

SPONGEdb: a pan-cancer resource for competing endogenous RNA interactions

Markus Hoffmann^{1,†}, Elisabeth Pachi^{1,†}, Michael Hartung^{1,†}, Veronika Stiegler^{1,†}, Jan Baumbach¹, Marcel H. Schulz^{2,3,4} and Markus List^{1,*}

¹Chair of Experimental Bioinformatics, TUM School of Life Sciences, Technical University of Munich, 85354 Freising, Germany, ²Institute for Cardiovascular Regeneration, Goethe University, 60596 Frankfurt am Main, Germany, ³German Center for Cardiovascular Research, Partner site Rhein-Main, 60590 Frankfurt am Main, Germany and ⁴Cardio-Pulmonary Institute, Goethe University Hospital, 60596 Frankfurt am Main, Germany

Received August 11, 2020; Revised November 12, 2020; Editorial Decision December 03, 2020; Accepted December 04, 2020

ABSTRACT

microRNAs (miRNAs) are post-transcriptional regulators involved in many biological processes and human diseases, including cancer. The majority of transcripts compete over a limited pool of miRNAs, giving rise to a complex network of competing endogenous RNA (ceRNA) interactions. Currently, gene-regulatory networks focus mostly on transcription factor-mediated regulation, and dedicated efforts for charting ceRNA regulatory networks are scarce. Recently, it became possible to infer ceRNA interactions genome-wide from matched gene and miRNA expression data. Here, we inferred ceRNA regulatory networks for 22 cancer types and a pan-cancer ceRNA network based on data from The Cancer Genome Atlas. To make these networks accessible to the biomedical community, we present SPONGEdb, a database offering a user-friendly web interface to browse and visualize ceRNA interactions and an application programming interface accessible by accompanying R and Python packages. SPONGEdb allows researchers to identify potent ceRNA regulators via network centrality measures and to assess their potential as cancer biomarkers through survival, cancer hallmark and gene set enrichment analysis. In summary, SPONGEdb is a feature-rich web resource supporting the community in studying ceRNA regulation within and across cancer types.

INTRODUCTION

microRNAs (miRNAs) are important non-coding, post-transcriptional regulators that are involved in many biological processes and human diseases (1). miRNAs regulate their target RNA transcripts by either degrading them or

by preventing their translation (2), thus acting as rheostats that regulate gene expression and maintain the functional balance of gene networks (3). The competing endogenous RNA (ceRNA) hypothesis suggests that RNA transcripts sharing binding sites for the same miRNA are in competition (4). Hence, miRNA-ceRNA interactions follow a many-to-many relationship where one miRNA can affect multiple ceRNA targets, and one ceRNA can contain multiple binding sites for various miRNAs (5), leading to complex cross-talk.

Since failures of these complex regulatory systems may lead to cancer (6), it is crucial to infer and compare gene-regulatory networks across various cancer types. However, previous network inference efforts focused mostly on transcription factor regulation (7–9), neglecting the influence of miRNA regulation and ceRNA competition. To close this gap, several efforts have been directed at developing methods for ceRNA network inference (10), which were applied to individual cancer types (11) but also across cancer types (12).

A number of resources have been proposed for studying ceRNAs, including miRTissue_{ce} (13), LnCeVar (14), Pan-ceRNADB (15), miRTarBase (16), ceRDB (17), lnCeDB (18), miRSponge (19), LncACTdb (20), miRcode (21), and starBase v2.0 (22) (see Supplementary Section S7 and Tables S1–3 for a systematic comparison). Only a few of these databases are dedicated to ceRNA networks. Pan-ceRNADB covers 20 cancer types (15) and considers messenger RNAs (mRNAs) in addition to lncRNAs as potential ceRNAs. In contrast, the LnCeVar database focuses on lncRNAs with experimental evidence for ceRNA interactions and adds additional analyses on the effect of single nucleotide and copy number variants on ceRNA activity (14). Existing resources currently do not afford genome-wide coverage, i.e. are limited to lncRNAs, or employ simplistic methods, which lead to a large number of false positives. For instance, only considering if two genes are en-

*To whom correspondence should be addressed. Markus List. Email: markus.list@wzw.tum.de

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

riched for shared miRNA regulation in a hypergeometric test neglects that both genes and the miRNA have to be co-expressed. Examining the correlation of gene expression with a shared miRNA does not quantify the effect on gene-gene correlation.

This highlights the need for a comprehensive and easily accessible resource of pan-cancer ceRNA networks that leverage state-of-the-art methods that account for the triangle relationship of ceRNA–miRNA–ceRNA triplets. Since methods leveraging conditional mutual information such as CUPID (23) or JAMI (24) do not afford genome-wide coverage due to computational costs, we have previously developed the method SPONGE, which overcomes limitations of sensitivity correlation (25) and facilitates fast genome-wide ceRNA network inference (26).

Here, we present SPONGEdb, a database providing access to SPONGE-inferred genome-wide ceRNA networks of 22 cancer types in TCGA (27) and a pan-cancer ceRNA network considering all available data (Figure 1). In contrast to most other resources, SPONGEdb results account for the contribution of multiple miRNAs and address previously neglected confounding factors. Specifically, we use the SPONGE R package (26) to infer empirical P -values that limit the false positive rate and avoid spurious correlations. SPONGEdb offers additional insights into the results through providing user-friendly features for accessibility not found in other databases: (i) SPONGEdb reported ceRNA interactions are easily accessible via a well-documented application programming interface (API); (ii) SPONGEdb allows investigating the effect of individual or multiple ceRNAs on a global ceRNA network, highlighting their relevance in a cancer background; a user-friendly web interface allows for searching and filtering ceRNA interactions and to visualize them as an interactive network enriched with additional information about patient survival as well as gene expression or functional annotations from WikiPathways (28), GeneCards (29), Quick GO (30), and Cancer Hallmarks (31). Functional enrichment analysis can be performed via g:Profiler (32). In addition to the web interface, we provide R and Python packages to allow third-party developers, data scientists, and biomedical researchers to carry out in-depth analyses. Here, we give an overview of the features of SPONGEdb and present a series of use cases demonstrating how this unique resource can be used for explorative analysis and hypothesis generation when studying ceRNA competition in cancer. Furthermore, we present results of SPONGE with experimentally validated data retrieved from Tay *et al.* (33) and miR-Sponge (19) using the features of SPONGEdb.

