

# A comprehensive method protocol for annotation and integrated functional understanding of lncRNAs

Meik Kunz\*, Beat Wolf\*, Maximilian Fuchs\*, Jan Christoph, Ke Xiao, Thomas Thum, David Atlan, Hans-Ulrich Prokosch and Thomas Dandekar

Corresponding authors: Thomas Dandekar, Functional Genomics and Systems Biology Group, Department of Bioinformatics, Biocenter, Würzburg, Germany. Tel.: ++49-931-3184551; Fax: ++49-931-3184552; E-mail: dandekar@biozentrum.uni-wuerzburg.de; Meik Kunz, Chair of Medical Informatics, Friedrich-Alexander University of Erlangen-Nürnberg, Erlangen, Germany. Tel.: ++49-9131-85-26767; Fax: ++49-9131-85-26754; E-mail: meik.kunz@fau.de

\*Equal contribution.

## Abstract

Long non-coding RNAs (lncRNAs) are of fundamental biological importance; however, their functional role is often unclear or loosely defined as experimental characterization is challenging and bioinformatic methods are limited. We developed a novel integrated method protocol for the annotation and detailed functional characterization of lncRNAs within the genome. It combines annotation, normalization and gene expression with sequence-structure conservation, functional interactome and promoter analysis. Our protocol allows an analysis based on the tissue and biological context, and is powerful in functional characterization of experimental and clinical RNA-Seq datasets including existing lncRNAs. This is demonstrated on the uncharacterized lncRNA GATA6-AS1 in dilated cardiomyopathy.

**Key words:** method protocol; lncRNAs; annotation; gene expression analysis; sequence-structure analysis; integrated functional analysis

## Introduction

In recent years, long non-coding RNAs (lncRNAs) have been identified as functional players that regulate important biological processes and signaling pathways associated with pathogenesis and disease progression [1–6]. Hundreds of lncRNAs were annotated, but only few are experimentally characterized due to the fact that lncRNAs have complex regulatory functions, such as

binding of mRNAs, miRNAs and proteins [3, 7]. Bioinformatics approaches, such as phylogenetic and functional interaction analysis, are helpful to comprehensively understand lncRNA functions for experimental characterization [3].

Several databases, such as LNCipedia [7], NONCODE [8] and LncRBase [9], have been developed, but all these mainly provide general information about already published lncRNAs without considering the biological context. Consequently, several

Meik Kunz is group leader at the Chair of Medical Informatics, Friedrich-Alexander University of Erlangen-Nürnberg, Erlangen, Germany.

Beat Wolf is postdoc at the University of Applied Sciences and Arts of Western Switzerland, Fribourg, Switzerland.

Maximilian Fuchs is PhD student at the Department of Bioinformatics, University of Würzburg, Germany.

Jan Christoph is postdoc at the Chair of Medical Informatics, Friedrich-Alexander University of Erlangen-Nürnberg, Erlangen, Germany.

Ke Xiao is group leader at the Institute of Molecular and Translational Therapeutic Strategies, Hannover Medical School, Hannover, Germany.

Thomas Thum is chair and full professor at the Institute of Molecular and Translational Therapeutic Strategies and member of the REBIRTH Excellence Cluster (Hannover Medical School, Hannover, Germany) as well as visiting professor at the National Heart and Lung Institute, Imperial College London, London, UK.

David Atlan is CEO of the NGS data analysis company Phenosystems SA, Braine le Château, Belgium.

Hans-Ulrich Prokosch is chair and full professor at the Chair of Medical Informatics, Friedrich-Alexander University of Erlangen-Nürnberg, Erlangen, Germany.

Thomas Dandekar is chair and full professor at the Department of Bioinformatics, University of Würzburg, Germany.

Submitted: 21 December 2018; Received (in revised form): 29 April 2019

© The Author(s) 2019. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

integrated bioinformatics tools have been designed for *de novo* analysis. Examples include starBase, which focuses on Pan-cancer analysis and interaction networks of lncRNAs based on large-scale tumor datasets [10], and the lncRNAtor tool, which helps in functional understanding of lncRNAs using data sets collected from The Cancer Genome Atlas as an example [11]. However, these tools focus on specific diseases and only provide annotated and already analyzed data from publicly available datasets. This renders them limited, especially for analysis of in-house experimental datasets as well as user-specific functional characterization of newly detected lncRNAs. Improved methods for comprehensive annotation and integrated functional understanding of lncRNAs are therefore needed.

