


DATABASE

Open Access



BDdb: a comprehensive platform for exploration and utilization of birth defect multi-omics data

Dengwei Zhang^{1,2,3,4†}, Si Zhou^{1,5†}, Ziheng Zhou^{2†}, Xiaosen Jiang^{1,2†}, Dongsheng Chen^{2†}, Hai-Xi Sun^{3,4,6}, Jie Huang⁷, Shoufang Qu⁷, Songchen Yang², Ying Gu^{2,3,4}, Xiuqing Zhang², Xin Jin^{2,8*}, Ya Gao^{2,9*} , Yue Shen^{2,3,4*} and Fang Chen^{2,10*}

Abstract

Background: Birth defects pose a major challenge to infant health. Thus far, however, the causes of most birth defects remain cryptic. Over the past few decades, considerable effort has been expended on disclosing the underlying mechanisms related to birth defects, yielding myriad treatises and data. To meet the increasing requirements for data resources, we developed a freely accessible birth defect multi-omics database (BDdb, <http://t21omics.cn/gb.org>) consisting of multi-omics data and potential disease biomarkers.

Results: In total, omics datasets from 136 Gene Expression Omnibus (GEO) Series records, including 5245 samples, as well as 869 biomarkers of 22 birth defects in six different species, were integrated into the BDdb. The database provides a user-friendly interface for searching, browsing, and downloading data of interest. The BDdb also enables users to explore the correlations among different sequencing methods, such as chromatin immunoprecipitation sequencing (ChIP-Seq) and RNA sequencing (RNA-Seq) from different studies, to obtain the information on gene expression patterns from diverse aspects.

Conclusion: To the best of our knowledge, the BDdb is the first comprehensive database associated with birth defects, which should benefit the diagnosis and prevention of birth defects.

Keywords: Birth defects, Chromosomal abnormality, Omics, Biomarker

Background

Birth defects refer to abnormalities present at birth in form, function, biochemistry, and mentality [1]. More than 8.14 million children are born with severe birth defects each year, which are among the principal causes of infant mortality [2]. The impacts of birth defects on human society are widespread, not only affecting survival

and quality of life for those affected, but also resulting in emotional and economic burdens on the family [3]. Although many studies have attempted to unveil the causes of birth defects, most remain vague [2, 3]. The established causes of birth defects can be roughly divided into three categories: i.e., genetic factors, environmental factors, and their interactions [4]. Genetic factors include chromosomal aberrations and genetic mutations, which can render severe intellectual disabilities and deformity [2, 5].

Chromosomal abnormalities are one of the main causes of birth defects with a known etiology [3], with nearly one in 200 newborns affected [6]. Most of these children

*Correspondence: jinjin@genomics.cn; gaoya@genomics.cn; shengyue@genomics.cn; chenfang1@genomics.cn

[†]Dengwei Zhang, Si Zhou, Ziheng Zhou, Xiaosen Jiang and Dongsheng Chen have contributed equally to this work

²BGI-Shenzhen, Shenzhen 518083, People's Republic of China
Full list of author information is available at the end of the article



© 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.

suffer severe intellectual disability and tissue and organ deformities. The most common chromosomal abnormalities include trisomy 21 (Down syndrome), trisomy 18 (Edwards syndrome), trisomy 13 (Patau syndrome), and sex chromosome aneuploidies such as monosomy X (Turner syndrome), conferring more than 80% of prenatal diagnosis of chromosomal abnormalities [6]. Among them, Down syndrome is the most common disease, with an occurrence of 1/319 to 1/1 000 in different populations [3, 7].

Birth defects resulting from chromosomal abnormalities have almost no effective treatment, although they are identifiable at early pregnancy through prenatal screening and diagnosis. To facilitate the study of human diseases, many databases (e.g., MalaCards, DisGeNET, DGA, RareDDB, miR2Disease, HEDD) have been developed with a focus on the annotation of diseases and related genes [8–13]. However, there is currently no multi-omics database for collating, storing, integrating, and displaying birth defect-related research datasets. To fill this gap, we developed a freely accessible birth defect multi-omics database (BDdb, <http://t21omics.cngb.org>).