MATERIALS AND METHODS

Data processing

Using the SPONGE (26) bioconductor package, we identified significant ceRNA interactions for 10 019 samples for which paired gene, and miRNA expression data were available in TCGA (27). Data were processed by the TOIL project (35) and downloaded from the TOIL data hub in the Xena browser (36). Expression values were log2 transformed, and genes and miRNAs were discarded if not expressed in 80% or more of the samples or when

variance was <0.5 . To cover both non-coding and coding miRNA interactions, we used sequence-based predictions of several methods, namely TargetScan 7.1 (37) and miRcode v.11 (21). We further consider experimental evidence reported by miRTarBase 7.0 (38,39) and DIANA-LncBase v2 (40,41). SPONGE infers a ceRNA interaction network from the paired gene and miRNA expression data. SPONGE was applied to 22 cancer types in TCGA with >100 samples to obtain robust results for the correlations as well as for the aggregated pan-cancer dataset using default settings. For more details, we refer to the original method publication (26). Survival information from TCGA was used to produce Kaplan-Meier plots, where patients were split into two groups based on the median expression level. We used a log-rank test for assessing if the two survival curves differ significantly.

Implementation

Centrality measures. Centrality measures provide important information about the organization of complex systems in network analysis (42,43). Del Rio *et al.* showed that a combination of at least two centrality measures achieve reliable performance in biomarker detection (44). Hence, we provide degree, betweenness and eigenvector centrality (45) in SPONGEdb.

Degree centrality. The degree centrality corresponds to the number of edges connected to a node and highlights hub nodes (42), which are known to be important in biological systems (46–48).

Betweenness centrality. Here, we consider bottlenecks rather than hubs of a network to be important, i.e. betweenness centrality $c_i^{BC}(g)$ of node i in a graph g is high if i falls on a large fraction of shortest paths between other nodes j and k (49). Formally,

$$c_i^{BC}(g) = \sum_{(j,k), j \neq i, j \neq k} \frac{v_g(i : j, k)}{v_g(j, k)} \quad (1)$$

with $v_g(i : j, k)$, the number of shortest paths between j and k that visit i and $v_g(j, k)$ the number of all shortest paths between j and k . Raman *et al.* demonstrated that betweenness centrality correlate with lethality of organisms in protein networks (50).

Eigenvector centrality. Here, the importance of a node i is related to the importance of neighbors j connected by edge a_{ij} , i.e. eigenvector centrality c_i^{EC} is defined as:

$$c_i^{EC} = \frac{1}{\lambda} \sum_{j=1}^n a_{ij} c_j^{EC} \quad (2)$$

where a_{ij} is 1 if i and j are connected and else 0. λ corresponds to the eigenvalue which is unique under the requirement that all values of the eigenvector are positive. Negre *et al.* showed the importance of eigenvector centrality for biological pathways by creating a protein allosteric pathway network (51).

The SPONGEdb database contains expression and survival data from TCGA, gene annotations from EN-