To address this, we developed a method protocol for combined annotation and functional analysis of lncRNAs. We applied different bioinformatics approaches, such as sequence-structure conservation and regulatory network analysis, and functional context including experimentally validated genome and interactome data. The analysis protocol is innovative, time efficient and easy to use. It can therefore be used in labs as a stand-alone method for sequencing data analysis and functional characterization of lncRNAs. A detailed protocol is given in the supplementary material. Moreover, the complete workflow with all tools is implemented as a free academic trial in the Next Generation Sequencing (NGS) data analysis pipeline GensearchNGS.

## Methods

Detailed methods are given in the supplementary material. A step-by-step protocol including Graphical User Interface (GUI) screenshot examples guides the users how to use our method, and illustrates which steps are automatically and semi-automatically implemented depending on the user interest and focus (supplementary material). Figure 1 summarizes our method protocol. It focuses on (i) annotation and gene expression analysis, (ii) sequence-structure conservation and (iii) functional analysis (see supplementary material for details).

- (i) Experimental data were automatically annotated with their gene ontology (GO) terms, phenotypes and pathway associations. Read counting is based on the Gencode gene model from the Ensembl database. Expressed genes are normalized using our own implementation of the DESeq2 normalization method, whereas differential expressed genes were calculated using a t-test. The results of the gene expression analysis are shown as an interactive list, and for a more visual overview as a heatmap.
- (ii) Orthologs from Ensembl and LNCipedia are used. As lncRNAs are not well conserved among species [12, 13], both databases do not provide information about lncRNA orthologous sequences. Therefore, if no orthologs are available, a BLAST analysis is implemented to manually determine the existence of orthologous sequences across species (parameters: e-value <0.01, identity  $\geq$ 80%; when e-value <0.01, P-value and e-value are nearly identical, [14]). The resulting orthologous sequences can be uploaded to the LocARNA webserver to analyze the phylogenetic sequence-structure conservation. This gives an overview of conserved sequence and structure parts among species and thus can provide information about the interaction context.
- (iii) The function analysis includes a gene enrichment analysis, promoter and regulatory network analysis.

The functional gene enrichment analysis using HPO, OMIM, Orphanet and WikiPathways helps to find overrepresented GO, pathway and phenotype terms for selected differentially expressed genes.

For network reconstruction we use experimentally validated interaction partners (mRNA, miRNA, protein) from several databases, in which a highly confident lncRNA interaction network is generated (see supplementary material). More importantly, since most lncRNAs have no experimentally validated interaction partners, potential mRNA (GATA6-AS1 sequence was split into snippets of 50 bases) and miRNA interaction partners can be predicted using miRanda [15] (parameter miRNA: scoring threshold [-sc 120]; mRNA: [-sc 160], energy [-en -72]), and protein interaction partners using catRAPID (parameter: star rating 2.5; discriminative power >75% represent high confidence predictions, [16]). We added a link to IntaRNA for further lncRNA-mRNA interaction prediction validation [17].

Furthermore, the predicted interactions can be extended using experimentally validated interaction partners from the interactome databases (String, BioGRID, CCSB and NPInter) to build up a knowledge-based interaction network. The script also allows export of the gene list, with its interactions, to Cytoscape using GML files for visualization and advanced functional network analyses. However, these various interactome annotations allow the user to interactively filter the differentially expressed genes according to criteria. Example criteria could include gene association with pathways or ontology terms, either directly or through neighboring genes on the interactome, as well as the tissue-specific expression context. This is an essential step for understanding a gene's tissue-specific and functional role and will help users to find the best candidates for further experimental validation.

For the promoter analysis, the script uses the lncRNA promoter sequence from -2000 to +500 bp relative to the transcription start site from Gencode and further analyzes it for transcription factor binding sites (TFBS) using a link to the Allgen PROMO tool (parameter: 0% matrix dissimilarity value to reduce false positives) [18]. Finally, the identified TFBS are combined with experimentally validated and predicted protein interaction partners. This further helps to understand the expression and complex regulatory effects of lncRNAs, for example cooperative working with an mRNA or a transcription factor (TF) in a function- and tissue-specific manner.