The current version of the BDdb includes a total of 326 records derived from 136 GSE Series records, involving 12 tissues and 35 cell types for human, as well as 12 tissues and 17 cell lines for mouse. This database provides an interactive platform that allows quick retrieval of datasets of interest with pertinent analysis. Furthermore, the BDdb includes 869 manually curated biomarkers from six species, involving 22 birth-defects. In summary, the BDdb provides a comprehensive resource for researchers and clinicians.

Construction and content

Data collection

The datasets used to build the BDdb were collected from the Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>) database [14]. We first employed a list of keywords such as “trisomy 21”, “trisomy 13”, “trisomy 18”, “trisomy 8”, “monosomy X”, and “XXY” to retrieve pertinent information on *Homo sapiens* and *Mus musculus*, resulting in 328 GSE Series records (Fig. 1). All datasets were collected before August 2019. We then manually selected the relevant series records by careful reading of the retrieved data. Eventually, a total of 136 GSE Series records were selected and re-analyzed.

Single-cell RNA sequencing (scRNA-seq) holds tremendous potential for studying cell phenotype and cell behavior at single-cell resolution [15]. As such, we considered single-cell sequencing studies focusing on birth defects. We searched birth-disease related scRNA-seq datasets in humans and mice from the Sequence Read Archive (SRA) and GEO databases. We found one GSE

Series (GSE127257) record [16], which included 13,766 cells. This dataset focused on the Ts65Dn mouse model for Down syndrome.

To obtain putative biomarkers associated with different birth defects from published studies, we searched PubMed using keywords, including disease names with “marker(s)”, “gene(s)”, and “genetic(s)”. More than 3 000 publications were obtained. We manually surveyed the abstracts of these publications, and if they were considered to include pertinent information, we downloaded the paper and its supplementary materials. We then manually checked the full text of the selected publications to obtain biomarkers with reliable evidence and gain useful information, including the title of the paper, PMID, disease type, sequencing type, tissue type, cell type, and species. Among the final results, more than 500 relevant publications were included, and marker gene lists were manually extracted.

Data processing

For each dataset, we carefully read the original paper if available. When a dataset contained different karyotypes or models (or different types of sources, i.e., tissues and cell lines), we manually divided it into multiple sub-datasets in the BDdb (Fig. 1), with 326 (sub)-datasets thus generated (Additional file 1: Table S1).

