Olgun *et al. BMC Bioinformatics* (2021) 22:294
https://doi.org/10.1186/s12859-021-04112-9

BMC Bioinformatics

SOFTWARE Open Access

NoRCE: non-coding RNA sets cis enrichment tool

Gulden Olgun[1,3], Afshan Nabi[2] and Oznur Tastan[2,3*]

*Correspondence: otastan@sabanciuniv.edu
[2] Present Address: Faculty of Engineering and Natural Sciences, Sabanci University, 34956 Istanbul, Turkey
Full list of author information is available at the end of the article

Abstract

***Background*:** While some non-coding RNAs (ncRNAs) are assigned critical regulatory roles, most remain functionally uncharacterized. This presents a challenge whenever an interesting set of ncRNAs needs to be analyzed in a functional context. Transcripts located close-by on the genome are often regulated together. This genomic proximity on the sequence can hint at a functional association.

***Results*:** We present a tool, NoRCE, that performs *cis* enrichment analysis for a given set of ncRNAs. Enrichment is carried out using the functional annotations of the coding genes located proximal to the input ncRNAs. Other biologically relevant information such as topologically associating domain (TAD) boundaries, co-expression patterns, and miRNA target prediction information can be incorporated to conduct a richer enrichment analysis. To this end, NoRCE includes several relevant datasets as part of its data repository, including cell-line specific TAD boundaries, functional gene sets, and expression data for coding & ncRNAs specific to cancer. Additionally, the users can utilize custom data files in their investigation. Enrichment results can be retrieved in a tabular format or visualized in several different ways. NoRCE is currently available for the following species: human, mouse, rat, zebrafish, fruit fly, worm, and yeast.

***Conclusions*:** NoRCE is a platform-independent, user-friendly, comprehensive R package that can be used to gain insight into the functional importance of a list of ncRNAs of any type. The tool offers flexibility to conduct the users' preferred set of analyses by designing their own pipeline of analysis. NoRCE is available in Bioconductor and https://github.com/guldenolgun/NoRCE.

Keywords: Non-coding gene, Enrichment analysis, Multi-species R package, Co-expression analysis, TAD

Background

The advent of next-gen sequencing technologies and their application to transcriptomes have shown that the vast majority of the human genome is transcribed [1, 2] and the non-coding RNAs (ncRNAs) represent the largest class of transcripts in the human genome [3, 4]. NcRNAs are categorized into different groups based on length, location, or function: long non-coding RNAs (lncRNAs), microRNAs (miRNAs), small interfering RNAs (siRNAs), small nucleolar RNAs (snoRNAs), small nuclear RNAs (snRNAs), and Piwi-interacting RNAs (piRNAs).

© The Author(s), 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

NcRNAs have been implicated in a wide array of cellular processes [2, 5–7] and emerging evidence further reinforces that they have crucial functional importance for normal development and disease [8]. For example, lncRNAs, the largest class of ncRNAs, are reported to control nuclear architecture and transcription, modulate mRNA stability, translation, and post-translational modifications [7, 9]. Nevertheless, only a small fraction of ncRNAs have been functionally characterized today, and most ncRNAs' functions remain unknown. The lack of functional annotation of ncRNAs presents a challenge when an ncRNA set of interest is available and needs to be functionally investigated for further analysis.

Most of the available ncRNAs functional enrichment tools are limited to miRNAs. In the first step of these tools, they make a list of genes that are targeted by at least one of the miRNAs in the input set, which is followed by an enrichment analysis on this target gene set [10–12]. The target set is derived from experimentally validated interaction databases or produced by target prediction algorithms. Among them, Corna [10], miRTar [12], and Diana-miRPath v.3 [11] differ from varied features such as the source of the targets or the functional sets on which the analysis is conducted. Since the predicted target interactions might include high false positives and are not context-specific, some methods also take into account the changes in mRNA levels. MiRComb [13] conducts a miRNA-mRNA expression analysis followed by miRNA target prediction on the negatively correlated mRNA targets. miRFA [14] considers both the negatively and positively correlated using TCGA data. miTALOS [15, 16] additionally provides a tissue-specific filtering of the targets.

There is also a limited number of tools that offer functional annotation and enrichment analysis on lncRNA sets. Similar to miRNA methods, these methods first find a set of coding genes that are co-expressed genes with the given lncRNA or the lncRNAs in the collection and conduct analysis on these coding genes [17–19]. With regards to other ncRNAs, only a few studies provide analysis for ncRNAs other than lncRNA and miRNA. StarBase v2 first constructs a regulatory network based on experimentally identified RNA binding sites and their interactions; next, they perform functional enrichment on the interacting coding genes of the ncRNAs [20]. Starbase v2 offers analysis on miRNAs, lncRNAs, and the pseudogenes. CircFunBase [21] is not an enrichment tool but provides manually curated functions of circular RNAs that can be used for enrichment analysis.

The available tools are limited to the type of input ncRNA they support and do not take into account genomic neighborhood information. In this work, we present NoRCE (Non-coding RNA Sets Cis Enrichment Tool), which offers broad applicability and functionality for enrichment analysis of all types of ncRNAs sets using genomic proximity. NoRCE first finds nearby coding genes on the genome of the ncRNAs in the input set and uses the functional annotations of this coding gene set to perform functional enrichment on the ncRNA set. The motivation of using coding genes for annotation is based on the evidence presented earlier that genes nearby can be linked functionally. Thevenin et al. [22] show that functionally related coding genes are co-localized on the genome. Engreitz et al. [23] report that both coding and non-coding genes can regulate the expression of neighboring genes on the genome. There are several instances of lncRNAs that influence the nearby genes' expressions [24–26]. For example, Ørom et al. [27]

report that the depletion of some ncRNAs led to decreased expression of their neighboring protein-coding genes. Others also support the involvement of lncRNAs in the *cis* regulation, where both the regulatory ncRNA and the target gene are transcribed from the same or nearby genomic locus [28]. Based on these findings, in this work, we take into account the coding genes nearby to functionally assess a given ncRNA set. The transfer of functional annotation from nearby coding genes has been used in the general genomic interval set enrichment tools [29–32].

To offer broad functionality and applicability, NoRCE allows several additional features. The identified neighborhood coding gene set can be filtered or expanded with coding genes found to be co-expressed with the input ncRNAs. For this, NoRCE allows users to input their expression data or make use of pre-computed correlation results for The Cancer Genome Atlas (TCGA) project expression data. Since TAD boundaries affect the expression of neighboring genes [33], NoRCE also allows analysis that takes into account the topologically associated domain regions (TAD) boundaries on the genome. NoRCE provides miRNA specific options as well; the user can filter the neighbor set with predicted targets of the input miRNAs. Moreover, the input ncRNA set can be filtered based on ncRNA biotype (such as sense, antisense, lincRNA). NoRCE supports various commonly used statistical tests for enrichment.