## Results and Discussion

To validate our protocol as an all-in-one method for lncRNA characterization, we performed analysis on a RNA-Seq dataset from GEO (GSE99321). The dataset contains seven samples from patients with dilated cardiomyopathy (DCM) and seven healthy controls. Each sample contains around 40 million reads. The expression analysis shows significant deregulation of 29 lncRNAs (Supplementary Table S1;  $\log_{2}FC > 0.75 / < -0.75$ , P-value < 0.05, False Discovery Rate (FDR) adjusted P-value < 0.45).

In the following we focus on the experimentally uncharacterized lncRNA GATA6-AS1. The sequence-structure analysis shows a sequence-structure conservation of GATA6-AS1 among different mammals (Supplementary Figure S1; Supplementary Table S2; e-value <0.01, identity  $\geq$ 80%).

lncRNAs have been implicated in a wide range of biological processes such as chromatin remodeling, recruitment of transcription factors, regulation of miRNAs, mRNA processing

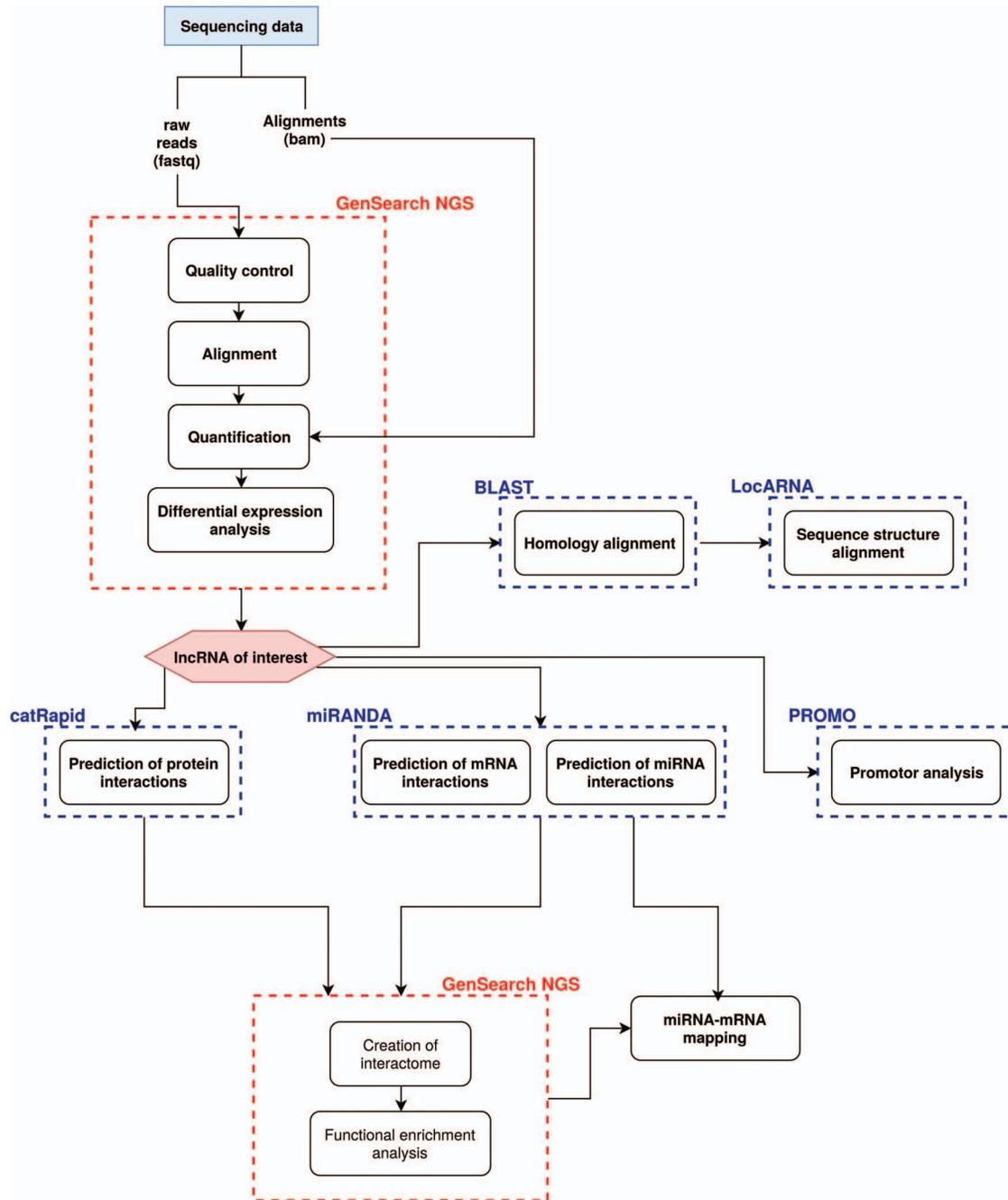


Figure 1. Workflow of our method protocol. Red dot boxes indicate the automatic modules, blue dot boxes indicate the semi-automatic modules (external reference data and programs) of the protocol flow with the analysis steps (black boxes). The input and output node indicate that this is also a stand-alone workflow.

and translation. For this reason, the analysis of potential interaction partners of GATA6-AS1 is essential for its functional characterization. Since we are interested in the functional role of GATA6-AS1 in DCM, we focus on interactions of GATA6-AS1 with mRNAs, miRNAs and proteins, as well as their association with