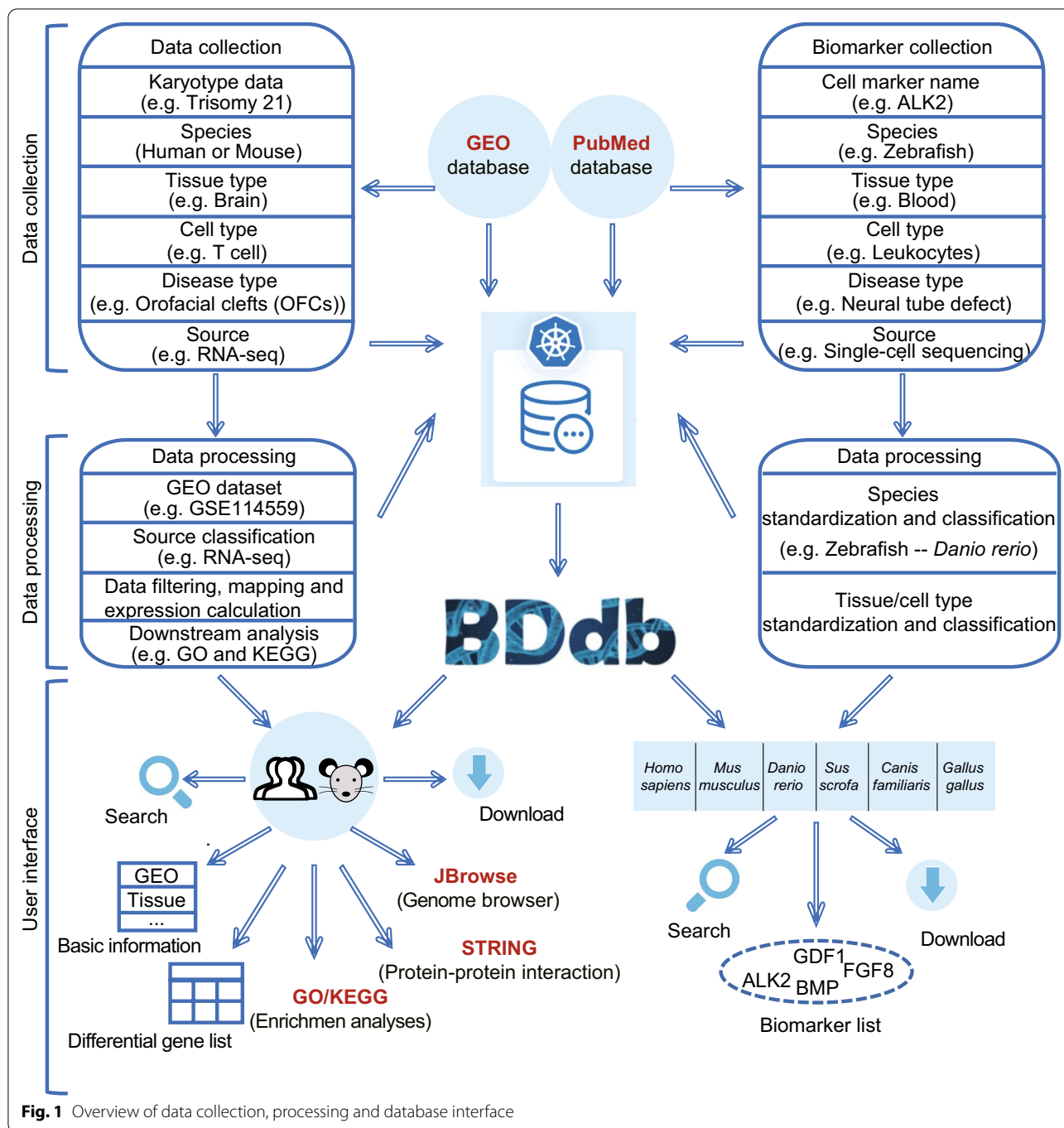
Raw data from next-generation sequencing and microarrays were downloaded. These data were re-analyzed using a uniform pipeline due to the lack of complete analysis results. Some microarray datasets were difficult to re-analyze or contained problems. Thus, we used the integrated GEO2R tool provided by GEO website to obtain a list of differentially expressed genes (DEGs). Detailed analyses are as follows.

RNA-Seq

Raw reads were filtered by SOAPnuke [17] with the parameters “-l 15 -q 0.2 -n 0.05”. Clean reads were then aligned to human (GRCh38.p12, downloaded from GENCODE [18]) or mouse (GRCm38.p6, downloaded from GENCODE [18]) reference genomes using HISAT2 [19]. StringTie [20] was used to compute gene expression. To screen DEGs, DESeq2 [21] was adopted for samples containing replicates; otherwise, DEGs were obtained using edgeR [22]. DEGs were screened based on absolute \log_2 -fold-change ≥ 0.5 , $p < 0.05$, and false discovery rate (FDR) < 0.1 . Gene Ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) annotation were performed using clusterProfiler [23].

DNA methylation

Low-quality reads from Bisulfite sequencing (BS-Seq) or reduced-representation bisulfite sequencing (RRBS-Seq)



were filtered with Trim Galore [24]. Subsequently, alignment of reads and extraction of methylation information were performed using Bismark [25]. Polymerase chain reaction (PCR) duplications were removed for the BS-Seq samples but not for RRBS-Seq sample as per the Bismark instructions. The R package methylKit [26] was used to extract the differential methylation regions among diverse samples.

DNA-protein interactions

An identical pipeline was adopted for analysis of chromatin immunoprecipitation sequencing (ChIP-Seq) and DNase I hypersensitive sites sequencing (DNase-Seq). SOAPnuke [17] was employed to remove low-quality reads, with the resulting clean reads then mapped against the human or mouse genome references using Bowtie2 [27]. Peak calling was performed using MACS2 [28, 29]

and peak annotation was accomplished using CHIP-seeker [30]. Lastly, DiffBind [31] was used to identify differential peaks.

