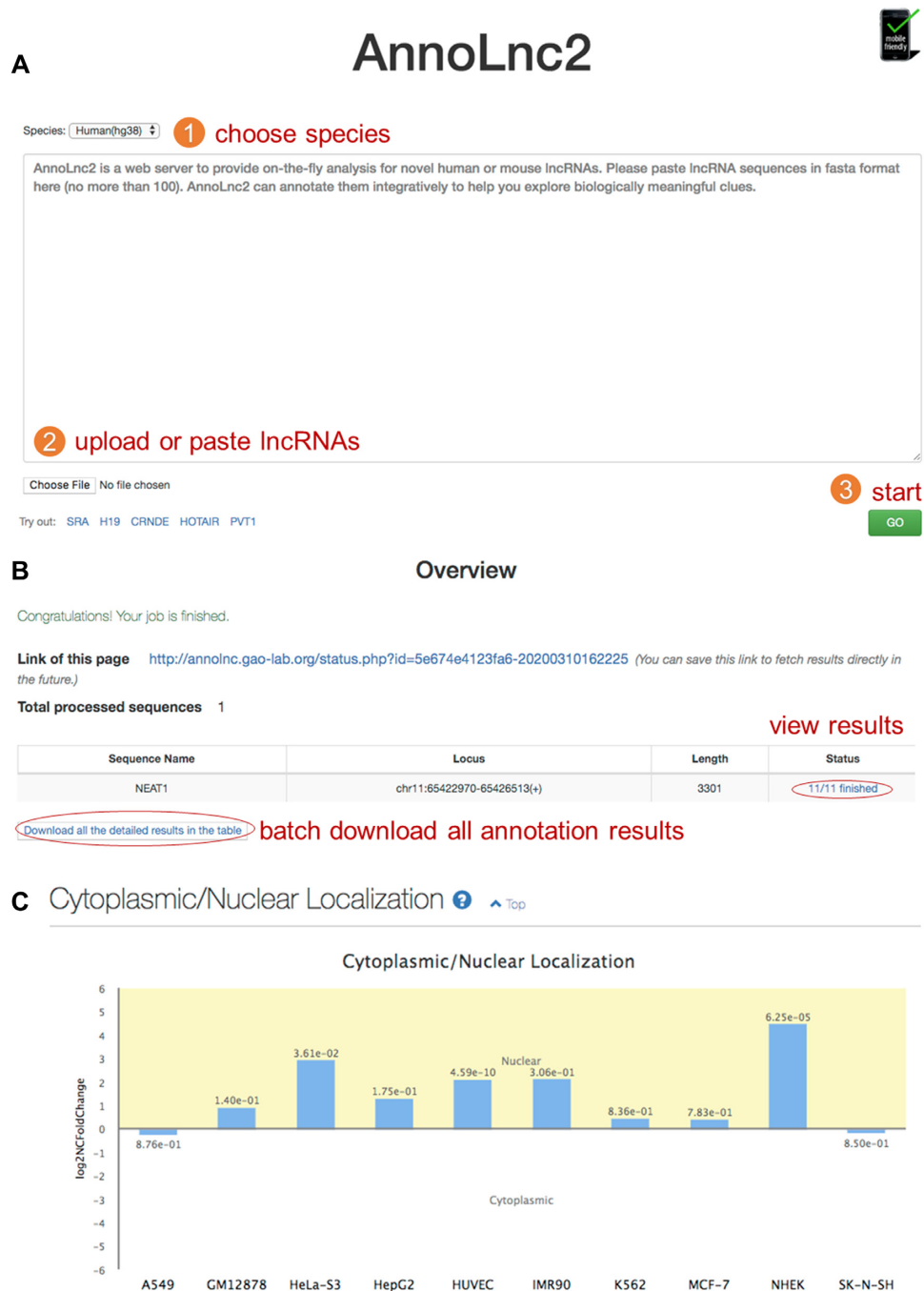


Figure 2. AnnoLnc2 web server. Users can run AnnoLnc2 through a three-step operation (A) and view detailed annotation results, as well as download all results by one click (B). A well-studied lncRNA, NEAT1, is an important component of nuclear paraspeckles (63). (C) Annotation results from the subcellular localization module show that NEAT1 is strongly located in the nucleus in multiple cell lines. Each bar represents logarithm scaled nuclear/cytoplasmic localization ratio, with corresponding FDR marked upon it.