heart diseases. No experimentally validated interaction partners are known for GATA6-AS1. However, prediction analysis identifies 207 mRNA, 41 miRNA and 10 potential protein interaction partners, from which 7 mRNAs (SFRP1, LTBP3, SLC27A3, FAM20A, FCGBP, HS3ST2, GRB14; out of 605 mRNAs in the RNA-Seq) and 1 miRNA (miRNA-4750; out of 3 miRNAs in the

RNA-Seq) show also a significant deregulation in the RNA-Seq data (Supplementary Table S3). This is a small overlap of mRNAs with the RNA-Seq data; however, it is known that interaction prediction tools are typically giving too many false positives (over-prediction). Thus, a mapping of the predicted interactions to the RNA-Seq data is always recommended and indicates strong candidate interactions for further analysis as they show an expression in the RNA-Seq.

Using the direct, experimentally validated interaction partners from NPInter and all predicted GATA6-AS1 interactors results in a knowledge-based interaction network of 2747

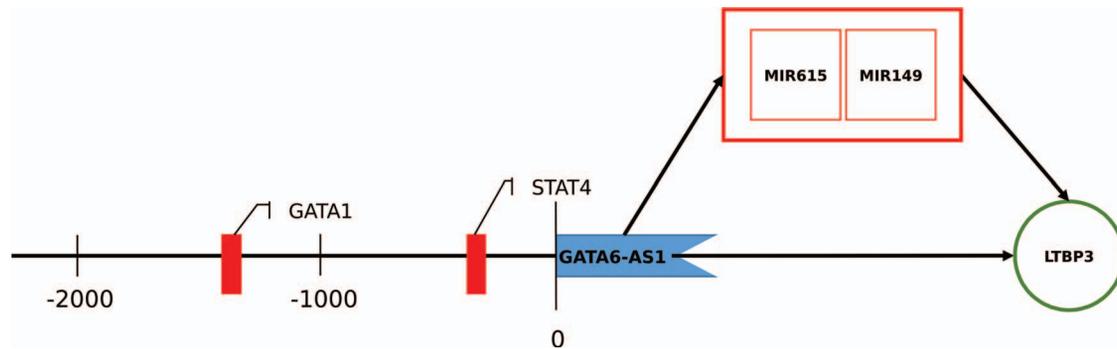


Figure 2. Schematic overview of the regulatory effect of GATA6-AS1 highlighted with our method protocol. GATA6-AS1 binds directly the mRNA LTBP3 or indirectly via sponging of miRNA-149 and miRNA-615. GATA6-AS1 can be regulated by promoter binding of GATA1 and STAT4.

nodes and 66 898 edges (here not shown). Interestingly, we found an overlap with 115 RNAs from the RNA-Seq data (Supplementary Table S4). However, the functional enrichment analysis (HPO, OMIM, Orphanet, WikiPathways) of the interactome shows 55 biological pathways and phenotypes related to cardiac diseases and heart failure (Supplementary Table S5).