Small RNA (smRNA) sequencing

Trim Galore [24] was adopted to trim the adaptors of the raw reads. Only smRNA reads that ranged in length from 18 to 30 nt were retained, with the parameters “-quality 20 -gzip -small_rna -max_length 30”. Eligible reads were then aligned against the human reference genome (GRCh38.p12, downloaded from GENCODE) and the miRBase [32] (release 22.1, <http://www.mirbase.org/>) to be annotated as miRNA, tRNA, rRNA, snoRNA, snRNA, piRNA, or other non-coding RNAs.

Microarray

For DNA microarray datasets with raw signal files, an in-house bioinformatics pipeline was adopted for quality control and expression quantification. Briefly, we read the *cel* file by using the R package *affy* (v1.8 release) [33], then, used the R packages *simpleaffy* [34] and *affyPLM* [35] to assess chips quality. Low-quality chips ($\text{actin3}/\text{actin5} > 3$, $\text{gapdh3}/\text{gapdh5} > 1.25$ and non-detected BioB) and *lowvar* probe sites were excluded. We next adopted the “rma” method for pretreatment. The expression abundance values of genes, including protein-coding genes (PCGs) and long non-coding RNAs (lncRNAs), were summarized using *t*-tests. Expression values were \log_2 transformed with an offset of 1. We resolved the up- and down-regulated genes, and then performed KEGG annotation and GO enrichment analysis with clusterProfiler [23]. Finally, we constructed a network of DEGs using STRING [36].

sci-ATAC-seq

We extracted corresponding metadata, including cell type and source (i.e., tissue, cell line). We also obtained the cell groups (referring to patient ID, culture condition or cell phenotype) from the supplementary tables of the papers and labeled each cell with the cell group information. Marker genes were then collected.

Website construction

The website was constructed using CodeIgniter, a powerful PHP framework. CodeIgniter provides an Application Programming Interface (API) for connecting the website to the MySQL database. We also used JavaScript libraries, including jQuery (3.4.1), jQuery-Labelauty, and additional plugins, to perform dynamic web services.

Data access

The BDdb website is accessible to users at <http://t21omics.cngb.org>, and offers a concise, well-organized

interface. Users are welcome to add comments regarding their requirements and suggestions to improve the database. All provided data can be downloaded. The BDdb will be continually updated with additional data from pertinent research. Figure 1 shows the overall features of the database.