In the following sections, we first detail the NoRCE's capabilities and the technical details. We also exemplify the NoRCE on two different functional analyses. In the first use case, we analyze the set of ncRNAs differentially expressed in brain disorder, while the second one showcases miRNA specific analysis on cancer patient data.

Implementation

Capabilities of NoRCE and workflow are summarized in Fig. 1. For a given set of ncRNAs, NoRCE first recognizes the coding genes close to ncRNA genes on the linear genome. Based on user-specified options, these genes are expanded or filtered using co-expressed genes, target predictions, or using the information on the TAD regions. Once the genes of interest are gathered, several gene enrichment analyses are performed. The details of these steps are provided in the following sections.

Species supported

NoRCE supports analysis for *Homo sapiens*, *Mus musculus* (house mouse), *Rattus norvegicus* (brown rat), *Danio rerio* (zebrafish), *Drosophila melanogaster* (fruit fly), *Caenorhabditis elegans* (worm) and *Saccharomyces cerevisiae* (yeast). For *Homo sapiens*, it handles human hg19 and hg38 assemblies. For the other species, it uses the most recent assembly of the species. Supported assemblies for different species are provided in Additional file 1: Table S1.

Curating the *cis* coding gene list

NoRCE accepts a set of any type of ncRNAs, $S = \{r_1, \ldots, r_n\}$. For each ncRNA, $r_i \in S$, in the input list, NoRCE identifies all proximal protein-coding genes in 1D genome. The proximal genes are considered as those that are within the base-pair limit of the genomic start coordinate of the input gene and/or within the base-pair limit of the genomic end coordinate of the input gene. If the coding gene r_i is located within the

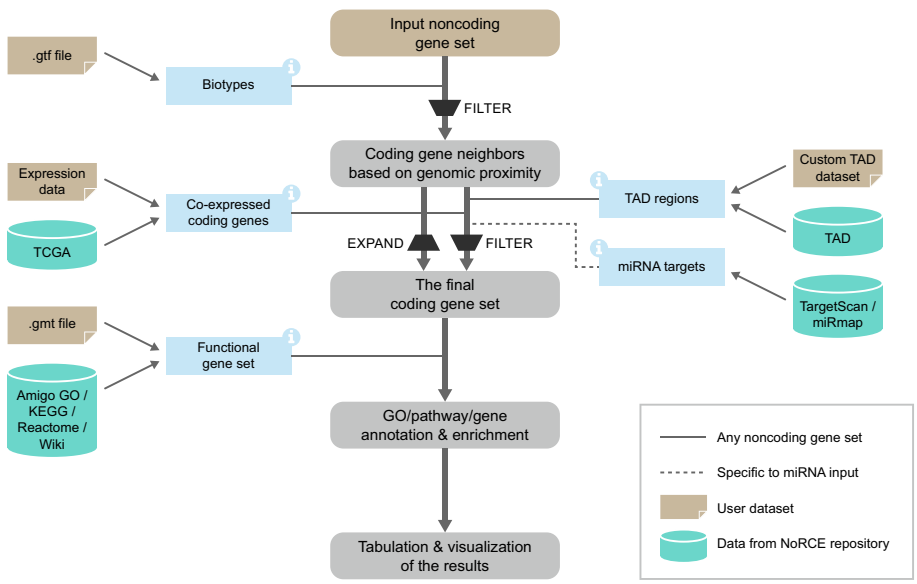

Fig. 1 The workflow of the NoRCE package

user-specified base-pair limit from the upstream and/or downstream of known transcription start and/or end position of the ncRNA gene, it is designated as a neighboring coding gene of r_i and added to the coding gene list pool of C_i. The union of the coding genes, constitute the final coding gene set to be tested for functional enrichment, $C = \cup_{i=1}^{n} C_i$. The pool of coding genes can be further filtered or expanded based on the additional biological evidence available, detailed in the next sections, with user-selected options. Users can also limit the analysis to the introns or exons of the neighboring coding genes. In that case, NoRCE applies the genomic proximity criterion on the intron or the exon of the genes based on the user's selection.

Input can be provided to NoRCE in the form of gene symbols, Ensembl genes and transcripts, Entrez IDs, or miRBase IDs. Since no single source contains information on all the transcripts, gene coordinates and their annotations are retrieved from two different databases: ENSEMBL [34] and UCSC [35]. We collect the ENSEMBL data via biomaRt package [36]. Genes are retrieved from UCSC using the rtracklayer package [37].

Incorporating co-expression information

Since coding genes that exhibit high co-expression patterns can hint to functional cooperation, NoRCE enables the user to incorporate co-expressed coding genes into the analysis. If the filtering option is set, each C_i is filtered such that only the neighboring coding genes that are also co-expressed with r_i are placed into C. If the expansion option is set, a coding gene is co-expressed with any of the $r_i \in S$ is added to C.

NoRCE enables the user to conduct the expression analysis with user input expression data. In this case, users are expected to load the expression data in TSV or TXT format; or they can use the *SummarizedExperiment* object in R. Before the correlation analysis, NoRCE executes a pre-processing step on expression data. The variance of each gene's

expression is calculated, and genes that vary lesser than the user-defined variance cutoff, 0.0025 by default, are excluded from the analysis. NoRCE supports commonly used correlation measures: Pearson Correlation, Kendall Rank Correlation, Spearman's Rank Correlation. The default values for correlation coefficient cutoff is 0.3, for significance p-value, 0.05 and confidence level 0.95. The user can set the correlation and significance cutoffs based on their need.

To assist analysis for cancer, NoRCE also allows using pre-computed co-expressed gene sets for ncRNAs measured in The Cancer Genome Atlas (TCGA) project [38]. Since TCGA contains the expression profiles for miRNA, mRNA, and lncRNA, this examination is limited to only the miRNA and lncRNA inputs. The co-expressed genes are defined using the Pearson correlation coefficient. Users can set the cutoff for the correlation coefficient.

Filtering genes with the TAD boundary information

The gene regulatory interactions are affected by the 3D chromatin structure of the genome [39]. On a single chromosome, chromatin compartmentalizes into sub-domains, named as topologically associating domain (TADs). TAD boundary regions insulate the *cis*-regulating elements [40]. NoRCE allows filtering based on TADs. If this option is selected, when curating the nearby genes of an ncRNA, NoRCE will only include the coding genes within the same TAD boundary with that of the ncRNA. We compile TAD regions for different cell-lines and species from various sources and made them available for use in conducting the analysis. These data sources and the species for which they are available are provided in Additional file 1: Table S2. NoRCE allows inputting BED formatted TAD boundary files. Thus, the user can conduct this analysis with other available TAD information.