DISCUSSION

As the successor of AnnoLnc, AnnoLnc2 supports annotating both human and mouse novel lncRNAs with wider and more comprehensive data resources and provides a standalone version for large-scale customized annotation, along with a more responsive and user-friendly Web interface. The major improvements are listed below (also see the full list at <http://annolnc.gao-lab.org/help.php#link-update>).

- AnnoLnc2 is now available for both *Mus Musculus* (mm10) and *Homo sapiens* (hg38), compared to *Homo sapiens* (hg19) only in AnnoLnc.
- A novel module has been added for assessing lncRNAs' subcellular localization based on nucleus/cytosol-separated profiling data (29).
- All annotations have been updated with more datasets incorporated. For example, the ChIP-seq dataset now cov-

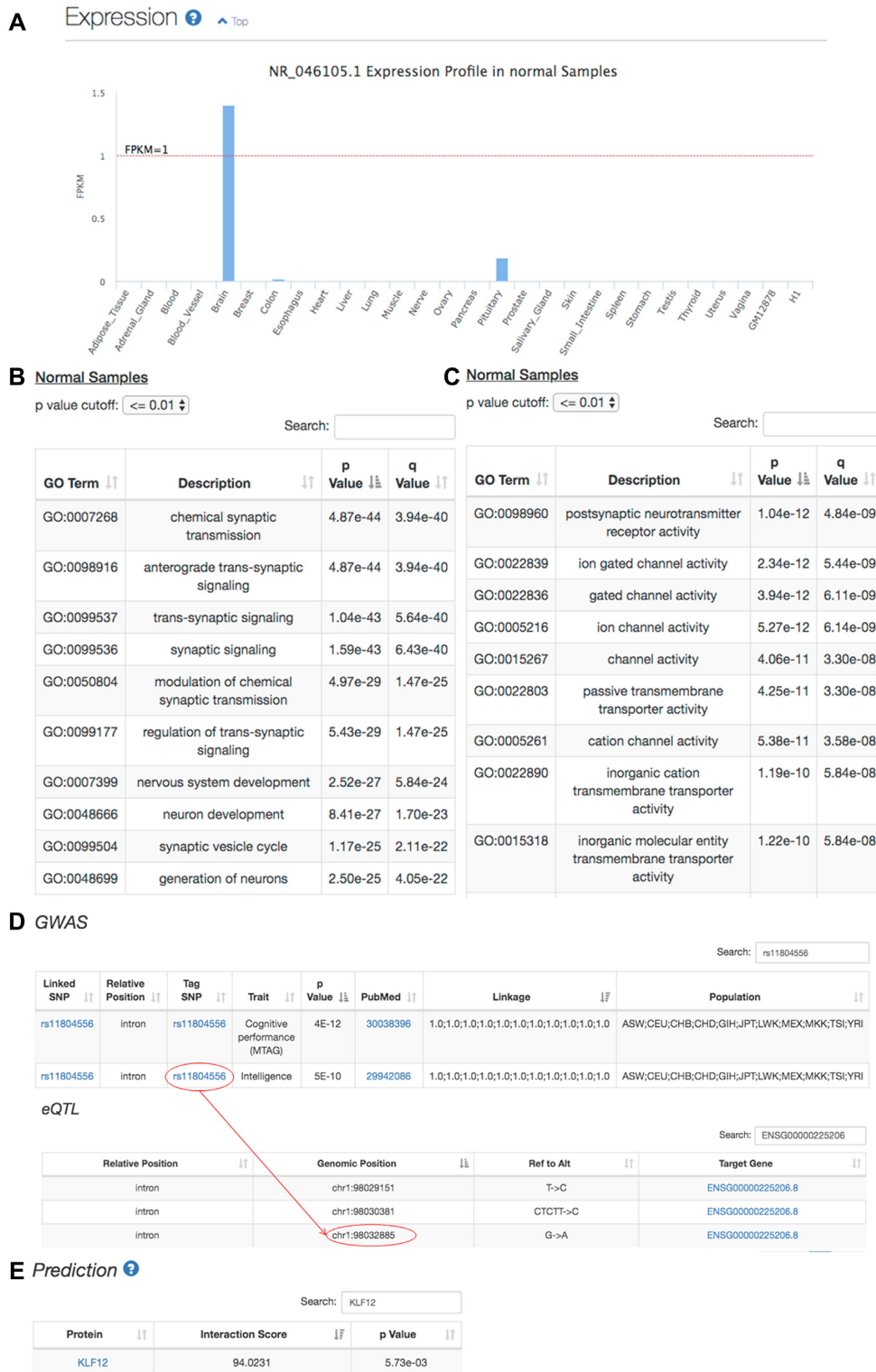


Figure 3. Case study of the human lncRNA NR_046105.1. (A) Expression profile of NR_046105.1 in normal samples, where it is exclusively highly expressed in the brain. (B-C) Functions of NR_046105.1 predicted by the coexpression-based functional annotation module at the biological process (B) or molecular function (C) level. (D) Annotation results from the genetic association module. The SNP rs11804556 is not only a tag SNP but also an eQTL. (E) KLF12 is predicted to interact with NR_046105.1 by the protein interaction module.

A miRNA Regulation Top

Search:

miRNA Families	Start	End	Conservation Score				CLIP Support	Binding Sites in Structure
			Glire	Euarchontoglires	Placental	Vetebrate		
miR-128-3p	1153	1158	0.12	0.00	0.00	0.00	View	

B Cell line Samples

p value cutoff:

Search:

GO Term	Description	p Value	q Value
GO:0002683	negative regulation of immune system process	3.36e-8	5.43e-04
GO:1903707	negative regulation of hemopoiesis	1.08e-6	8.69e-03
GO:0002691	regulation of cellular extravasation	1.75e-6	9.15e-03
GO:0045653	negative regulation of megakaryocyte differentiation	2.27e-6	9.15e-03
GO:0002682	regulation of immune system process	3.98e-6	1.28e-02
GO:0030097	hemopoiesis	1.20e-5	2.61e-02
GO:0002819	regulation of adaptive immune response	1.24e-5	2.61e-02
GO:0045638	negative regulation of myeloid cell differentiation	1.29e-5	2.61e-02
GO:0051606	detection of stimulus	1.79e-5	3.21e-02
GO:0002520	immune system development	2.48e-5	3.59e-02

D Tissue Samples

p value cutoff:

Search:

GO Term	Description	p Value	q Value
GO:0006412	translation	6.17e-16	8.60e-12
GO:0043043	peptide biosynthetic process	1.07e-15	8.60e-12
GO:0043604	amide biosynthetic process	1.96e-14	1.06e-10
GO:0006518	peptide metabolic process	4.56e-14	1.84e-10
GO:0043603	cellular amide metabolic process	4.44e-12	1.43e-08
GO:0009059	macromolecule biosynthetic process	1.07e-10	2.89e-07
GO:0034645	cellular macromolecule biosynthetic process	2.30e-10	5.31e-07
GO:0044271	cellular nitrogen compound biosynthetic process	2.80e-10	5.65e-07
GO:0009058	biosynthetic process	1.74e-9	3.11e-06
GO:1901566	organonitrogen compound biosynthetic process	2.87e-9	4.31e-06

C Cell line Samples

p value cutoff:

Search:

GO Term	Description	p Value	q Value
GO:0019763	immunoglobulin receptor activity	8.10e-5	2.44e-01
GO:0004984	olfactory receptor activity	1.23e-4	2.44e-01
GO:1901618	organic hydroxy compound transmembrane transporter activity	1.60e-4	2.44e-01