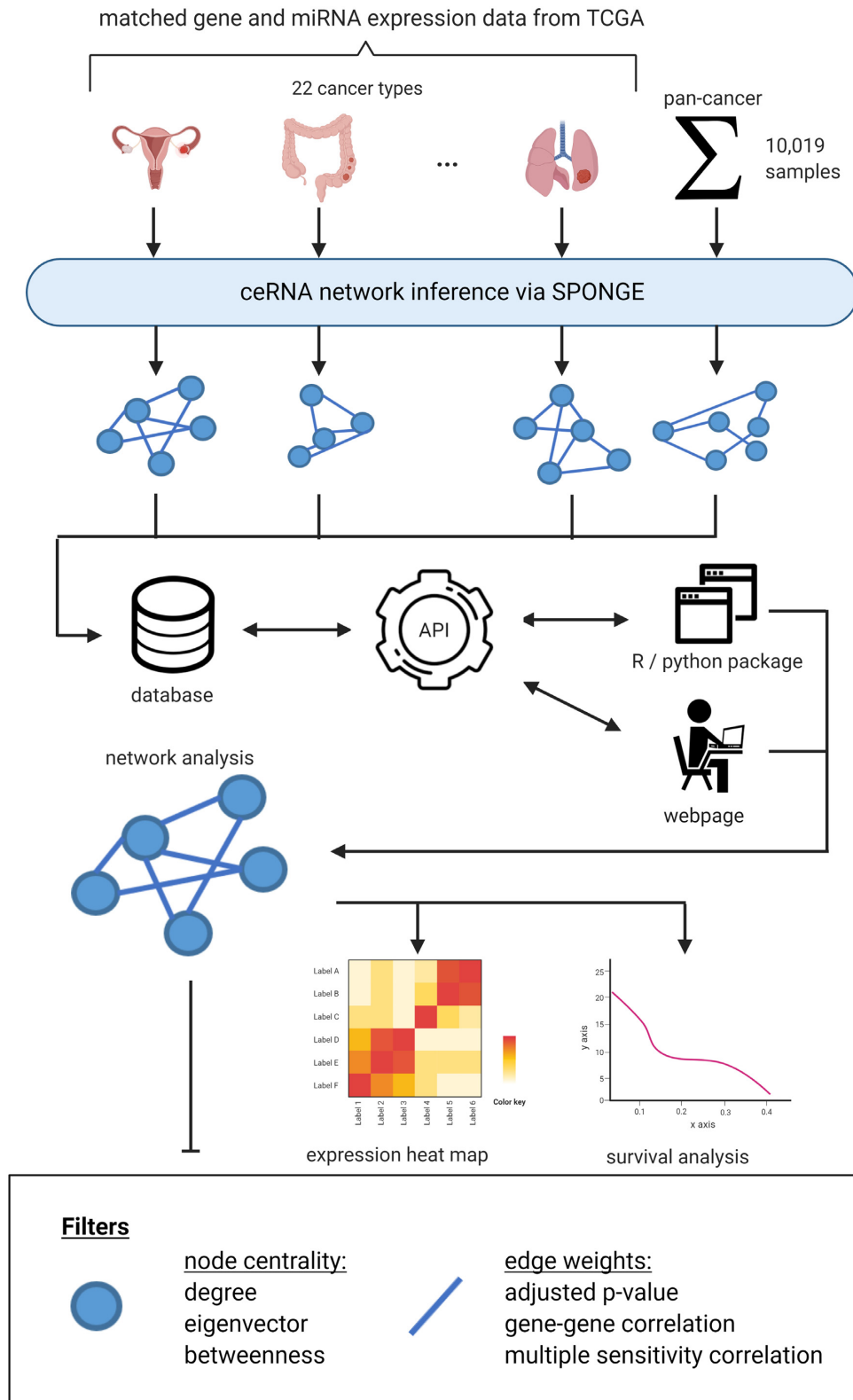


Figure 1. SPONGEdb overview: ceRNA networks for 22 cancer types as well as a pan-cancer ceRNA network for all 10 019 samples in TCGA have been inferred with SPONGE (26) and can be queried efficiently via a RESTful API through any programming language. We provide R and Python packages to simplify scripted access and offer a user-friendly and feature-rich web interface offering additional insights into ceRNA from external resources (ENCODE (34) for additional gene information, miRBase for additional miRNA information, Quick GO for correlating Gene Ontologies and the WikiPathways keys for each gene). The web interface offers visualization of the ceRNA networks, expression heat maps and survival analysis via Kaplan–Meier plots. Various filters for node centrality and edge weights allow refining the ceRNA network to focus on cancer-type-specific essential ceRNA regulators and their strongest or most significant interactions.

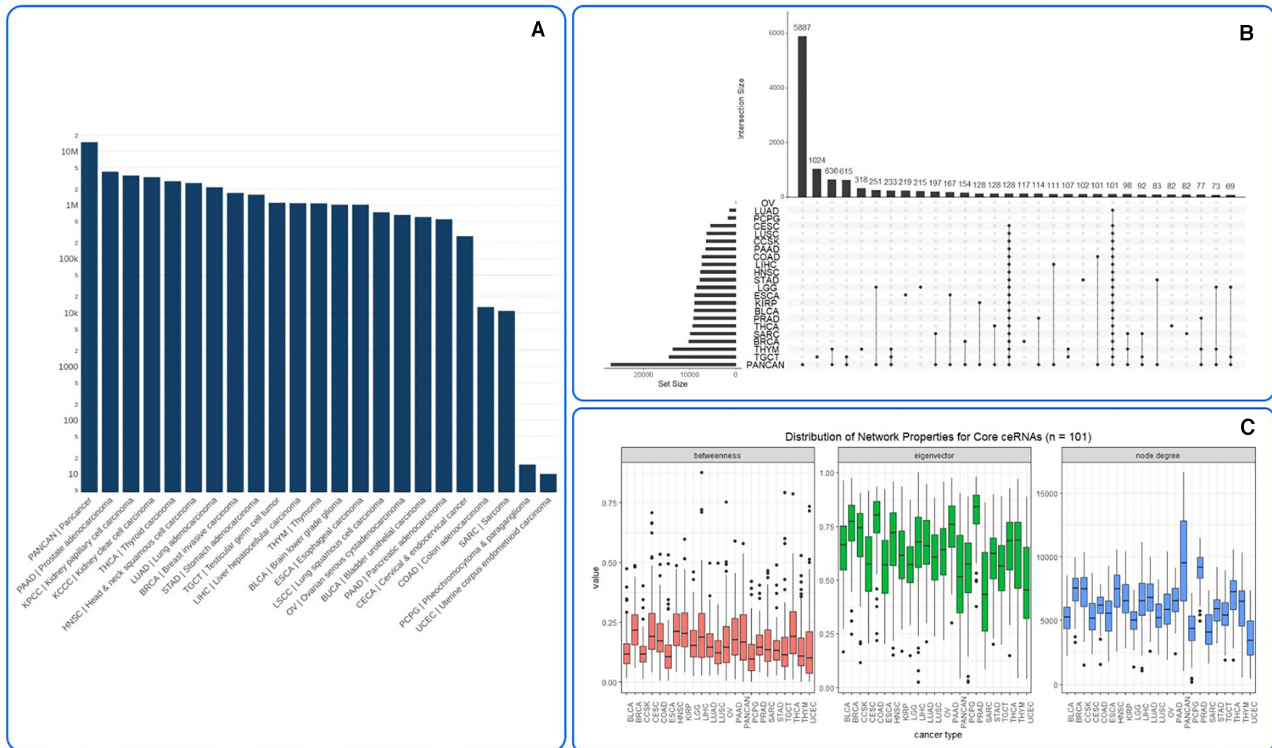


Figure 2. (A) Number of significant ceRNA interactions for each cancer type on a logarithmic scale. (B) Many ceRNAs are specific for a certain cancer type. Only two subsets of core ceRNA interactions are active in almost all cancer types (subsets with 101 and 128 ceRNAs, respectively). (C) Distribution of network properties for core ceRNAs for the subset of 101 ceRNA interactions, highlighting that core ceRNAs have only moderate regulatory strength.

CODE (34) and miRNA annotations from miRBase (52). For more details about the implementation of SPONGEdb, see Supplementary Sections S1–4 and Figures S1 and 2.

RESULTS AND DISCUSSION

SPONGEdb allows users to query ceRNA interactions for 22 cancer types, as well as a pan-cancer ceRNA network via an API for which we implement three use cases: (i) a user-friendly web interface, (ii) an R package, and (iii) a Python package. The general architecture and work-flow of SPONGEdb are shown in Figure 1. In the following, we present three application cases that illustrate the broad utility of SPONGEdb in ceRNA research.

Pan-cancer analysis of ceRNA activity

SPONGEdb reports many significant interactions across 20 of the 22 cancer types as well as a unique pan-cancer analysis revealing key ceRNAs that play a role across cancer types (Figure 2A). Many of the candidate ceRNAs implicated here have not been described in the literature and can thus serve to prioritize candidates for experimental validation. Only a fraction of ceRNAs appear to act across all cancer types and, conversely, many ceRNAs act in a cancer-specific fashion (Figure 2B and C). As expected, we observe a positive trend between the number of tumor samples (e.g. breast cancer has 1063 samples, while thymoma has only 119 samples) and the number of significant interactions (Figure 4), which can be explained by an increase

in statistical power. Fitting this hypothesis, the pan-cancer analysis comprising all cohorts provides the largest number of significant interactions. However, for some cancer types, we find only very few significant interactions, even if the sample numbers would suggest otherwise. Likewise, we observe a difference in the number of significant interactions for cancer types with similar cohort sizes.