Biotype specific analysis

If the user wants to conduct a biotype specific analysis, NoRCE can select the ncRNAs that are annotated with the given biotypes and use this biotype-filtered subset in the subsequent steps. Also, NoRCE allows extraction of ncRNAs of given biotypes S and performs analysis on the subset of genes that do not contain the genes annotated with given biotypes. NoRCE accepts GTF formatted GENCODE annotation files for biotype analysis.

miRNA target list

For miRNA specific inputs, NoRCE provides additional features. The coding gene set, C, can be restricted to the potential miRNA targets; thus, only neighboring coding genes that are also miRNA potential targets are included. The miRNA target list is curated from various sources. Computationally predicted miRNA-target interactions are obtained from the TargetScan [41] for the species except *Rattus norvegicus* as it is not available. Target predictions for *Rattus norvegicus* miRNAs are obtained from the miRmap [42]. No miRNA is reported for *Saccharomyces cerevisiae* [43]. Thus, NoRCE does not provide any miRNA analysis for *Saccharomyces cerevisiae*. Table 1 presents the details of the pre-computed target predictions.

Table 1 List of miRNA target prediction algorithms used for each species

Species	Database	Ver/date	References
Homo sapiens	TargetScan	v7.2	[41]
Mus musculus	TargetScan	v7.1	[89]
Rattus norvegicus	miRmap	2013	[42]
D. melanogaster	TargetScan	v6.2	[90]
Danio rerio	TargetScan	v6.2	[89]
C. elegans	TargetScan	v6.2	[91]

The version information is provided, which is the currently available version for the corresponding species

Enrichment analysis

Once the coding gene list, *C*, is curated, NoRCE conducts functional enrichment analysis. NoRCE supports analysis with various functional annotations: gene ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Reactome pathway, WikiPathways, genes, or GMT formatted integrated pathway dataset. For the annotation, we make use of biannually updated databases in Bioconductor. Gene ontologies and their annotated gene list are provided via GO.db [44] package. To increase the statistical analysis power, only the GO terms with at least 5 annotated protein-coding genes are considered as suggested by [17]. KEGG annotation is performed using KEGG.db [45], for Reactome enrichment analysis, NoRCE utilizes reactome.db [46]. NoRCE employs WikiPathways API to retrieve the pathways and annotated gene list [47].

NoRCE supports commonly used enrichment tests: hypergeometric test, Fisher Exact Test, Binomial Test, X^2 test. We refer readers to [48] for details of the statistical tests. The background gene set is all the genes in the functional annotation source that is selected. NoRCE provides the flexibility of providing a user-defined background gene set.

Presentations of the results

NoRCE provides different ways to export the results. All information in enrichment analysis can be retrieved in a tabular format. Also, users can set the number of top enrichment results to exported, and NoRCE outputs these results based on p-value or p adjusted values in a tabular format. Networks and dot plots can be used to visualize the enrichment results. The dot plot shows the top enriched terms, their p-values (or p adjusted values), and the number of enriched genes in the input neighbor set. In the network representation, the enriched terms are represented with nodes, and the ncRNA and coding transcripts related to the enriched terms are represented with edges. In this graph, the node size is proportional to the node degree. The nodes in the networks are clustered, and a color code distinguishes between node clusters. Modularity clustering is employed to cluster the nodes [49]. For network visualization features, NoRCE makes use of the igraph package [50].

NoRCE also offers specialized visualization options for pathway and GO analysis. GO enrichment results can be illustrated in a directed acyclic graph (GO-DAG). We derive the DAG information through the AmiGO API [51]. In this diagram, nodes

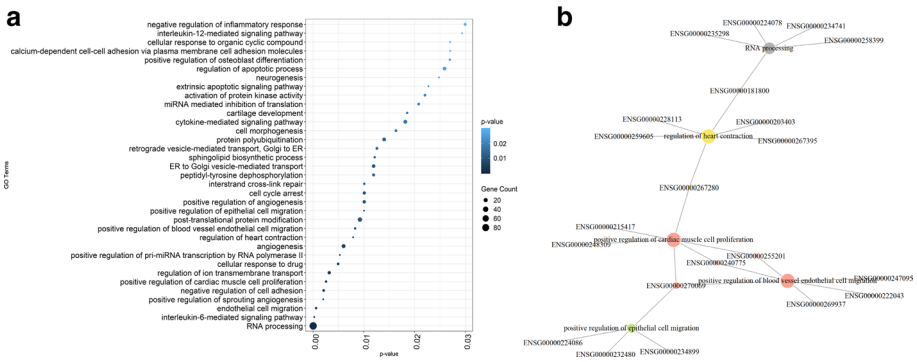

Fig. 2 **a** Top 35 GO biological processes enriched in the set of differentially expressed brain ncRNAs from [52]. The x-axis represents the *p*-values; the y-axis represents the GO terms. The dot area is proportional to the size of the overlapping gene set, and the color signifies the *p*-value of the enrichment test for the corresponding GO-term. **b** The functional enrichment analysis is repeated with TAD filtering and the GO-term:ncRNA network is provided. The size of the nodes is proportional to the degree of the node. Different colors are used to differentiate the clusters of nodes

are GO-terms, and edges indicate relation types between GO terms. Enriched GO-terms are colored according to their *p*-values or p-adjusted values. Users can export enriched GO-DAG diagrams in a PNG or SVG format. For pathway enrichment results, KEGG and Reactome enrichment results can be visualized within KEGG and Reactome maps. The enriched terms are marked with color using the KEGG and Reactome APIs, respectively. These visuals are displayed through the browser. In the results sections and the supplementary materials, we provide examples of these visualizations.

Results

To demonstrate how NoRCE could be used to analyze a list of ncRNAs functionally, we apply NoRCE on several problems and multiple independent datasets. We use the default parameter settings in the following analyses unless otherwise stated.

Case study 1: enrichment analysis of the ncRNAs for the psychiatric disorders

In this use case, we demonstrate the functional enrichment analysis of a set of ncRNAs related to brain disorders based on gene expression data measured by Gandal et al. [52]. These ncRNAs exhibit gene- or isoform-level differential expression in at least one of the following disorders: autism spectrum disorder (ASD), schizophrenia (SCZ), and bipolar disorder (BD). In total, the ncRNA gene set contains 1,363 differentially expressed human ncRNAs. We perform GO enrichment for biological processes and pathway enrichment analysis based on pathways provided by Bader Lab [53]. The number of pathways and the different pathway sources included in the Bader Lab set is provided in Additional file 1: Table S3. In these enrichment analyses, the background gene sets are described as the groups of all annotated genes in the corresponding GO or pathway dataset. The protein-coding genes that fall into this neighborhood region of the ncRNAs are input to the enrichment analysis. We also showcase NoRCE's ability to constrain the input set with protein-coding genes within the TAD boundaries.