E

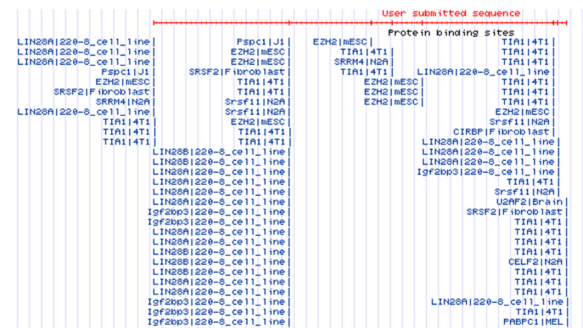


Figure 4. Case study of the mouse lncRNA Pvt1. (A) miRNA annotation results revealed that miR-128-3p might interact with Pvt1. (B–D) Functions of Pvt1 predicted by the coexpression-based functional annotation module at the biological process (B) or molecular function (C) level in cell line samples and at the biological process level in tissue samples (D). (E) Integrated view in UCSC Genome Browser displaying the binding spectrum of LIN28A and TIA1 on Pvt1.

ers 1339 human TFs and 738 mouse TFs (versus 159 human TFs in AnnoLnc), and the updated CLIP-seq dataset contains 188 human RBPs and 62 mouse RBPs (versus 51 human RBPs in AnnoLnc).

- In response to numerous requests from multiple users, a standalone version is available for large-scale offline analysis.

To the best of our knowledge, AnnoLnc2 is currently the most comprehensive annotation tool for analyzing novel lncRNAs in both human and mouse. These updates significantly enhance AnnoLnc2 as an effective and efficient tool for annotating lncRNAs and inspiring hypotheses for further investigation, which we would like to demonstrate below with two cases.

The polymorphisms in human lncRNA NR_046105.1 were reported to influence schizophrenia (48). Based on the results from the genetic association module, we found 72 SNPs (71 intronic) enriched in the NR_046105.1 gene locus. Among them, 69 SNPs are in linkage disequilibrium with schizophrenia-associated tag SNPs, e.g. rs1625579 (48,49) (Supplementary Table S5). Consistently, the expression profile (Figure 3A) indicates that NR_046105.1 is exclusively highly expressed in the brain, which supports the important roles of NR_046105.1 in the brain.

We then explored how NR_046105.1 performs its function. The coexpression-based functional annotation module found that NR_046105.1 is involved in ‘chemical synaptic transmission’ and ‘synaptic signaling’ from the ‘biological process’ category (Figure 3B), which is consistent with the finding of a previous study (50). In addition, its enriched ‘molecular function’ terms suggested that NR_046105.1 regulates the activity of neurotransmitter reporters and ion channels (Figure 3C), which may partially explain how it affects synaptic transmission and leads to brain disease. Further inspection of GWAS SNP and eQTL annotations showed that the SNP rs11804556 is associated with ‘cognitive performance’ and the expression of NR_046105.1 (Figure 3D), suggesting that this variant affects cognition by regulating gene expression. Based on the results from the genetic association module, we also found that most eQTLs (98/101) in NR_046105.1 influence DPYD expression. Considering that DPYD was reported to be involved in pancreatic cancer (51), as is miR-137 (product of NR_046105.1) (52,53), these eQTLs may provide an explanation for their relationship. Moreover, He *et al.* (53) found that miR-137 plays a role in pancreatic cancer by targeting KLF12, and the protein interaction module of AnnoLnc2 also discovered this interaction (Figure 3E).

Similarly, the well-studied mouse lncRNA Pvt1 (Ensembl ID: ENSMUST00000180432.8) was reported to regulate atrial fibrosis by acting as a sponge of miR-128-3p (54), which was also identified by the miRNA regulation module of AnnoLnc2 (Figure 4A). Moreover, Pvt1 was predicted to play roles in the regulation of immune system processes based on enriched functions at the biological process level in cell lines (Figure 4B), in agreement with the finding of a previous study (55). According to the results from the functional annotation module, this process may be achieved by regulating ‘immunoglobulin receptor activity’ (Figure 4C). In addition, we found that Pvt1 might regulate translation

(Figure 4D), which was further supported by the annotated multiple binding sites of LIN28A and TIA1 on Pvt1 (Figure 4E), both of which have been reported to act as regulators of mRNA translation (56,57). Overall, AnnoLnc2 leads to the new hypothesis that Pvt1 is involved in translation regulation by acting as a sponge of the RNA binding proteins LIN28A and TIA1.