To further investigate whether GATA6-AS1 indirectly regulates mRNAs via miRNA sponging, we used experimentally validated mRNA targets from NPInter of the predicted GATA6-AS1 interacting miRNAs and mapped them against the differentially expressed mRNAs from the RNA-Seq and the predicted GATA6-AS1 mRNA interactions (Supplementary Table S6). This shows 40 mRNAs, which are deregulated in the RNA-Seq and experimentally validated miRNA targets. These could be indirectly regulated by GATA6-AS1 as an effect of miRNA sponging. Moreover, 10 experimentally validated mRNA targets of the predicted miRNAs show an overlap with the predicted GATA6-AS1 mRNA interactions. Interestingly, latent transforming growth factor beta binding protein 3 (LTBP3) is also deregulated in the RNA-Seq data, which might indicate a direct regulation of GATA-AS1 via mRNA and miRNA-149 and miRNA-615 targeting (Figure 2). Further prediction validation with IntaRNA shows a potential interaction between GATA-AS1 and LTBP3, thus strengthening the probability of this interaction (Supplementary Table S6).

Finally, the promoter analysis identifies binding sites for 20 potential TFs (0% matrix dissimilarity) in the promoter of GATA6-AS1, in which heart and stress associated factors such as C/EBPbeta, AP-2alphaA, GATA-1 and STAT4 were found (Supplementary Table S7). Detailed results are given in the supplementary materials.

Together, our results suggest a functional role of GATA6-AS1 in extracellular matrix remodeling in DCM via a direct and indirect regulatory effect, e.g. targeting of LTBP3, miRNA-149 and miRNA-615, and activation through promoter binding of STAT4 and GATA1 (Figure 2). Latent-TGF $\beta$ -binding proteins such as LTBP3 are known to mediate the targeting of Transforming Growth Factor (TGF)-beta complexes into extracellular matrix structure [19, 20]. Extracellular matrix remodeling is involved in DCM progression [21], but the role of lncRNAs has not been systematically analyzed and remains still unclear [22, 23]. There is co-expression of GATA6-AS1 and LTBP3, but not regarding miRNA-149 and miRNA-615 (no independent validation dataset for co-expression analysis is available). This might be due to the fact that only three miRNAs are significantly deregulated in the RNA-Seq data. Nevertheless, our analysis strategy allows us to filter the high number of predicted interaction partners to a functional working hypothesis: GATA6-AS1 interactions involve

the deregulated and predicted LTBP3 as well as miRNA-149 and miRNA-615 as predicted GATA6-AS1 miRNAs and known experimentally validated LTBP3 targets. These demonstration results need to be analyzed by validation experiments to examine the role of GATA6-AS1 in DCM. The sequence-structure conservation of GATA6-AS1 in mammals allows development of experimental models.

Interestingly, a recent paper also supports a role of GATA6-AS1 in cardiac development. This was demonstrated by a computational method that detects triplex-forming lncRNAs and DNA targets [24]. This independently supports the ability of our tool to functional characterize lncRNAs.

Notably, our scripts can also be used for analysis of other RNAs, for instance for mRNAs, in particular when combined with suitable refined analysis steps (in particular separate analysis of 5' UTR, CDS and 3' UTR for motifs). We tested this on several miRNAs and mRNAs (examples include [25–27]). As our approach is generic and modular implemented, the protocol is even suitable for genomic analysis to identify variants relevant for disease development, e.g. heart and tumor. Moreover, the method pipeline could be integrated in the IT infrastructure of data integration centers of university hospitals, or molecular tumor boards to improve the university medicine [28].

### Study limitations

Our aim was to develop a general method protocol for lncRNA analysis and functional characterization that is not restricted to specific disease types. We focus on quite common lncRNA interactions with RNAs (miRNAs, mRNAs) and proteins, but we want to point out that more mechanistic interactions exist to describe the complex lncRNA effects [29]. The modularity of our protocol allows to extend it for further analysis depending on the interests of the users, for instance regarding R-loop binding (e.g. R-loopDB [30]) and RNA:DNA triplex formation (several RNA-RNA and RNA-DNA interaction prediction tools can be found in [31]). This makes the protocol useful to a broad community.