Functional enrichment results

The dot plot in Fig. 2 shows the top 35 enriched BP GO-terms, sorted based on the significance of enrichment. The number of annotated genes with the corresponding GO-terms are provided in the graph. We detect RNA related GO terms such as the *positive regulation of pri-miRNA transcription by RNA polymerase II*, *miRNA mediated inhibition of translation*, and *RNA processing*. Additionally, various GO terms are pertinent to various neurological functions such as *response to ischemia*, *sensory perception of pain*, and *neurogenesis*. It has been reported that cerebral ischemia-induced genes are upregulated in schizophrenia [54], and it is common to have chronic pain in bipolar patients [55]. Interestingly, we observe that the enriched terms include cardiac and vascular-related functions. Several studies exhibit interactions between neural diseases and changes in blood vessel pathology and blood flow [56–58]. Others reveal that patients with bipolar disorder have low heart rate variability, which is a physiological measure of variation in the time interval between each heartbeat [59, 60].

Alternative visualizations of these functional enrichment results are provided in the Additional file 1 section. Additional file 1: Figure S1 shows the top 7 GO terms in a network visualization format. Additional file 1: Table S4 lists the top enriched GO BP terms to showcase the tabular format output capabilities.

Functional enrichment results with TAD filtering

We repeat the previous functional analysis when the TAD filtering is on. When this filter is applied, only the protein-coding genes near the ncRNAs in the input list, and at the same time reside within the same TAD regions are included in the enrichment analysis. In this analysis, we use custom defined TAD regions for the adult dorsolateral prefrontal cortex that are provided by the [52] study, and we keep all the other parameters in their default values.

Figure 2 illustrates the GO-term network for the top 7 enriched GO terms. Alternative representations of these results are provided in the Supplementary Materials (Additional file 1: Table S5, Figs. S2 (A), and S2 (B)). Interestingly, in this analysis, we identify cell cycle regulation related GO-terms. Cell cycle regulating genes have been associated with autism in GWAS studies [61]. In DNA derived from the pre-frontal cortex, cell cycle regulating genes show autism-specific CNVs [61]. In schizophrenia and bipolar disorder, many genes participate in cell cycle regulation and they have been shown to have differential expression levels [62]. We also identify cell adhesion in this enrichment analysis. Cell adhesion has been reported to be disrupted in autism [63] and schizophrenia [64]. Moreover, in schizophrenia and bipolar disorders, cell adhesion pathways have been reported to contribute to disease susceptibility [65].

Comparison of enrichment analysis with and without TAD-based filtering

We compare the enrichment analyses with and without TAD-based filtering to understand the effect of TAD filtering. When the enrichment analysis is based on only neighborhood genes, we detect 48 enriched biological processes. When we repeat our analysis with TAD filtering, we observe 29 enriched biological processes. The top 10 enriched terms are mostly the same for both analysis; these include cell cycle and cell adhesion-related terms, as well as several cardiac and vascular-related functional terms (*positive*

regulation of cardiac muscle cell proliferation, *positive regulation of blood vessel endothelial cell migration* and *angiogenesis*).

Running the enrichment analysis with TAD filtering allows us to uncover brain disorder-related GO-terms that are not identified by the enrichment analysis solely based on neighborhood genes. Using TAD filtering, we distinguish 6 enriched biological processes that have been reported to relate to a brain disorder in the literature, Table 3. Krishnan et al. [66] report that *Regulation of Rho protein signal transduction (GO:0035023), somatic stem cell population maintenance (GO:0035019), calcium ion transport (GO:0006816), ubiquitin-dependent protein catabolic process (GO:0006511)* are potential ASD related GO-terms. Rho GTPases are important regulators of the neural system, and mutations in Rho GTPases' regulators and effectors can cause neural diseases, including ALS [67]. Moreover, we also observe *positive regulation of phosphatidylinositol 3-kinase signaling (GO:0014068)* after TAD-filtering. Phosphoinositide 3-kinase is a well-known pathway that regulates several processes, including proliferation, growth, apoptosis, and cytoskeletal rearrangement [68]. It is linked with several diseases and it is considered as a hallmark of cancer [69]. Kurek et al. [70] show that cancer-associated PIK3CA mutations cause epilepsy, and there is a strong correlation between epilepsy and autism [71]. Moreover, Krishnan et al. [66] report that this GO-term is related to ASD. Detecting neural disease linked GO-terms by employing TAD filtering shows that TAD filtering might help arrive at more precise enrichment results. In conclusion, using different approaches can lead to more nuanced enrichment analysis results.

Pathway enrichment using predefined pathway gene sets

NoRCE enables pathway enrichment analysis for various sources, including KEGG, Reactome pathway, and WikiPathways. Also, NoRCE supports pathway enrichment using custom pathway databases such as MSigDb [53], or other user-curated data provided in GMT format. To showcase the NoRCE capability, we utilize the Bader Lab dataset as the user-defined pathway gene set analysis. In this analysis, we only consider the genes in the neighborhood of the differentially expressed ncRNAs in brain disorders.

Interestingly, we find many enriched pathways that are related to neural diseases. Some of these pathways directly related to ASD, schizophrenia, and bipolar disorder, including *Synaptic signaling pathway* associated with an *autism spectrum disorder, WP4539*; *Amyotrophic lateral sclerosis, WP2447*; *Alzheimer's disease, WP2059*. Additionally, we found that many of the signaling pathways, such as *G-Protein Signaling*, *mTOR signalling*, *MAPK Signaling* pathway and those pathways are associated with at least one of the brain disorder: autism spectrum disorder, schizophrenia, and bipolar disorder [72–74]. Due to the space limit, we list a subset of enriched disease-related pathways in Additional file 1: Table S6 and S7.

Comparison between ASD associated GO-terms and NoRCE enrichment results

Krishnan et al. [66] predict novel ASD risk genes based on the brain-specific functional network. In their work, they also identified functions potentially dysregulated by ASD-associated mutations. We compare our findings with this set of ASD associated GO function terms [66]. We observe that most of the enriched GO terms reported in NoRCE

are listed as potential ASD-related GO terms in Krishnan et al. [66] study. The ncRNA enrichment analysis of NoRCE without TAD filtering identifies 48 enriched GO terms, and 32 of these terms are also in the list of ASD related GO terms [66], corresponding to 67% overlap. When TAD filtering is applied, there are 29 enriched terms, and 21 are in the ASD GO term list, corresponding to 73% overlap (Additional file 1: Table S9). We test the significance of these overlap ratios. We randomly select ncRNA set with the same size of the input from the all gene population. We find the enriched term with this random gene set and checked if the overlap ratio is equal or higher in the randomized case. A p-value is calculated by repeating this procedure 1000 times.

Case study 2: functional enrichment analysis of variably expressed miRNAs in brain cancer using miRNA targets

NoRCE offers a filtering option for the input miRNA's targets. Users can choose to filter ncRNA neighbors, such that only those that are the targets of the miRNA are included. To demonstrate this option, we use a set of miRNAs that are differentially expressed ncRNAs in brain cancer obtained from dbDEMC 2.0 [75] for the functional analysis. This set contains 407 miRNAs and is provided in NoRCE with the name *brain_miRNA*. We choose the Reactome pathway as the functional gene set.