During past years, the AnnoLnc web server has been widely used by the community, serving 50 000+ users around the world, with millions of sequences analyzed. In the future, we will continue our efforts to improve the services provided by AnnoLnc to research communities. More specifically, we will keep updating backend data sources and support more species, including fruit fly and nematode, as well as plants, such as *Arabidopsis*. Beyond the current modules, other types of helpful annotations, such as RNA modification and machine learning-based function or localization prediction, could also be incorporated to better understand lncRNA functionalities and potential mechanisms. Finally, we recognize the urgent need for a user-friendly and interactive interface for a wider research community and will continue improving the server interface based on users’ feedback.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank multiple scientists at UCSC Genome Browse for their help with conservation score computation. The analyses are supported by the High-performance Computing Platform of Peking University, and we thank Dr Chun Fan and Yin-Ping Ma for their assistance during the analysis.

FUNDING

National Key Research and Development Program [2016YFC0901603]; China 863 Program [2015AA020108]; State Key Laboratory of Protein and Plant Gene Research and the Beijing Advanced Innovation Center for Genomics (ICG) at Peking University; National Program for Support of Top-notch Young Professionals (to G.G.). Funding for open access charge: National Key Research and Development Program [2016YFC0901603]; China 863 Program [2015AA020108]; State Key Laboratory of Protein and Plant Gene Research and the Beijing Advanced Innovation Center for Genomics (ICG) at Peking University; National Program for Support of Top-notch Young Professionals (to G.G.).

Conflict of interest statement. None declared.

REFERENCES

1. Ulitsky, I. and Bartel, D.P. (2013) lincRNAs: genomics, evolution, and mechanisms. *Cell*, **154**, 26–46.
2. Marchese, F.P., Raimondi, I. and Huarte, M. (2017) The multidimensional mechanisms of long noncoding RNA function. *Genome Biol.*, **18**, 206.
3. Ponjavic, J., Ponting, C.P. and Lunter, G. (2007) Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res.*, **17**, 556–565.