Utility and discussion

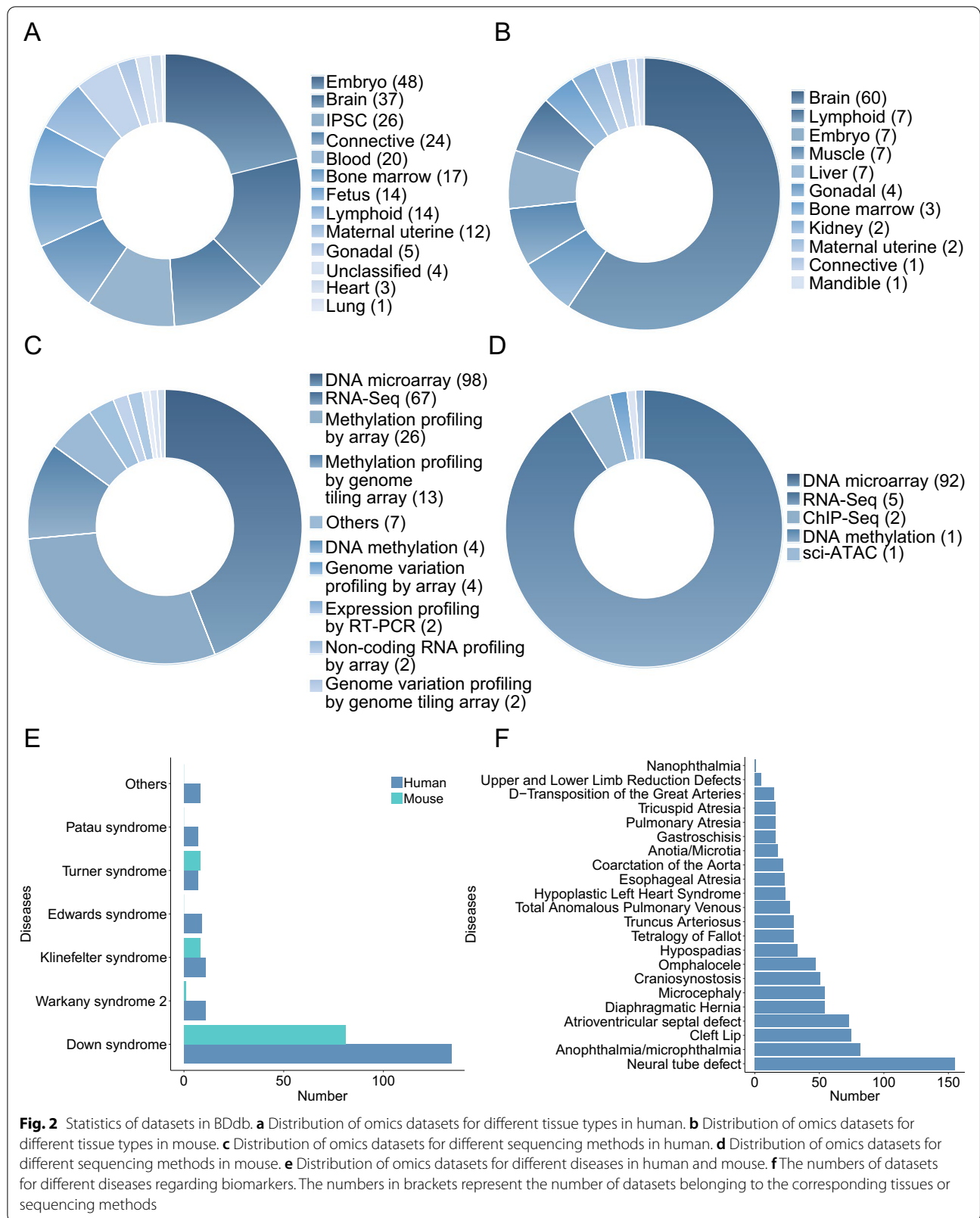
BDdb statistics

At present, the BDdb contains 101 and 37 GSE Series records from humans and mice, respectively. The obtained datasets include 15 diseases, 12 tissues, and 35 cell lines in humans, and four diseases, 12 tissues, and 17 cell lines in mice (Fig. 2a–e) obtained from multi-omics studies, e.g., genomics, transcriptomics, epigenomics, and single-cell omics (Fig. 2c, d). Moreover, the BDdb contains 869 potential biomarkers pertinent to 22 types of birth defects, such as microcephaly and neural tube defects (Fig. 2f). These markers were obtained from more than 500 studies involving six species, i.e., *Homo sapiens*, *Danio rerio*, *Mus musculus*, *Sus scrofa*, *Canis familiaris*, and *Gallus gallus*.

In the BDdb, embryonic tissues are among the most abundant samples for humans, followed by brain and blood; brain tissue samples are among the most abundant for mice, followed by lymphoid and embryonic samples (Fig. 2a, b). Correspondingly, the most abundant cell lines are blastocysts belonging to human embryo tissue and cortex belonging to mouse brain tissue. In terms of the diversity of sequencing types, DNA microarray data are dominant for both humans (44%) and mice (91%), followed by RNA-Seq, methylation profiling by array, ChIP-Seq, and DNA methylation (Fig. 2c, d). Most datasets are linked to Down syndrome, followed by Klinefelter syndrome, Turner syndrome, Warkany syndrome 2, and Edwards syndrome (Fig. 2e). In addition to datasets of diseases associated with chromosomal abnormalities, those related to diseases such as orofacial clefts, and open myelomeningocele are also included. For biomarkers collection, the top five diseases regarding related datasets include neural tube defects, anophthalmia/microphthalmia, cleft lip, atrioventricular septal defects, and diaphragmatic hernia (Fig. 2f). A summary of these datasets and biomarkers can be found in Additional files 1 and 2: Tables S1 and S2, respectively.

Database features and utility

The BDdb contains multi-omics datasets and allows users to query the subsequent analysis results with five functional states. The easy-to-use interface provides access for searching, browsing, visualizing, and downloading (Fig. 3). The online user guide illustrates several cases of BDdb usage.