We identify *lysosome vesicle biogenesis* (p-value $= 7.1\text{e}{-05}$), *trans-golgi network vesicle budding* (p-value $= 0.0006$), *ion channel transport* (p-value $= 0.0091$), and *axon guidance* (p-value $= 0.0382$) pathways as enriched. Previous studies report that the *axon guidance* and *ion channel transport* pathways are related to the Glioblastoma Multiforme [76, 77]. Other evidence also suggests that miRNAs could be acting as key fine-tuning regulatory elements in *axon guidance* [78].

Case study 3: functional enrichment analysis with co-expression analysis

NoRCE also supports filtering based on a co-expression analysis. When defining coding gene neighborhoods for an ncRNA, the user can choose to include a coding gene only if it is co-expressed. Alternatively, the users can choose to augment the coding genes list with the co-expressed coding gene set. To demonstrate this option, we use NoRCE on the brain cancer patient data obtained from TCGA.

The TCGA data include expression levels for mRNA and miRNA for matched primary tumor solid samples from 527 tumor patients. miRNA-seq data are measured as per million mapped reads (RPM) values, and RNA-seq data are measured as Fragments per Kilobase of transcript per Million mapped reads upper quartile normalization (FPKM–UQ). We apply the same pre-processing step as in our previous method [79]. Genes and miRNAs that have very low expression levels (RPKM < 0.05) in many patients (more than 20% of the samples) are filtered out. The gene expression values are $log2$ transformed, and those with high variability are retained for co-expression analysis. For this aim, only the genes with median absolute deviation (MAD) above 0.5 are used. The final expression dataset contains 444 miRNA and 12,643 mRNA genes on 527 tumor patients on which we perform Pearson correlation analysis. The mRNAs which have more than 0.1 correlation with a miRNA are retained.

Table 2 Pathway enrichment results for nearby co-expressed genes with miRNAs

Pathway ID	Pathway term	*p*-value	GeneRatio	BGRatio	EGNo	ncGeneList
R-HSA-9006934	Signaling by receptor tyrosine kinases	0.0006	8/44	473/10654	8	hsa-mir-199b, hsa-mir-214, hsa-mir-3934, hsa-mir-455, hsa-mir-483, hsa-mir-141, hsa-mir-338, hsa-mir-3689a
R-HSA-422475	Axon guidance	0.0017	8/44	553/10654	8	hsa-mir-199b, hsa-mir-214, hsa-mir-4684, hsa-mir-585, hsa-mir-95, hsa-mir-150, hsa-mir-3689a
R-HSA-1266738	Developmental biology	0.0327	9/44	1097/10654	9	hsa-mir-199b, hsa-mir-214, hsa-mir-4684, hsa-mir-585, hsa-mir-935, hsa-mir-95, hsa-mir-150, hsa-mir-3689a

The GeneRatio is computed by dividing the overlapping with the coding genes with the functional gene set to the number of all protein-coding genes within the input set neighbourhood. The BGRatio column represents the ratio of the number of genes found in the enriched GO term set to the size of the background gene set. The EGNo refers to the size of the overlap between the corresponding GO term gene set and the neighboring coding gene set. ncGeneList column contains ncRNA genes that are enriched with the corresponding GO-term

Table 3 Brain disorder related biological process GO term enrichment results that show the TAD analysis enhancement

ID	Term	*p*-value	References
GO:0035023	Regulation of Rho protein signal transduction	0.0187	[66, 67]
GO:0035019	Somatic stem cell population maintenance	0.0205	[66]
GO:0006816	Calcium ion transport	0.0360	[66]
GO:0014068	Positive regulation of phosphatidylinositol 3-kinase signaling	0.0219	[66, 69, 92]
GO:0006511	Ubiquitin-dependent protein catabolic process	0.0277	[66, 93]
GO:0007417	Central nervous system development	0.0255	[94, 95]

The listed enriched GO terms are detected by enrichment analysis based on neighboring genes and TAD filtering but are undetected by enrichment analysis solely based on neighboring genes. The TAD information is obtained from adult dorsolateral prefrontal cortex data [52] study. For a given enriched term, we provide the *p*-value and a list of literature evidence that shows the link between brain disorders and the enriched term

When we examine the enriched pathways, *Signaling by Receptor Tyrosine Kinases* emerges as an important pathway. *Receptor Tyrosine Kinases* is a cell surface receptor family, and its members are responsible for growth factors, hormones, cytokines, neurotrophic factors [80]. Following their activation, they can signal through downstream pathways responsible for survival, differentiation, and angiogenesis [80]. Inhibition on *Receptor Tyrosine Kinases* and their signal pathways are utilized as target therapy on brain cancer [81]. Also, miRNAs are reported to take a role as mediators or suppressors in these pathways and promote tumor cell death [81]. The Reactome diagram for this pathway is illustrated in Additional file 1: Figure S4. In both target-based and co-expression-based analyses on the differentially expressed miRNAs, *Axon guidance pathway* is enriched. MiRNAs' role in axon guidance have been reported elsewhere [78, 82–84]. For example, Baudet et al. [84] report that miR-124 controls Sema3A, which is essential for normal axon guidance. Accumulating evidence also points out that axonogenesis is stimulated by malignant cells and contributes to cancer growth

and metastasis [85]. These findings also support the NoRCE capability for finding interesting functional inferences. The details of the enrichment results are provided in Table 2.

Case study 4: functional enrichment analysis of pan-cancer driver lncRNAs

As a fourth case study, we conduct an analysis where the input list comprises lncRNAs. High-throughput sequencing technologies have revealed that there are thousands of lncRNAs whose aberrant expressions are associated with different cancer types [86]. As part of the ICGC/TCGA Pan-Cancer Analysis of Whole Genomes (PCAWG) Consortium, the Cancer LncRNA Census provides a dataset of 122 high-confidence lncRNAs with causal roles in cancer phenotypes [87]. We utilize these known tumor suppressor or oncogene lncRNAs and conduct enrichment analysis with NoRCE. For this enrichment analysis, we use the neighborhood genes filtered based on TAD boundaries obtained from 3D Genome Browser [88]. All other choices and parameters are set to their default values.

Enrichment analysis of this set of 122 cancer associated lncRNAs yields 11 enriched biological processes (Additional file 1: Table S11). Interestingly, 2 of the 11 enrichment biological processes are related to miRNA processes. This could indicate that some of the cancer related lncRNAs are located in the same TAD regions as the miRNA host genes. Moreover, we detect developmental processes such as *anatomical structure development, multicellular organism development, anterior/posterior pattern specification*. This may indicate that these cancer related lncRNAs have a role in developmental processes. Also, analysis on the network for the enriched GO-term and their annotated genes, Additional file 1: Figure S5, demonstrate that the RNA process and their annotated genes form a separate graph from other enriched terms and genes.