4. Zhang, K., Shi, Z.M., Chang, Y.N., Hu, Z.M., Qi, H.X. and Hong, W. (2014) The ways of action of long non-coding RNAs in cytoplasm and nucleus. *Gene*, **547**, 1–9.
5. Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P. *et al.* (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, **458**, 223–227.
6. Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M. *et al.* (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*, **149**, 1393–1406.
7. Yoon, J.-H., Abdelmohsen, K. and Gorospe, M. (2014) Functional interactions among microRNAs and long noncoding RNAs. *Semin. Cell Dev. Biol.*, **34**, 9–14.
8. Li, J., Ma, W., Zeng, P., Wang, J., Geng, B., Yang, J. and Cui, Q. (2015) LncTar: a tool for predicting the RNA targets of long noncoding RNAs. *Brief. Bioinform.*, **16**, 806–812.
9. Zhang, W., Yue, X., Tang, G., Wu, W., Huang, F. and Zhang, X. (2018) SFPEL-LPI: sequence-based feature projection ensemble learning for predicting LncRNA-protein interactions. *PLOS Comput. Biol.*, **14**, e1006616.
10. Zhou, J., Huang, Y., Ding, Y., Yuan, J., Wang, H. and Sun, H. (2018) lncFunTK: a toolkit for functional annotation of long noncoding RNAs. *Bioinformatics*, **34**, 3415–3416.
11. Cao, Z., Pan, X., Yang, Y., Huang, Y. and Shen, H.-B. (2018) The lncLocator: a subcellular localization predictor for long non-coding RNAs based on a stacked ensemble classifier. *Bioinformatics*, **34**, 2185–2194.
12. Su, Z.-D., Huang, Y., Zhang, Z.-Y., Zhao, Y.-W., Wang, D., Chen, W., Chou, K.-C. and Lin, H. (2018) iLoc-lncRNA: predict the subcellular location of lncRNAs by incorporating octamer composition into general PseKNC. *Bioinformatics*, **34**, 4196–4204.
13. Ma, L., Cao, J., Liu, L., Du, Q., Li, Z., Zou, D., Bajic, V.B. and Zhang, Z. (2019) LncBook: a curated knowledgebase of human long non-coding RNAs. *Nucleic Acids Res.*, **47**, 2699.
14. Fang, S., Zhang, L., Guo, J., Niu, Y., Wu, Y., Li, H., Zhao, L., Li, X., Teng, X., Sun, X. *et al.* (2018) NONCODEV5: a comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Res.*, **46**, D308–D314.
15. Hou, M., Tang, X., Tian, F., Shi, F., Liu, F. and Gao, G. (2016) AnnoLnc: a web server for systematically annotating novel human lncRNAs. *BMC Genomics*, **17**, 931.
16. Tarailo-Graovac, M. and Chen, N. (2009) Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinform.*, doi:10.1002/0471250953.bi0410s25.
17. Johnson, R. and Guigo, R. (2014) The RIDL hypothesis: transposable elements as functional domains of long noncoding RNAs. *RNA*, **20**, 959–976.
18. Lee, H., Zhang, Z. and Krause, H.M. (2019) Long noncoding RNAs and repetitive elements: junk or intimate evolutionary partners? *Trends Genet.*, **35**, 892–902.
19. Fabbri, M., Girmata, L., Varani, G. and Calin, G.A. (2019) Decrypting noncoding RNA interactions, structures, and functional networks. *Genome Res.*, **29**, 1377–1388.
20. Qian, X., Zhao, J., Yeung, P.Y., Zhang, Q.C. and Kwok, C.K. (2019) Revealing lncRNA Structures and interactions by sequencing-based approaches. *Trends Biochem. Sci.*, **44**, 33–52.
21. Lorenz, R., Bernhart, S.H., Höner zu Siederdisen, C., Tafer, H., Flamm, C., Stadler, P.F. and Hofacker, I.L. (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.*, **6**, 26.
22. Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G. *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.
23. Goff, L.A., Groff, A.F., Sauvageau, M., Traves-Gibson, Z., Sanchez-Gomez, D.B., Morse, M., Martin, R.D., Elcavage, L.E., Liapis, S.C., Gonzalez-Celeiro, M. *et al.* (2015) Spatiotemporal expression and transcriptional perturbations by long noncoding RNAs in the mouse brain. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 6855–6862.
24. Hou, M., Tian, F., Jiang, S., Kong, L., Yang, D. and Gao, G. (2016) LocExpress: A web server for efficiently estimating expression of novel transcripts. *BMC Genomics*, **17**, 1023.