Ovarian serous cystadenocarcinoma and uterine corpus endometrial carcinoma appear as outliers with <1% of significant interactions (Figure 2), suggesting that in these cancer types, ceRNA regulation may either be less pronounced than in other cancer types or confounded by heterogeneity, i.e. introduced by the tumor microenvironment or by subtype-specific differences. To further investigate this, we performed SPONGE analysis on an independent Australian cohort of ovarian cancer patients (<https://dcc.icgc.org/releases/current/Projects/OV-AU>) where we also found <1% of significant interactions (see Supplementary Section S8 and Table S4), confirming the results on TCGA data.

There are other factors to consider that may limit the sensitivity of SPONGE, e.g. the high number of significant ceRNA interactions in testicular germ cell tumors could be explained by a relatively high complexity of the transcriptome in this tissue (53). Moreover, we consider only gene–miRNA interactions with a negative regression coefficient and may thus miss cases where translational repression rather than degradation of the target transcript occurs. The presence of possible biological confounders highlights that

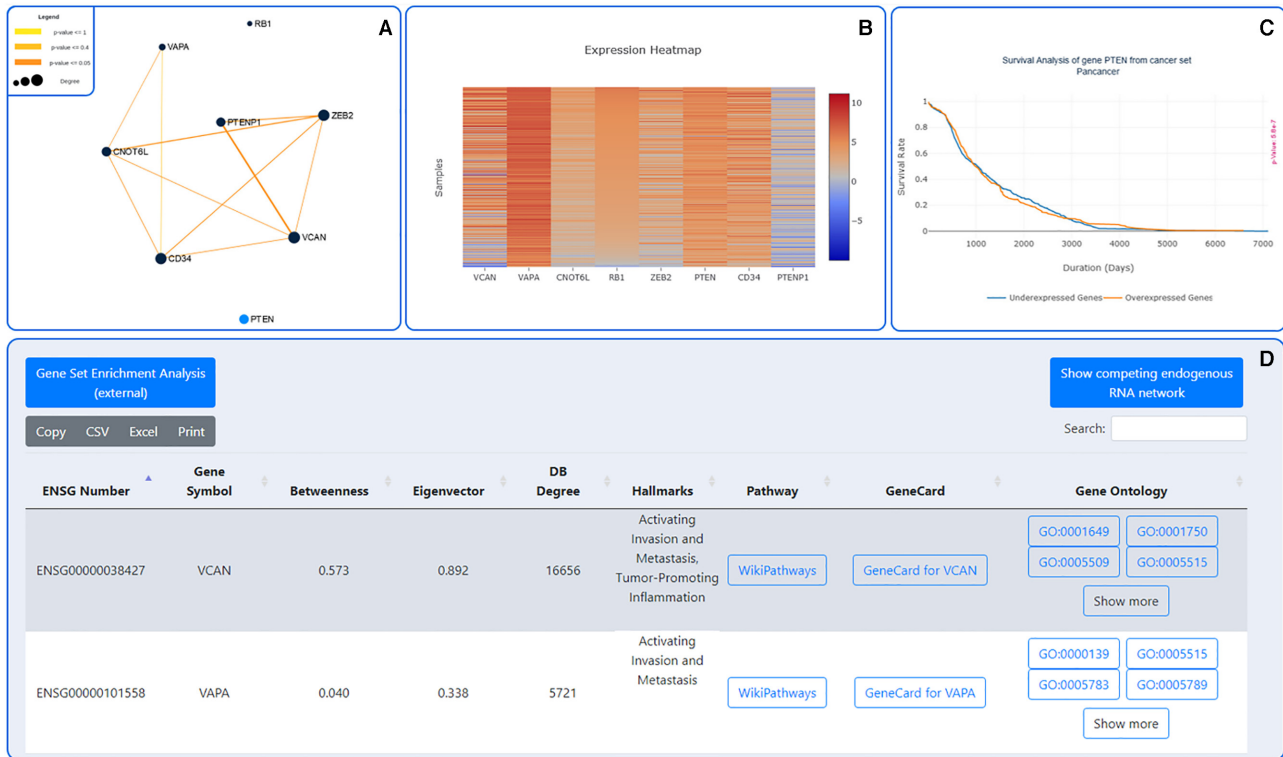


Figure 3. The reproduced figures of PTEN, PTENP1, VCAN, CD34, CNOT6L, VAPA, ZEB2 and RB1. The SPONGE algorithm found most of the described interactions in the Tay *et al.* paper and found additional interactions between the genes, which need to be experimentally validated. Additionally, the website provides following information: (A) Network visualizing the interactions between the genes. The edge thickness and color, as well as the node size, corresponds to the *P*-value/node degree. *P*-value ≤ 1 : yellow, *P*-value ≤ 0.4 : light orange, *P*-value ≤ 0.05 : dark orange. (B) Expression heatmap. (C) Survival Analysis of PTEN. *P*-value = $5.8e^{-7}$. (D) External links of the shown genes for one-click further investigation to important web resources. We included WikiPathways, GeneCards, Quick GO and g:Profiler.