We repeat the same analysis without considering the neighborhood gene information. In this case, we only consider the coding genes that partially overlap with the input lncRNA set and fall into the same TAD boundary with the lncRNA genes. This way we are able to measure the effects of including genes nearby on the genomic sequence for enrichment analysis. We detect 8 enriched biological processes. When we compare our findings with results obtained for enrichment analysis based on neighborhood genes filtered with TAD boundaries, we are unable to detect two developmental process (*anterior/posterior pattern specification and anatomical structure morphogenesis* and one miRNA related GO-term (*miRNA mediated inhibition of translation*. This finding is a subset of results that are obtained by cis-based gene enrichment filtering with TAD boundaries. Thus, we recommend carrying out enrichment analysis by combining multiple information sources such as cis genes, TAD boundaries, co-expression analyses.

Discussion

In showcasing NoRCE, we analyzed sets of ncRNAs implicated in diseases, including brain disorder related ncRNAs and cancer-related lncRNAs and miRNAs. Functional enrichment of these ncRNAs yielded interesting biological findings highlighting how NoRCE could be useful in answering a wide range of questions. The datasets and examples showcased here are also provided in the R\Bioconductor package.

NoRCE uses functional sets such as those derived from GO and pathway databases and miRNA prediction tools. Improvements in these databases and tools allow NoRCE conduct more accurate analysis. NoRCE is designed for non-coding RNAs, but can use both coding and non-coding RNAs as input. Currently, the user can use NoRCE to conduct analysis in human and mouse, rat, zebrafish, fruit fly, worm, and yeast. As a future direction, NoRCE can be extended to support analysis for other species. Moreover, the current version of the package contains only miRNA target predictions. However, NoRCE can be enhanced by including target prediction for other ncRNAs, including sRNAs and snoRNAs.

Conclusions

NoRCE is a comprehensive, flexible, and user-friendly tool for enrichment analysis of all types of ncRNAs. It works for multi-species and is available as an R package. NoRCE, unlike existing tools, conducts enrichment by taking into account the genomic neighborhood of the ncRNAs in the input set and transfers functional annotations of these coding genes. We should note that although cis-regulation has been reported for many ncRNA types, it may not hold for all types of ncRNAs. Therefore, in addition to the genomic neighborhood-based analysis, NoRCE allows the standard approaches of using coding genes co-expressed with the input ncRNAs in detecting the enriched functions. Another unique feature of NoRCE that it allows an option for making use of TAD regions. NoRCE provides flexibility to the user; the user can perform analysis with different options and use the library's readily available datasets to conduct the analysis or input custom datasets. It is also possible to include or exclude any analysis that NoRCE contains.

Availability and requirements

Project name: NoRCE
Project home page: http://bioconductor.org/packages/release/bioc/html/NoRCE.html
Operating system(s): Platform independent
Programming language: R
Other requirements: Listed in http://bioconductor.org/packages/release/bioc/html/NoRCE.html
License: MIT license
Any restrictions to use by non-academics: None

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12859-021-04112-9.

Additional file 1. Suplementary information and results for the package and further analysis.

Acknowledgements
The results shown here are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga. We also thank Marcel Ramos, our R\Bioconductor package code reviewer, Martin Morgan, Michael Lawrence, and Lori Shepherd for their valuable feedback, comments, and time on our R/Bioconductor package.

Authors' contributions
G.O and O.T. designed the study. G.O. implemented the package with feedback by A. N. All authors contributed to the results, discussions, and writing the manuscript. All authors read and approved the final manuscript.

Funding
We thank Sabanci University and Bilkent University for internal funding support. The funders had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials
Code and Tutorial are in https://github.com/guldenolgun/NoRCE

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Computer Engineering, Bilkent University, Ankara, Turkey. [2]Present Address: Faculty of Engineering and Natural Sciences, Sabanci University, 34956 Istanbul, Turkey. [3]Cancer Data Science Lab, National Cancer Institute, National Institute of Health, Bethesda, MD, USA.

Received: 29 November 2020 Accepted: 30 March 2021
Published online: 02 June 2021

References

1. Djebali S, Davis AMCA, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, et al. Landscape of transcription in human cells. Nature. 2012;489(7414):101–8.
2. Ulitsky I, Bartel DP. lincrnas: genomics, evolution, and mechanisms. Cell. 2013;154(1):26–46.
3. Batista PJ, Chang HY. Long noncoding RNAs: cellular address codes in development and disease. Cell. 2013;152(6):1298–307.
4. Iyer MK, Niknafs YS, Malik R, Singhal U, et al. The landscape of long noncoding RNAs in the human transcriptome. Nat Genet. 2015;47(3):199.
5. Cech TR, Steitz JA. The noncoding RNA revolution-trashing old rules to forge new ones. Cell. 2014;157(1):77–94.
6. Geisler S, Coller J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. Nat Rev Mol Cell Biol. 2013;14(11):699–712.
7. Quinodoz S, Guttman M. Long noncoding RNAs: an emerging link between gene regulation and nuclear organization. Trends Cell Biol. 2014;24(11):651–63.
8. Esteller M. Non-coding RNAs in human disease. Nat Rev Genet. 2011;12(12):861.
9. Yao R-W, Wang Y, Chen L. Cellular functions of long noncoding RNAs. Nat Cell Biol. 2019;21(5):542–51.
10. Wu X, Watson M. CORNA: testing gene lists for regulation by microRNAs. Bioinformatics. 2009;25(6):832–3.
11. Vlachos IS, Zagganas K, Paraskevopoulou MD, Georgakilas G, et al. Diana-mirpath v3. 0: deciphering microRNA function with experimental support. Nucl Acids Res. 2015;43(W1):460–6.
12. Hsu JBK, et al. miRTar: an integrated system for identifying miRNA-target interactions in human. BMC Bioinform. 2011;12:300.
13. Vila-Casadesús M, Gironella M, Lozano JJ. Mircomb: An r package to analyse miRNA-mRNA interactions. Examples across five digestive cancers. PLoS ONE. 2016;11(3):0151127.
14. Borgm"astars E, de Weerd HA, Lubovac-Pilav Z. miRFA: an automated pipeline for microRNA functional analysis with correlation support from TCGA and TCPA expression data in pancreatic cancer. BMC Bioinformatics. 2019;420:393.
15. Kowarsch A, Preusse M, Marr C, Theis FJ. miTALOS: analyzing the tissue-specific regulation of signaling pathways by human and mouse microRNAs. RNA. 2011;17(5):809–19.
16. Preusse M, Theis FJ, Mueller NS. miTALOS v2: analyzing tissue specific microRNA function. PLoS ONE. 2016;11:3.