25. Yevshin, I., Sharipov, R., Kolmykov, S., Kondrakhin, Y. and Kolpakov, F. (2019) GTRD: a database on gene transcription regulation-2019 update. *Nucleic Acids Res.*, **47**, D100–D105.
26. Agarwal, V., Bell, G.W., Nam, J.W. and Bartel, D.P. (2015) Predicting effective microRNA target sites in mammalian mRNAs. *Elife*, **4**, e05005.
27. Cabili, M.N., Dunagin, M.C., McClanahan, P.D., Biaesch, A., Padovan-Merhar, O., Regev, A., Rinn, J.L. and Raj, A. (2015) Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol.*, **16**, 20.
28. Chen, L.-L. (2016) Linking long noncoding RNA localization and function. *Trends Biochem. Sci.*, **41**, 761–772.
29. Djebali, S., Davis, C.A., 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.
30. Gandhi, M., Caudron-Herger, M. and Diederichs, S. (2018) RNA motifs and combinatorial prediction of interactions, stability and localization of noncoding RNAs. *Nat. Struct. Mol. Biol.*, **25**, 1070–1076.
31. Zhang, B., Gunawardane, L., Niazi, F., Jahanbani, F., Chen, X. and Valadkhan, S. (2014) A novel RNA motif mediates the strict nuclear localization of a long noncoding RNA. *Mol. Cell. Biol.*, **34**, 2318–2329.
32. Lubelsky, Y. and Ulitsky, I. (2018) Sequences enriched in Alu repeats drive nuclear localization of long RNAs in human cells. *Nature*, **555**, 107–111.
33. Shukla, C.J., McCorkindale, A.L., Gerhardinger, C., Korthauer, K.D., Cabili, M.N., Shechner, D.M., Irizarry, R.A., Maass, P.G. and Rinn, J.L. (2018) High-throughput identification of RNA nuclear enrichment sequences. *EMBO J.*, **37**, e98452.
34. Carlevaro-Fita, J., Polidori, T., Das, M., Navarro, C., Zoller, T.I. and Johnson, R. (2019) Ancient exapted transposable elements promote nuclear enrichment of human long noncoding RNAs. *Genome Res.*, **29**, 208–222.
35. Yin, Y., Lu, J.Y., Zhang, X., Shao, W., Xu, Y., Li, P., Hong, Y., Cui, L., Shan, G., Tian, B. *et al.* (2020) U1 snRNP regulates chromatin retention of noncoding RNAs. *Nature*, **580**, 147–150.
36. Rinn, J.L. and Chang, H.Y. (2012) Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.*, **81**, 145–166.
37. Lu, Q., Ren, S., Lu, M., Zhang, Y., Zhu, D., Zhang, X. and Li, T. (2013) Computational prediction of associations between long non-coding RNAs and proteins. *BMC Genomics*, **14**, 651.
38. Zhu, Y., Xu, G., Yang, Y.T., Xu, Z., Chen, X., Shi, B., Xie, D., Lu, Z.J. and Wang, P. (2019) POSTAR2: Deciphering the post-Transcriptional regulatory logics. *Nucleic Acids Res.*, **47**, D203–D211.
39. Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A. and Rinn, J.L. (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.*, **25**, 1915–1927.
40. Hindorf, L.a., Sethupathy, P., Junkins, H.a., Ramos, E.M., Mehta, J.P., Collins, F.S. and Manolio, T.a. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 9362–9367.
41. Ardlie, K.G., Deluca, D.S., Segre, A. V., Sullivan, T.J., Young, T.R., Gelfand, E.T., Trowbridge, C.A., Maller, J.B., Tukiainen, T., Lek, M. *et al.* (2015) The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science*, **348**, 648–660.
42. Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E. *et al.* (2019) The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.*, **47**, D1005–D1012.
43. Bult, C.J., Blake, J.A., Smith, C.L., Kadin, J.A., Richardson, J.E., Agnostopoulos, A., Asabor, R., Baldarelli, R.M., Beal, J.S., Bello, S.M. *et al.* (2019) Mouse genome database (MGD) 2019. *Nucleic Acids Res.*, **47**, D801–D806.
44. Church, D.M., Goodstadt, L., Hillier, L.W., Zody, M.C., Goldstein, S., She, X., Bult, C.J., Agarwala, R., Cherry, J.L., DiCuccio, M. *et al.* (2009) Lineage-specific biology revealed by a finished genome assembly of the mouse. *PLoS Biol.*, **7**, e1000112.
45. Hubisz, M.J., Pollard, K.S. and Siepel, A. (2011) PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief. Bioinform.*, **12**, 41–51.

46. Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R. and Siepel, A. (2010) Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.*, **20**, 110–121.
47. 1000 Genomes Project Consortium, Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E. and McVean, G.A. (2010) A map of human genome variation from population-scale sequencing. *Nature*, **467**, 1061–1073.
48. Wright, C., Gupta, C.N., Chen, J., Patel, V., Calhoun, V.D., Ehrlich, S., Wang, L., Bustillo, J.R., Perrone-Bizzozero, N.I. and Turner, J.A. (2016) Polymorphisms in MIR137HG and microRNA-137-regulated genes influence gray matter structure in schizophrenia. *Transl. Psychiatry*, **6**, e724.
49. Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium (2011) Genome-wide association study identifies five new schizophrenia loci. *Nat. Genet.*, **43**, 969–976.
50. He, E., Lozano, M.A.G., Stringer, S., Watanabe, K., Sakamoto, K., den Ouden, F., Koopmans, F., Giamberardino, S.N., Hammerschlag, A., Cornelisse, L.N. *et al.* (2018) MIR137 schizophrenia-associated locus controls synaptic function by regulating synaptogenesis, synapse maturation and synaptic transmission. *Hum. Mol. Genet.*, **27**, 1879–1891.
51. Elander, N.O., Aughton, K., Ghaneh, P., Neoptolemos, J.P., Palmer, D.H., Cox, T.F., Campbell, F., Costello, E., Halloran, C.M., Mackey, J.R. *et al.* (2018) Expression of dihydropyrimidine dehydrogenase (DPD) and hENT1 predicts survival in pancreatic cancer. *Br. J. Cancer*, **118**, 947–954.
52. Lei, S., He, Z., Chen, T., Guo, X., Zeng, Z., Shen, Y. and Jiang, J. (2019) Long noncoding RNA 00976 promotes pancreatic cancer progression through OTUD7B by sponging miR-137 involving EGFR/MAPK pathway. *J. Exp. Clin. Cancer Res.*, **38**, 470.
53. He, Z., Guo, X., Tian, S., Zhu, C., Chen, S., Yu, C., Jiang, J. and Sun, C. (2019) MicroRNA-137 reduces stemness features of pancreatic cancer cells by targeting KLF12. *J. Exp. Clin. Cancer Res.*, **38**, 126.
54. Cao, F., Li, Z., Ding, W., Yan, L. and Zhao, Q. (2019) LncRNA PVT1 regulates atrial fibrosis via miR-128-3p-SPI1-TGF- β 1-Smad axis in atrial fibrillation. *Mol. Med.*, **25**, 7.
55. Zheng, Y., Tian, X., Wang, T., Xia, X., Cao, F., Tian, J., Xu, P., Ma, J., Xu, H. and Wang, S. (2019) Long noncoding RNA Pvt1 regulates the immunosuppression activity of granulocytic myeloid-derived suppressor cells in tumor-bearing mice. *Mol. Cancer*, **18**, 61.
56. Cho, J., Chang, H., Kwon, S.C., Kim, B., Kim, Y., Choe, J., Ha, M., Kim, Y.K. and Kim, V.N. (2012) LIN28A is a suppressor of ER-associated translation in embryonic stem cells. *Cell*, **151**, 765–777.
57. Díaz-Muñoz, M.D., Kiselev, V.Y., Le Novère, N., Curk, T., Ule, J. and Turner, M. (2017) Tia1 dependent regulation of mRNA subcellular location and translation controls p53 expression in B cells. *Nat. Commun.*, **8**, 530.
58. Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N. *et al.* (2013) The genotype-tissue expression (GTEx) project. *Nat. Genet.*, **45**, 580–585.
59. Wilks, C., Cline, M.S., Weiler, E., Diehkans, M., Craft, B., Martin, C., Murphy, D., Pierce, H., Black, J., Nelson, D. *et al.* (2014) The Cancer Genomics Hub (CGHub): overcoming cancer through the power of torrential data. *Database*, **2014**, bau093.
60. Feingold, E.A., Good, P.J., Guyer, M.S., Kamholz, S., Liefer, L., Wetterstrand, K., Collins, F.S., Gingeras, T.R., Kampa, D., Sekinger, E.A. *et al.* (2004) The ENCODE (ENCyclopedia of DNA Elements) project. *Science*, **306**, 636–640.
61. Sherry, S.T. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.
62. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, A.D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.
63. Bond, C.S. and Fox, A.H. (2009) Paraspeckles: nuclear bodies built on long noncoding RNA. *J. Cell Biol.*, **186**, 637–644.
64. Sharov, A.A. and Ko, M.S. (2009) Exhaustive search for over-represented DNA sequence motifs with CisFinder. *DNA Res.*, **16**, 261–273.