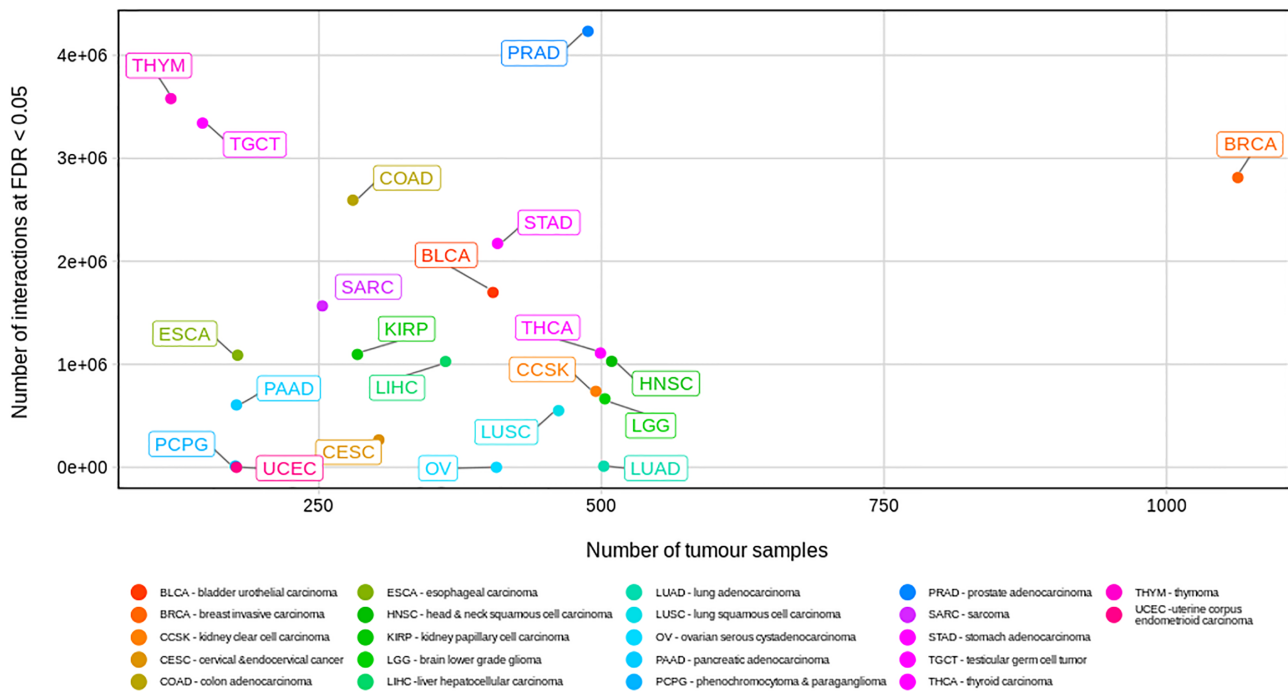


Figure 4. Pan-cancer analysis of the number of significant ceRNA interactions at FDR < 0.05 compared to the number of tumor samples. This figure was created with the SPONGEdb R package, see Supplementary Section S6 for details.

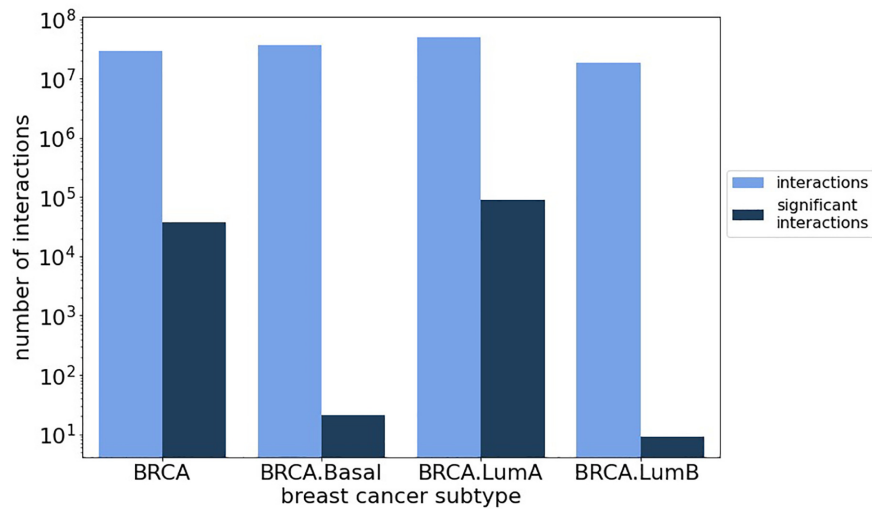


Figure 5. Comparison of results for BRCA and its subtypes luminal A, luminal B and basal after subsampling to 132 samples.

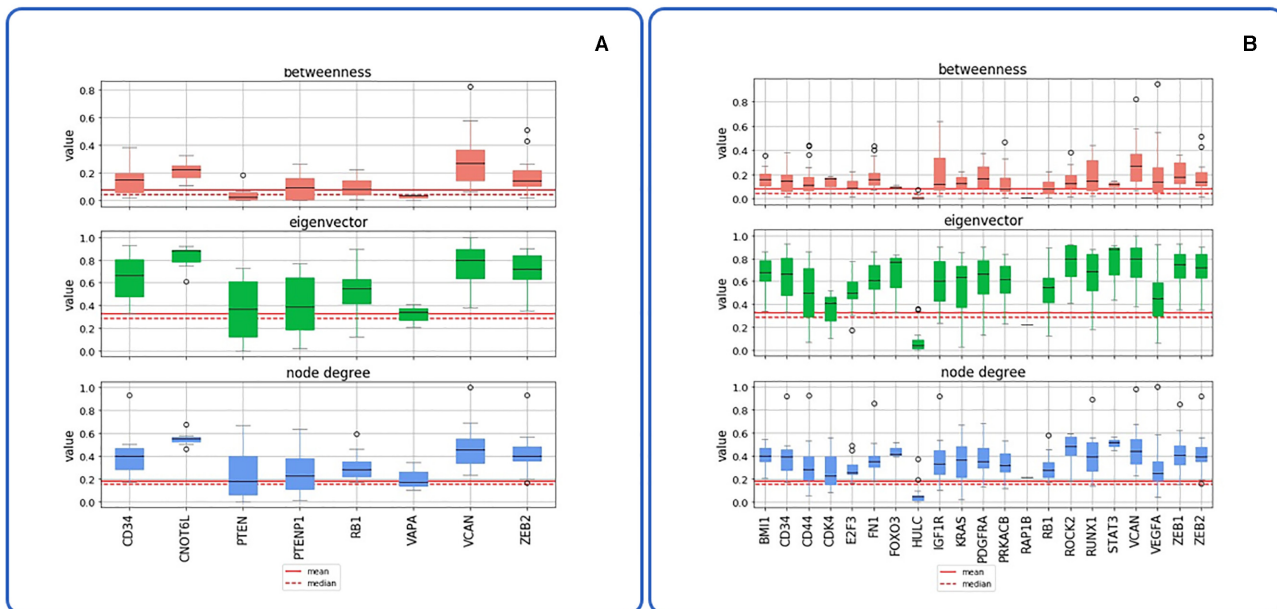


Figure 6. Different network centrality measures summarized across cancer types for experimentally validated ceRNAs from (A) Tay *et al.* (33) and (B) miRSponge (19). This plot was produced using the SPONGEdb python package (see Supplementary Section S12). Note that the mean and median were computed across all cancer types.

comparisons between cancer types should focus on commonalities rather than differences. We further note that the size of the resulting ceRNA networks in SPONGEdb can be increased by adjusting the FDR cutoff, which defaults to 0.05 to provide suitable candidates for experimental validation.