RNA sequencing analysis always raises questions regarding sample size and power estimation. Considering different experimental settings that influence the statistical power, we are using DESeq2 for differential expression analysis, which is known to give high performance power, also when the sample size is  $\leq 5$  [32]. We applied multiple-testing correction using the FDR method. We used as cut-offs  $P$ -value  $< 0.05$  and FDR adjusted  $P$ -value  $< 0.45$ . Since high-throughput sequencing methods such as RNA-Seq bear a high error rate, in particular due to sample preparation, small sample sizes and heterogeneous data

from adult and young patient cohorts, detection of differential expressed genes remains difficult. For instance, many genes detected with low FDR adjusted P-value cannot be validated in experiments [33]. In this context, to avoid underestimation of FDRs we used a P-value  $< 0.05$  as significant cut-off criteria to select top candidates for further comprehensive functional characterization in independent low-throughput approaches.

Moreover, it is also challenging to estimate power and satisfactory sample size, especially for lncRNAs that show lower expression and power relative to mRNAs from the same experiment (extensively reviewed in [32]). As the power and sample size from experiments directly influence the conclusions from our analysis protocol, users are alerted that an analysis pipeline is no substitute for good quality data and functional validation experiments. We recommend that users should carefully plan their experiments e.g. using approaches such as paired-sample RNA-Seq and multifactor experimental design, which significantly enhances the statistical power (e.g. use the RNASeqPowerCalculator [32] to estimate the sample size).

Conservation analysis is widely used for functional characterization of newly discovered lncRNAs. Unlike protein-coding genes and miRNAs, lncRNAs are highly variable and less conserved among species, thus hampering translational approaches [12, 29, 34–36]. For instance, lncRNA functional conservation can be obtained across large evolutionary distances even with no sequence and structure similarity [12]. This multifaceted levels of conservation was demonstrated for the well-known lncRNAs Malat1 and Hotair [13]. Therefore, lncRNA characterization should focus on sequence, structure and function analysis [12, 13]. As most lncRNAs are functional uncharacterized, we integrated approaches such as interaction prediction, which have per se limitations regarding reliability and specificity as they typically tend to over-predictions. Moreover, prediction tools require significant amount of computational power especially for lncRNA interaction analysis with whole transcriptomes. For this, we used miRanda after advantageously splitting lncRNAs into snippets of 50 nt to reduce the computational burden of lncRNA–mRNA interaction and secondary structure calculation. However, we also tried tools such as IntaRNA, but the computational burden was too high for the whole transcriptome. Similarly, studies from Antonov *et al.* [31] compared different lncRNA–mRNA prediction tools, but limited the analysis on 17 interactions between transcripts shorter than 5000 nt due to the computation burden and limited computation power. We know that our strategy raises potential problems as the secondary structure of lncRNAs is instrumental for the function, whereas prediction tools such as miRanda are based on seed base-pairing and take secondary structure, hence only quite locally into account. However, several functional lncRNAs lack sequence and structural conservation [12, 31]. In this context, our aim was to allow prediction of lncRNA interactions with whole transcriptomes, but this strategy needs further validation steps as included in our pipeline. For instance, mapping of the predicted interactions to the RNA-Seq data indicates strong candidate interactions for further analysis as they show an expression in the RNA-Seq data. Moreover, further validation of the mRNA interaction candidates with general lncRNA–mRNA prediction tools such as IntaRNA will strengthen the interaction prediction accuracy (several prediction tools can be found in [31]). As an alternative to the whole transcriptome approach with miRanda, users can calculate lncRNA–mRNA interactions with a defined set of mRNA transcripts (e.g. well-known mRNAs for cardiac disease) using IntaRNA.

Taken together, we combine several analyses and filtering steps (sequence, structure, functional interaction context) to comprehensively investigate the regulatory role of the identified lncRNAs. In our example dataset we used stringent thresholds e.g. for e-value [14] and prediction power [16] as recommended for best quality prediction accuracy from these programs. Nevertheless, analysis can be performed also using different cut-offs (high/low stringent). This different analysis steps and filter cut-offs strengthen the robustness and reduce false positive, thus allowing to select most reliable candidates for further experimental testing. We applied this analysis strategy as useful software and support in testing RNA functionalities for the hypertrophy associated lncRNA Chast [6] and human hypoxia-sensitive lncRNAs LINC00323–003 and MIR503HG [29].