17. Jiang Q, Ma R, Wang J, Wu X, Jin S, Peng J, Tan R, Zhang T, Li Y, Wang Y. Lncrna2function: a comprehensive resource for functional investigation of human lncRNAs based on RNA-seq data. BMC Genom. 2015;16(3):52.
18. Park C, Yu N, Choi I, Kim W, Lee S. lncRNAtor: a comprehensive resource for functional investigation of long non-coding RNAs. Bioinformatics. 2014;30(17):2480–5.
19. Zhao Z, Bai J, Wu A, Wang Y, et al. Co-lncRNA: investigating the lncRNA combinatorial effects in go annotations and KEGG pathways based on human RNA-seq data. Database. 2015;2015:082.
20. Li JH, Liu S, Zhou H, Qu LH, Yang JH. starbase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale clip-seq data. Nucleic Acids Res. 2014;42:92–7.
21. Meng X, Hu D, Zhang P, Chen Q, Chen M. CircFunBase: a database for functional circular RNAs. Database 2019;2019.
22. Thévenin A, Ein-Dor L, Ozery-Flato M, Shamir R. Functional gene groups are concentrated within chromosomes, among chromosomes and in the nuclear space of the human genome. Nucl Acids Res. 2014;42(15):9854–61.
23. Engreitz JM, Haines JE, Perez EM, Munson G, et al. Local regulation of gene expression by lncRNA promoters, transcription and splicing. Nature. 2016;539(7629):452.
24. Lee JT. Lessons from x-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. Genes Dev. 2009;23(16):1831–42.
25. Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R, Fraser P. The air noncoding RNA epigenetically silences transcription by targeting g9a to chromatin. Science. 2008;322(5908):1717–20.
26. Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA, et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature. 2011;472(7341):120–4.
27. Ørom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytnicki M, Notredame C, Huang Q, et al. Long noncoding RNAs with enhancer-like function in human cells. Cell. 2010;143(1):46–58.
28. Guil S, Esteller M. Cis-acting noncoding RNAs: friends and foes. Nat Struct Mol Biol. 2012;19(11):1068–75.
29. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. Great improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010;28(5):495.
30. Welch RP, Lee C, Imbriano PM, Patil S, Weymouth TE, Smith RA, Scott LJ, Sartor MA. Chip-enrich: gene set enrichment testing for chip-seq data. Nucl Acids Res. 2014;42(13):105.
31. Chicco D, Bi HS, Reimand J, Hoffman MM. Behst: genomic set enrichment analysis enhanced through integration of chromatin long-range interactions. biorxiv. bioRxiv 2019
32. Otlu B, Firtina C, Keleş S, Tastan O. Glanet: genomic loci annotation and enrichment tool. Bioinformatics. 2017;33(18):2818–28.
33. Lupiánez DG, Spielmann M, Mundlos S. Breaking tads: how alterations of chromatin domains result in disease. Trends in Genet. 2016;32(4):225–37.
34. Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R, et al. Ensembl. Nucleic Acids Res. 2020;48(D1):682–8.
35. Haeussler M, et al. The UCSC genome browser database: 2019 update. Nucleic Acids Res. 2019;47(D1):853–8.
36. Durinck S, Moreau Y, Kasprzyk A, Davis S, Moor BD, Brazma A, Huber W. Biomart and bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics. 2005;21(16):3439–40.
37. Lawrence M, Gentleman R, Carey V. rtracklayer: an R package for interfacing with genome browsers. Bioinformatics. 2009;25(14):1841–2.
38. Ahmed M, Nguyen H, Lai T, Kim DR. miRCancerdb: a database for correlation analysis between microRNA and gene expression in cancer. BMC Res Notes. 2018;11(1):103.
39. Boyan B, Cavalli G. Organization and function of the 3d genome. Nat Rev Genet. 2016;17(11):661.
40. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012;485(7398):376–80.
41. Shin C, Nam J-W, Farh KK-H, Chiang HR, Shkumatava A, Bartel DP. Expanding the microRNA targeting code: functional sites with centered pairing. Mol Cell. 2010;38(6):789–802.
42. Vejnar CE, Zdobnov EM. Mirmap: comprehensive prediction of microRNA target repression strength. Nucleic Acids Res. 2012;40(22):673–83.
43. Panni S, Prakash A, Bateman A, Orchard S. The yeast noncoding RNA interaction network. RNA. 2017;23(10):1479–92.
44. Carlson M, Falcon S, Pages H, Li N. Go. db: A set of annotation maps describing the entire gene ontology. R package version. 2017; 3, 568
45. Carlson M. Kegg. db: A set of annotation maps for kegg. r package version 3.1. 2 (2016)
46. Ligtenberg W. reactome. db: A set of annotation maps for reactome. R package version. 2018; 1
47. Slenter DN, Kutmon M, Hanspers K, Riutta A, Windsor J, Nunes N, Mélius J, Cirillo E, Coort SL, Digles D, et al. Wikipathways: a multifaceted pathway database bridging metabolomics to other omics research. Nucleic Acids Res. 2017;46(D1):661–7.
48. Krishnamoorthy K. Handbook of statistical distributions with applications. New York: Chapman and Hall; 2016.
49. Brandes U, Delling D, Gaertler M, Gorke R, Hoefer M, Nikoloski Z, Wagner D. On modularity clustering. IEEE Trans Knowl Data Eng. 2008;20(2):172–88.
50. Csardi G, Nepusz T. The igraph software package for complex network research. InterJ Compl Syst. 2006;1695:1–9.
51. Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., Lewis, S., Hub, A., Group, W.P.W. Amigo: online access to ontology and annotation data. Bioinformatics. 2008;25(2):288–9.
52. Gandal MJ, Zhang P, Hadjimichael E, Walker RL, et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science. 2018;362(6420):8127.
53. Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS ONE. 2010;5(11):13984.
54. Moises HW, Hess M, Binder H. Cerebral ischemia-induced genes are increased in acute schizophrenia: An opportunity for clinical translation of genomic research findings. bioRxiv, 2017; 158436
55. Failde I, Duenas M, Agüera-Ortíz L, Cervilla JA, Gonzalez-Pinto A, Mico JA. Factors associated with chronic pain in patients with bipolar depression: a cross-sectional study. BMC Psychiatry. 2013;13(1):112.