Cancer subtype analysis of BRCA

To further investigate the influence of cancer subtype-related heterogeneity, we performed independent analyses for the major breast cancer subtypes (luminal A, luminal B, HER2, and basal) from TCGA data. We chose breast cancer since its subtypes are clinically well defined, and since

this dataset has a sufficiently high sample number to consider subtypes. In Supplementary Section S9, (Supplementary Table S5 and Figures S3 and 4), we show that, despite the reduced heterogeneity, the number of significant interactions is dramatically reduced as a consequence of the lower sample number.

Next, we subsampled luminal A, luminal B, and basal subtype data as well as the entire BRCA cohort to 132 samples (owing to basal, the smallest of the subtype cohorts considered here). This analysis (Figure 5; Supplementary Section S9, Table S6, Figure S5 and 6) shows that the number of overall and significant interactions differs strongly between subtypes. Luminal A, which is the most frequent and least aggressive subtype, shows the largest number of

overall and significant interactions compared to the other subtypes as well as compared to a random sample drawn across all subtypes. This shows that subtype-specific analysis can reduce the influence of sample heterogeneity. Only a few significant interactions are reported in the other subtypes, which suggests that a subtype separation alone does not address all sources of heterogeneity. For example, we could recently show that the luminal and basal breast cancer subtypes in TCGA are confounded by tumor-infiltrating leukocytes (54).

Experimentally validated ceRNAs in a pan-cancer context

SPONGE can also be used to investigate previously reported or putative interactions of experimentally validated ceRNAs. For example, the pseudogene PTENP1 has been reported to regulate levels of the tumor suppressor gene PTEN by competing for shared miRNAs. Low amounts of PTEN are associated with a higher risk of developing cancer. Furthermore, Tay *et al.* reported a dense ceRNA network involving PTEN and PTENP1 together with VCAN, CNOT6L, CD34, VAPA, ZEB2 and RB1. To test if the reported ceRNA network can be reproduced in SPONGEdb, we selected the above query genes in the pan-cancer dataset. The resulting network (Figure 3 A) corroborates previously reported interactions by Tay *et al.* and highlights further putative ceRNA interactions between the selected genes, which need to be experimentally validated.

The web interface offers additional insights into the selected network, including a heatmap of expression values (Figure 3B), survival data shown as a Kaplan–Meier plot (Figure 3C), key network centrality measures including degree, betweenness and eigenvector centrality, as well as functional annotation of cancer hallmarks and links to external information about associated pathways and gene ontology terms, including gene set enrichment analysis (Figure 3D). All information shown in the web interface can conveniently be exported in tabular format or as an image.

Experimentally validated ceRNAs show better sponging capacity than random ceRNAs

Next, we investigated if experimentally validated ceRNAs outperform random sets of ceRNAs in terms of network centrality measures. For this analysis, we selected the eight ceRNAs reported by Tay *et al.* as well as another 21 experimentally validated ceRNAs reported by miRSponge and compared their centrality measures to an average of 1000 times randomly chosen gene set of equal size as well as to the median and mean of all genes available inside the network. Our results (Supplementary Section S10 and Figures S7–10) show that the betweenness, node degree and eigenvector centralities are considerably higher for these ceRNAs. We also investigated if these ceRNAs are only relevant for specific cancer types by systematically comparing their centrality measures. Our results (Supplementary Section S11, Figures S11 and 12) show that many experimentally validated ceRNAs are relevant in a larger number of cancer types. Figure 6 shows that all investigated ceRNAs but VAPA (Tay *et al.*) and HULC, RAP1B (miRSponge) have centrality measures larger than the mean of all ceRNAs re-

ported in SPONGEdb. This corroborates that network centrality measures reported by SPONGEdb capture meaningful ceRNA biology and that predicted ceRNAs are promising candidates for experimental validation.

CONCLUSION AND OUTLOOK

To date, TCGA is the most comprehensive resource of molecular profiling data in cancer research (27). The availability of paired gene and miRNA expression data offers a unique chance to comprehensively study ceRNA regulation in a network context. Here, we made SPONGE-inferred ceRNA networks for 22 cancer types available as an easy-to-access resource. To facilitate broad utility, SPONGEdb implements an API that allows biomedical researchers to query the results programmatically to embed our results in their own analyses and tools. Moreover, we offer an interactive web interface to browse, filter and visualize the results. In conclusion, SPONGEdb is an important resource for studying the role of ceRNAs in various cancer types and for prioritizing ceRNA candidates for experimental validation. For the future, we plan to extend SPONGEdb with more sophisticated network analysis features, e.g. for the detection of disease-relevant sub-networks via network enrichment (55). Although the paired gene and miRNA expression data are currently scarce, we envision that such data will become more prevalent, allowing us to extend SPONGEdb beyond the current application in cancer. Moreover, we expect that co-profiling of gene and miRNA expression in single cells will allow us to infer cell-type-specific ceRNA networks to further broaden the scope of SPONGEdb.

DATA AVAILABILITY

SPONGE website:

<http://sponge.biomedical-big-data.de>

API:

<http://sponge-api.biomedical-big-data.de>

R-package:

<https://github.com/biomedbigdata/SPONGE-web-R>

Python-package:

<https://github.com/biomedbigdata/spongeWebPy>

Static-file-server:

<http://sponge-files.biomedical-big-data.de>

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Cancer Online.

ACKNOWLEDGEMENTS

The results published here are in whole or part based upon data generated by the TCGA Research Network: <https://www.cancer.gov/tcga>. Figure 1 was created with BioRender.com.

FUNDING

This work was funded by the German Science Foundation (DFG) via the Collaborative Research Center

(SFB1371) [SFB1371 to M.L., J.B.]; This work was supported by the German Federal Ministry of Education and Research (BMBF) within the framework of the e:Med research and funding concept [01ZX1908A to M.L., J.B.]; Deutsches Zentrum für Herz-Kreislaufforschung [81Z0200101 to M.H.S.]. Funding for open access charge: Internal Funds.

Conflict of interest statement. None declared.