## Conclusion

In conclusion, we present a fast and efficient all-in-one method for annotation and functional analysis of lncRNAs. This is the first time that a method protocol follows such a comprehensive analysis strategy, combining sequence-structure conservation analysis and promoter analysis with the functional interaction context. Nevertheless, lncRNA analysis is challenging and the reader is alerted that a good analysis pipeline is no substitute for more good quality data and more functional validation experiments. In-depth studies are always required to fully understand and describe complex lncRNA functions and effects. However, we believe that our method is a powerful help in detailed understanding of the complex regulatory effects of lncRNAs within the genome. This in turn could open new windows for effective clinical strategies for better management of diseases. Our protocol works for uncharacterized as well as already experimentally characterized lncRNAs. Thus, our functional analysis strategy will help to better elucidate the complex regulatory effect of lncRNAs and is a powerful method protocol for daily scientific work.

## Data availability

The method protocol and source code are available in the supplementary material. Moreover, the protocol is implemented as an automatic and user-interactive graphical workflow in the NGS data analysis pipeline GensearchNGS available as a free academic trial under the link <http://www.phenosystems.com>.

### Key Points

- lncRNAs play an important biological role; however, the functional role is often unclear as experimental characterization is challenging and bioinformatic methods focus on specific diseases and mainly provide general information.
- We developed a powerful all-in-one method protocol for detailed understanding of the complex regulatory effects of lncRNAs combining sequence-structure conservation analysis and promoter analysis with the functional interaction context.
- The protocol allows analysis of experimental datasets including already experimental characterized as well as uncharacterized lncRNAs.
- It considers the tissue and biological context, and works for several disease types.

## Supplementary Data

Supplementary data are available online at <https://academic.oup.com/bib>.

## Acknowledgements

Linguistic proofreading by Aditi Joshi and Ulrike Rapp-Galmiche are gratefully acknowledged.

## Funding

Federal Ministry of Education and Research (BMBF grant FKZ 031L0129B and Era-Net grant 01KT1801 to M.K. and T.D.); European Research Council (ERC) grant Longheart to T.T.