56. Kealy J, Greene C, Campbell M. Blood-brain barrier regulation in psychiatric disorders. Neurosci Lett. 2018;726:133664.
57. Casas BS, Vitória G, do Costa MN, da Costa RM, Trindade P, Maciel R, Navarrete N, Rehen, SK, Palma V. hipsc-derived neural stem cells from patients with schizophrenia induce an impaired angiogenesis. Transl Psychiatry 2018;(1):1–15
58. Baruah J, Vasudevan A. The vessels shaping mental health or illness. Open Neurol J. 2019;13:1.
59. Henry BL, Minassian A, Paulus MP, Geyer MA, Perry W. Heart rate variability in bipolar mania and schizophrenia. J Psychiatr Res. 2010;44(3):168–76.
60. Cohen H, Kaplan Z, Kotler M, Mittelman I, Osher Y, Bersudsky Y. Impaired heart rate variability in euthymic bipolar patients. Bipolar Disord. 2003;5(2):138–43.
61. Chow ML, Pramparo T, Winn ME, Barnes CC, Li H-R, Weiss L, Fan J-B, Murray S, April C, Belinson H, et al. Age-dependent brain gene expression and copy number anomalies in autism suggest distinct pathological processes at young versus mature ages. PLoS Genet. 2012;8:3.
62. Benes FM, Lim B, Subburaju S. Site-specific regulation of cell cycle and DNA repair in post-mitotic GABA cells in schizophrenic versus bipolars. Proc Nat Acad Sci. 2009;106(28):731–6.
63. Betancur C, Sakurai T, Buxbaum JD. The emerging role of synaptic cell-adhesion pathways in the pathogenesis of autism spectrum disorders. Trends Neurosci. 2009;32(7):402–12.
64. Fan Y, Abrahamsen G, Mills R, Calderón CC, Tee JY, Leyton L, Murrell W, Cooper-White J, McGrath JJ, Mackay-Sim A. Focal adhesion dynamics are altered in schizophrenia. Biol Psychiatry. 2013;74(6):418–26.
65. O'Dushlaine C, Kenny E, Heron E, Donohoe G, Gill M, Morris D, Corvin A. Molecular pathways involved in neuronal cell adhesion and membrane scaffolding contribute to schizophrenia and bipolar disorder susceptibility. Mol Psychiatry. 2011;16(3):286–92.
66. Krishnan A, et al. Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. Nature Neurosci. 2016;19(11):1454.
67. Kasri NN, Van Aelst L. Rho-linked genes and neurological disorders. Pflügers Archiv-Eur J Physiol. 2008;455(5):787–97.
68. Vivanco I, Sawyers CL. The phosphatidylinositol 3-kinase-AKT pathway in human cancer. Nat Rev Cancer. 2002;2(7):489–501.
69. Fruman DA, Chiu H, Hopkins BD, Bagrodia S, Cantley LC, Abraham RT. The PI3K pathway in human disease. Cell. 2017;170(4):605–35.
70. Kurek KC, Luks VL, Ayturk UM, Alomari AI, Fishman SJ, Spencer SA, Mulliken JB, Bowen ME, Yamamoto GL, Kozakewich HP, et al. Somatic mosaic activating mutations in PIK3CA cause cloves syndrome. Am J Hum Genet. 2012;90(6):1108–15.
71. Besag FM. Epilepsy in patients with autism: links, risks and treatment challenges. Neuropsychiatric Disease Treat. 2018;14:1.
72. Godoy JA, Rios JA, Zollezzi JM, Braidy N, Inestrosa NC. Signaling pathway cross talk in Alzheimer's disease. Cell Commun Signal. 2014;12(1):23.
73. Tomita H, Ziegler ME, et al. G protein-linked signaling pathways in bipolar and major depressive disorders. Front Genet. 2013;4:297.
74. Vithayathil J, Pucilowska J, Landreth GE. ERK/MAPK signaling and autism spectrum disorders. Prog Brain Res. 2018;241:63–112.
75. Yang Z, Wu L, Wang A, Tang W, et al. dbDEMC 2.0: updated database of differentially expressed miRNAs in human cancers. Nucl Acids Res. 2016;45(D1):812–8.
76. Chedotal A, Kerjan G, Moreau-Fauvarque C. The brain within the tumor: new roles for axon guidance molecules in cancers. Cell Death Differ. 2005;12(8):1044–56.
77. Julia P, et al. Ion channel expression patterns in glioblastoma stem cells with functional and therapeutic implications for malignancy. PLoS ONE. 2017;12(3):e0172884.
78. Iyer AN, et al. microRNAs in axon guidance. Front Cell Neurosci. 2014;8:78.
79. Olgun G, Sahin O, Tastan O. Discovering lncRNA mediated sponge interactions in breast cancer molecular subtypes. BMC Genom. 2018;19(1):650.
80. Pearson JRD, Regad T. Targeting cellular pathways in glioblastoma multiforme. Signal Trans Target Therapy. 2017;2(1):1–11.
81. Pedro C, et al. High-throughput screening uncovers miRNAs enhancing glioblastoma cell susceptibility to tyrosine kinase inhibitors. Hum Mol Genet. 2017;26(22):4375–87.
82. Yi S, et al. Regulation of Schwann cell proliferation and migration by miR-1 targeting brain-derived neurotrophic factor after peripheral nerve injury. Sci Rep. 2016;6(1):1–10.
83. Li S, et al. MiR-340 regulates fibrinolysis and axon regrowth following sciatic nerve injury. Mol Neurobiol. 2017;54(6):4379–89.
84. Marie-Laure B, et al. miR-124 acts through CoREST to control onset of sema3a sensitivity in navigating retinal growth cones. Nat Neurosci. 2012;15(1):29.
85. Benoni B, et al. Nerve dependence: from regeneration to cancer. Cancer Cell. 2017;31(3):342–54.
86. Huarte M. The emerging role of lncRNAs in cancer. Nat Med. 2015;21(11):1253–61.
87. Carlevaro-Fita J, Lanzós A, Feuerbach L, Hong C, Mas-Ponte D, Pedersen JS, Johnson R. Cancer lncRNA census reveals evidence for deep functional conservation of long noncoding RNAs in tumorigenesis. Commun Biol. 2020;3(1):1–16.
88. Wang Y, Song F, Zhang B, Zhang L, Xu J, Kuang D, Li D, Choudhary MN, Li Y, Hu M, et al. The 3d genome browser: a web-based browser for visualizing 3d genome organization and long-range chromatin interactions. Genome Biol. 2018;19(1):151.
89. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. cell. 2005;120(1):15–20.
90. Ruby JG, Stark A, Johnston WK, Kellis M, Bartel DP, Lai EC. Evolution, biogenesis, expression, and target predictions of a substantially expanded set of drosophila microRNAs. Genome Res. 2007;17(12):1850–64.
91. Jan CH, Friedman RC, Ruby JG, Bartel DP. Formation, regulation and evolution of Caenorhabditis elegans 3' UTRs. Nature. 2011;469(7328):97.

92. Sánchez-Alegría K, Flores-León M, Avila-Muñoz E, Rodríguez-Corona N, Arias C. Pi3k signaling in neurons: a central node for the control of multiple functions. Int J Mol Sci. 2018;19(12):3725.
93. Rao G, Croft B, Teng C, Awasthi V. Ubiquitin-proteasome system in neurodegenerative disorders. J Drug Metab Toxicol. 2015;6(4):187.
94. Ziats MN, Rennert OM. Expression profiling of autism candidate genes during human brain development implicates central immune signaling pathways. PLoS ONE. 2011;6(9):24691.
95. Kumar A, Swanwick CC, Johnson N, Menashe I, Basu SN, Bales ME, Banerjee-Basu S. A brain region-specific predictive gene map for autism derived by profiling a reference gene set. PLoS ONE. 2011;6(12):28431.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

BMC