REFERENCES

- Kartha,R.V. and Subramanian,S. (2014) Competing endogenous RNAs (ceRNAs): new entrants to the intricacies of gene regulation. *Front. Genet.*, **5**, 8.
- Cancer Genome Atlas Research Network, Weinstein,J.N., Collisson,E.A., Mills,G.B., Shaw,K.R.M., Ozenberger,B.A., Ellrott,K., Shmulevich,I., Sander,C. and Stuart,J.M. (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.*, **45**, 1113–1120.
- Wang,Z., Jensen,M.A. and Zenklusen,J.C. (2016) A practical guide to The Cancer Genome Atlas (TCGA). *Methods Mol. Biol.*, **1418**, 111–141.
- Salmena,L., Polisenio,L., Tay,Y., Kats,L. and Pandolfi,P.P. (2011) A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell*, **146**, 353–358.
- Chiu,H.-S., Martínez,M.R., Komissarova,E.V., Llobet-Navas,D., Bansal,M., Paull,E.O., Silva,J., Yang,X., Sumazin,P. and Califano,A. (2018) The number of titrated microRNA species dictates ceRNA regulation. *Nucleic Acids Res.*, **46**, 4354–4369.
- Yang,C., Wu,D., Gao,L., Liu,X., Jin,Y., Wang,D., Wang,T. and Li,X. (2016) Competing endogenous RNA networks in human cancer: hypothesis, validation, and perspectives. *Oncotarget*, **7**, 13479–13490.
- Castro,D.M., de Veaux,N.R., Miraldi,E.R. and Bonneau,R. (2019) Multi-study inference of regulatory networks for more accurate models of gene regulation. *PLoS Comput. Biol.*, **15**, e1006591.
- Dewey,G.T. and Galas,D.J. (2013) In: *Gene Regulatory Networks*, Landes Bioscience, Austin, Texas.
- Sun,N. and Zhao,H. (2009) Reconstructing transcriptional regulatory networks through genomics data. *Stat. Methods Med. Res.*, **18**, 595–617.
- Le,T.D., Zhang,J., Liu,L. and Li,J. (2017) Computational methods for identifying miRNA sponge interactions. *Brief. Bioinform.*, **18**, 577–590.
- Sumazin,P., Yang,X., Chiu,H.-S., Chung,W.-J., Iyer,A., Llobet-Navas,D., Rajbhandari,P., Bansal,M., Guarnieri,P., Silva,J. et al. (2011) An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma. *Cell*, **147**, 370–381.
- Zhang,Y., Xu,Y., Feng,L., Li,F., Sun,Z., Wu,T., Shi,X., Li,J. and Li,X. (2016) Comprehensive characterization of lncRNA-mRNA related ceRNA network across 12 major cancers. *Oncotarget*, **7**, 64148–64167.
- Fiannaca,A., Paglia,L.L., Rosa,M.L., Rizzo,R. and Urso,A. (2020) miRTissue ce: extending miRTissue web service with the analysis of ceRNA-ceRNA interactions. *BMC Bioinformatics*, **21**, 199.
- Wang,P., Li,X., Gao,Y., Guo,Q., Ning,S., Zhang,Y., Shang,S., Wang,J., Wang,Y., Zhi,H. et al. (2020) LnCeVar: a comprehensive database of genomic variations that disturb ceRNA network regulation. *Nucleic Acids Res.*, **48**, D111–D117.
- Xu,J., Li,Y., Lu,J., Pan,T., Ding,N., Wang,Z., Shao,T., Zhang,J., Wang,L. and Li,X. (2015) The mRNA related ceRNA-ceRNA landscape and significance across 20 major cancer types. *Nucleic Acids Res.*, **43**, 8169–8182.
- Huang,H.-Y., Lin,Y.-C.-D., Li,J., Huang,K.-Y., Shrestha,S., Hong,H.-C., Tang,Y., Chen,Y.-G., Jin,C.-N., Yu,Y. et al. (2020) miRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. *Nucleic Acids Res.*, **48**, D148–D154.
- Sarver,A.L. and Subramanian,S. (2012) Competing endogenous RNA database. *Bioinformatics*, **8**, 731–733.
- Das,S., Ghosal,S., Sen,R. and Chakrabarti,J. (2014) InCeDB: database of human long noncoding RNA acting as competing endogenous RNA. *PLoS One*, **9**, e98965.
- Wang,P., Zhi,H., Zhang,Y., Liu,Y., Zhang,J., Gao,Y., Guo,M., Ning,S. and Li,X. (2015) miRSponge: a manually curated database for experimentally supported miRNA sponges and ceRNAs. *Database*, **2015**, doi:10.1093/database/bav098.
- Wang,P., Li,X., Gao,Y., Guo,Q., Wang,Y., Fang,Y., Ma,X., Zhi,H., Zhou,D., Shen,W. et al. (2019) LncACTdb 2.0: an updated database of experimentally supported ceRNA interactions curated from low- and high-throughput experiments. *Nucleic Acids Res.*, **47**, D121–D127.
- Jeggari,A., Marks,D.S. and Larsson,E. (2012) miRcode: a map of putative microRNA target sites in the long non-coding transcriptome. *Bioinformatics*, **28**, 2062–2063.
- Li,J.-H., Liu,S., Zhou,H., Qu,L.-H. and Yang,J.-H. (2013) starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.*, **42**, D92–D97.
- Chiu,H.-S., Llobet-Navas,D., Yang,X., Chung,W.-J., Ambesi-Impimbato,A., Iyer,A., Kim,H.R., Seviour,E.G., Luo,Z., Sehgal,V. et al. (2015) Cupid: simultaneous reconstruction of microRNA-target and ceRNA networks. *Genome Res.*, **25**, 257–267.
- Hornakova,A., List,M., Vreeken,J. and Schulz,M.H. (2018) JAMI: fast computation of conditional mutual information for ceRNA network analysis. *Bioinformatics*, **34**, 3050–3051.
- Paci,P., Colombo,T. and Farina,L. (2014) Computational analysis identifies a sponge interaction network between long non-coding RNAs and messenger RNAs in human breast cancer. *BMC Syst. Biol.*, **8**, 83.
- List,M., Dehghani Amirabad,A., Kostka,D. and Schulz,M.H. (2019) Large-scale inference of competing endogenous RNA networks with sparse partial correlation. *Bioinformatics*, **35**, i596–i604.
- Tomczak,K., Czerwińska,P. and Wiznerowicz,M. (2015) The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp. Oncol.*, **19**, A68–A77.
- Slenter,D.N., Kutmon,M., Hanspers,K., Riutta,A., Windsor,J., Nunes,N., MÃ©lius,J., Cirillo,E., Coort,S.L., Digles,D. et al. (2017) WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res.*, **46**, D661–D667.
- Stelzer,G., Rosen,N., Plaschkes,I., Zimmerman,S., Twik,M., Fishilevich,S., Stein,T.I., Nudel,R., Lieder,I., Mazor,Y. et al. (2016) The GeneCards Suite: from gene data mining to disease genome sequence analyses. *Curr. Protoc. Bioinform.*, **54**, 1.30.1–1.30.33.
- Binns,D., Dimmer,E., Huntley,R., Barrell,D., O'Donovan,C. and Apweiler,R. (2009) QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics*, **25**, 3045–3046.
- Zhang,D., Huo,D., Xie,H., Wu,L., Zhang,J., Liu,L., Jin,Q. and Chen,X. (2020) CHG: a systematically integrated database of cancer hallmark genes. *Front. Genet.*, **11**, 29.
- Reimand,J., Arak,T., Adler,P., Kolberg,L., Reisberg,S., Peterson,H. and Vilo,J. (2016) g:Profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res.*, **44**, W83–W89.
- Tay,Y., Rinn,J. and Pandolfi,P.P. (2014) The multilayered complexity of ceRNA crosstalk and competition. *Nature*, **505**, 344–352.
- ENCODE Project Consortium (2011) A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.*, **9**, e1001046.
- Vivian,J., Rao,A.A., Nothhaft,F.A., Ketchum,C., Armstrong,J., Novak,A., Pfeil,J., Narkizian,J., Deran,A.D., Musselman-Brown,A. et al. (2017) Toil enables reproducible, open source, big biomedical data analyses. *Nat. Biotechnol.*, **35**, 314–316.
- Goldman,M.J., Craft,B., Hastie,M., Repečka,K., McDade,F., Kamath,A., Banerjee,A., Luo,Y., Rogers,D., Brooks,A.N. et al. (2020) Visualizing and interpreting cancer genomics data via the Xena platform. *Nat. Biotechnol.*, **38**, 675–678.
- Agarwal,V., Bell,G.W., Nam,J.-W. and Bartel,D.P. (2015) Predicting effective microRNA target sites in mammalian mRNAs. *Elife*, **4**, e05005.
- Chou,C.-H., Shrestha,S., Yang,C.-D., Chang,N.-W., Lin,Y.-L., Liao,K.-W., Huang,W.-C., Sun,T.-H., Tu,S.-J., Lee,W.-H. et al. (2018) miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.*, **46**, D296–D302.