## References

- Kung JTY, Colognori D, Lee JT. Long noncoding RNAs: past, present, and future. *Genetics* 2013;**193**:651–69.
- Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet* 2009;**10**:155–9.
- Kunz M, Wolf B, Schulze H, et al. Non-coding RNAs in lung cancer: contribution of bioinformatics analysis to the development of non-invasive diagnostic tools. *Genes (Basel)* 2016;**8**(1):8.
- Nagano T, Fraser P. No-nonsense functions for long noncoding RNAs. *Cell* 2011;**145**:178–81.
- Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. *Cell* 2009;**136**:629–41.
- Viereck J, Kumarswamy R, Foinquinos A, et al. Long noncoding RNA Chast promotes cardiac remodeling. *Sci Transl Med* 2016;**8**:326ra322.
- Volders P-J, Helsens K, Wang X, et al. LNCipedia: a database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Res* 2013;**41**:D246–51.
- Fang S, Zhang L, Guo J, et al. NONCODEV5: a comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Res* 2018;**46**:D308–14.
- Chakraborty S, Deb A, Maji RK, et al. LncRBase: an enriched resource for lncRNA information. *PLoS One* 2014;**9**: e108010.
- Li J-H, Liu S, Zhou H, et al. starBase v2.0: decoding miRNA-ncRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res* 2014;**42**:D92–7.
- Park C, Yu N, Choi I, et al. lncRNAtor: a comprehensive resource for functional investigation of long non-coding RNAs. *Bioinformatics* 2014;**30**:2480–5.
- Ulitsky I. Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nat Rev Genet* 2016;**17**:601.
- Diederichs S. The four dimensions of noncoding RNA conservation. *Trends Genet* 2014;**30**:121–3.
- Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool. *J Mol Biol* 1990;**215**:403–10.
- Betel D, Wilson M, Gabow A, et al. The [microRNA.org](http://microRNA.org) resource: targets and expression. *Nucleic Acids Res* 2008;**36**:D149–53.
- Agostini F, Zanzoni A, Klus P, et al. catRAPID omics: a web server for large-scale prediction of protein-RNA interactions. *Bioinformatics* 2013;**29**:2928–30.
- Smith C, Heyne S, Richter AS, et al. Freiburg RNA tools: a web server integrating INTARNA, EXPARNA and LOCARNA. *Nucleic Acids Res* 2010;**38**:W373–7.
- Messeguer X, Escudero R, Farre D, et al. PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics* 2002;**18**:333–4.
- Brooke JS, Cha JH, Eidels L. Latent transforming growth factor beta-binding protein-3 and fibulin-1C interact with the extracellular domain of the heparin-binding EGF-like growth factor precursor. *BMC Cell Biol* 2002;**3**:2.
- Koli K, Hyytiainen M, Ryyanen MJ, et al. Sequential deposition of latent TGF-beta binding proteins (LTBPs) during formation of the extracellular matrix in human lung fibroblasts. *Exp Cell Res* 2005;**310**:370–82.
- Louzao-Martinez L, Vink A, Harakalova M, et al. Characteristic adaptations of the extracellular matrix in dilated cardiomyopathy. *Int J Cardiol* 2016;**220**:634–46.
- Huang Z-P, Ding Y, Chen J, et al. Long non-coding RNAs link extracellular matrix gene expression to ischemic cardiomyopathy. *Cardiovasc Res* 2016;**112**:543–54.
- Hermans-Beijnsberger S, van Bilsen M, Schroen B. Long non-coding RNAs in the failing heart and vasculature. *Noncoding RNA Res* 2018;**3**:118–30.
- Kuo CC, Hanzelmann S, Senturk Cetin N, et al. Detection of RNA-DNA binding sites in long noncoding RNAs. *Nucleic Acids Res* 2019;**47**:e32.
- Kunz M, Gottlich C, Walles T, et al. MicroRNA-21 versus microRNA-34: lung cancer promoting and inhibitory microRNAs analysed in silico and in vitro and their clinical impact. *Tumour Biol* 2017;**39**:1010428317706430.
- Kunz M, Pittroff A, Dandekar T. Systems biology analysis to understand regulatory miRNA networks in lung cancer. *Methods Mol Biol* 2018;**1819**:235–47.
- Temme S, Friebe D, Schmidt T, et al. Genetic profiling and surface proteome analysis of human atrial stromal cells and rat ventricular epicardium-derived cells reveals novel insights into their cardiogenic potential. *Stem Cell Res* 2017;**25**:183–90.
- Prokosch HU, Acker T, Bernarding J, et al. MIRACUM: medical informatics in research and care in university medicine. *Methods Inf Med* 2018;**57**:e82–91.
- Fiedler J, Breckwoldt K, Remmele CW, et al. Development of long noncoding RNA-based strategies to modulate tissue vascularization. *J Am Coll Cardiol* 2015;**66**:2005–15.
- Jenjaroenpun P, Wongsurawat T, Sutheeworapong S, et al. R-loopDB: a database for R-loop forming sequences (RLFS) and R-loops. *Nucleic Acids Res* 2017;**45**:D119–27.
- Antonov IV, Mazurov E, Borodovsky M, et al. Prediction of lncRNAs and their interactions with nucleic acids: benchmarking bioinformatics tools. *Brief Bioinform* 2019;**20**(2):551–64.
- Ching T, Huang S, Garmire LX. Power analysis and sample size estimation for RNA-Seq differential expression. *RNA* 2014;**20**:1684–96.
- Lai Y. A statistical method for the conservative adjustment of false discovery rate (q-value). *BMC Bioinformatics* 2017;**18**:69.
- Chen J, Shishkin AA, Zhu X, et al. Evolutionary analysis across mammals reveals distinct classes of long non-coding RNAs. *Genome Biol* 2016;**17**:19.
- Chodroff RA, Goodstadt L, Sirey TM, et al. Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes. *Genome Biol* 2010;**11**:R72.
- Rivas E, Clements J, Eddy SR. A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat Methods* 2017;**14**:45–8.