39. Chou, C.-H., Chang, N.-W., Shrestha, S., Hsu, S.-D., Lin, Y.-L., Lee, W.-H., Yang, C.-D., Hong, H.-C., Wei, T.-Y., Tu, S.-J. *et al.* (2016) miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. *Nucleic Acids Res.*, **44**, D239–D247.
40. Paraskevopoulou, M.D., Georgakilas, G., Kostoulas, N., Reczko, M., Maragkakis, M., Dalamagas, T.M. and Hatzigeorgiou, A.G. (2013) DIANA-LncBase: experimentally verified and computationally predicted microRNA targets on long non-coding RNAs. *Nucleic Acids Res.*, **41**, D239–D245.
41. Paraskevopoulou, M.D., Vlachos, I.S., Karagkouni, D., Georgakilas, G., Kanellos, I., Vergoulis, T., Zagganas, K., Tsanakas, P., Floros, E., Dalamagas, T. *et al.* (2016) DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts. *Nucleic Acids Res.*, **44**, D231–D238.
42. Rodrigues, F.A. (2019) Network centrality: an introduction. In: Macau, E.E.N. (ed) *A Mathematical Modeling Approach from Nonlinear Dynamics to Complex Systems*. Springer International Publishing, Cham, pp. 177–196.
43. Zheng, F., Wei, L., Zhao, L. and Ni, F. (2018) Pathway network analysis of complex diseases based on multiple biological networks. *Biomed. Res. Int.*, **2018**, 5670210.
44. del Rio, G., Koschützki, D. and Coello, G. (2009) How to identify essential genes from molecular networks? *BMC Syst. Biol.*, **3**, 102.
45. Bloch, F., Jackson, M.O. and Tebaldi, P. (2016) Centrality measures in networks. arXiv doi: <https://arxiv.org/abs/1608.05845>, 17 June 2017, preprint: not peer reviewed.
46. He, X. and Zhang, J. (2006) Why do hubs tend to be essential in protein networks? *PLoS Genet.*, **2**, e88.
47. Zhu, X., Gerstein, M. and Snyder, M. (2007) Getting connected: analysis and principles of biological networks. *Genes Dev.*, **21**, 1010–1024.
48. Lu, X., Jain, V.V., Finn, P.W. and Perkins, D.L. (2007) Hubs in biological interaction networks exhibit low changes in expression in experimental asthma. *Mol. Syst. Biol.*, **3**, 98.
49. Freeman, L.C. (1977) A set of measures of centrality based on betweenness. *Sociometry*, **40**, 35–41.
50. Raman, K., Damaraju, N. and Joshi, G.K. (2014) The organisational structure of protein networks: revisiting the centrality-lethality hypothesis. *Syst. Synth. Biol.*, **8**, 73–81.
51. Negre, C. F.A., Morzan, U.N., Hendrickson, H.P., Pal, R., Lisi, G.P., Loria, J.P., Rivalta, I., Ho, J. and Batista, V.S. (2018) Eigenvector centrality for characterization of protein allosteric pathways. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, E12201–E12208.
52. Griffiths-Jones, S. (2006) miRBase: the microRNA sequence database. *Methods Mol. Biol.*, **342**, 129–138.
53. Soumillon, M., Necsulea, A., Weier, M., Brawand, D., Zhang, X., Gu, H., Barthès, P., Kokkinaki, M., Nef, S., Gnirke, A. *et al.* (2013) Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep.*, **3**, 2179–2190.
54. Lazareva, O., Van Do, H., Canzar, S., Yuan, K., Baumbach, J., Blumenthal, D.B., Tieri, P., Kacprowski, T. and List, M. (2020) BiCoN: Network-constrained biclustering of patients and omics data. *Bioinformatics*, doi:10.1093/bioinformatics/btaa1076.
55. Batra, R., Alcaraz, N., Gitzhofer, K., Pauling, J., Ditzel, H.J., Hellmuth, M., Baumbach, J. and List, M. (2017) On the performance of de novo pathway enrichment. *NPJ. Syst. Biol. Appl.*, **3**, 1–8.