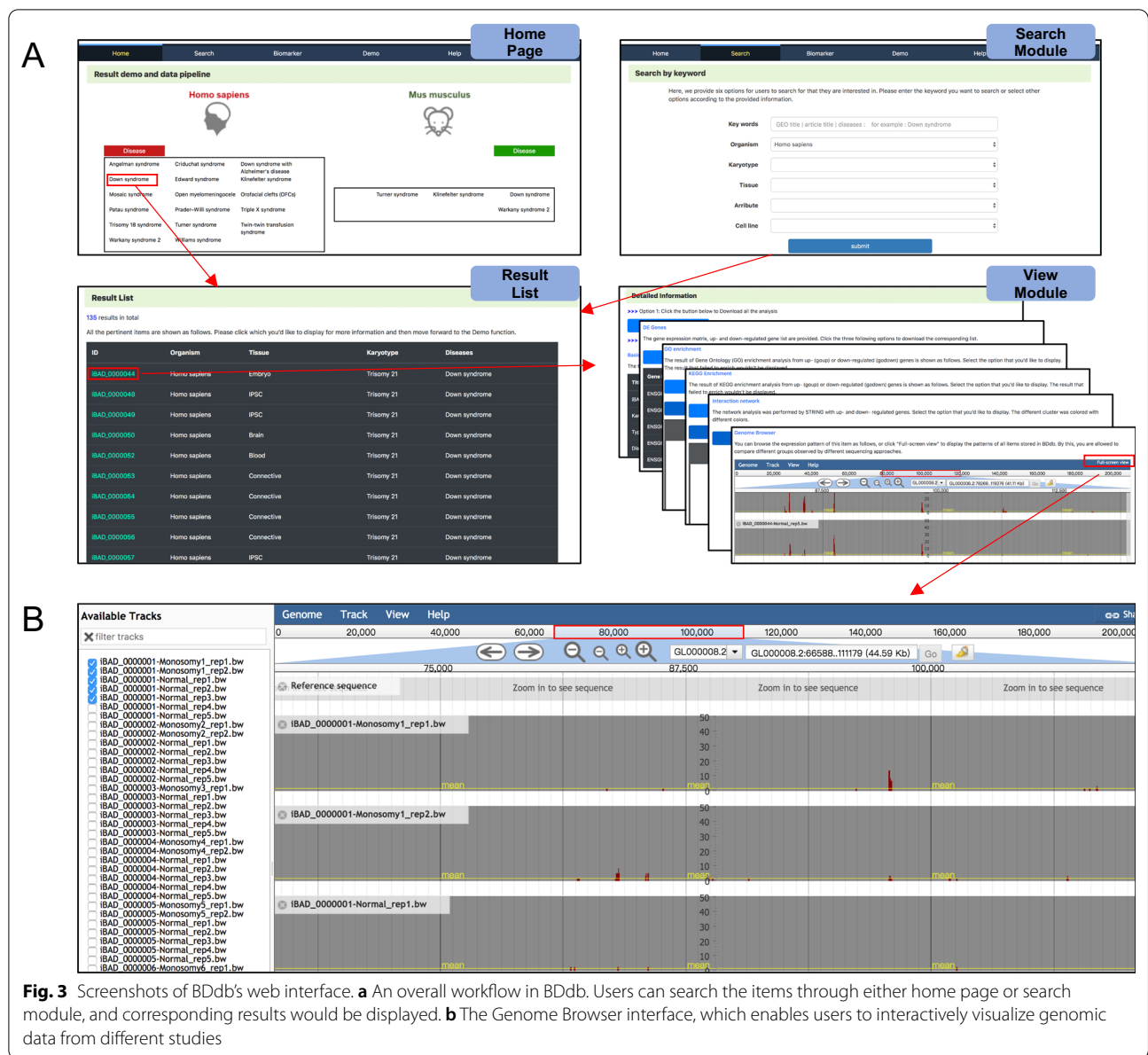


Fig. 3 Screenshots of BDdb's web interface. **a** An overall workflow in BDdb. Users can search the items through either home page or search module, and corresponding results would be displayed. **b** The Genome Browser interface, which enables users to interactively visualize genomic data from different studies

Information search

For the search module, users can search by inputting keywords or choosing the provided options. As shown in Fig. 3a, users can select one or more options, including the organism, karyotype, tissue, and cell line. This enables users to opt for certain kinds of karyotypes, such as trisomy 21, trisomy 18, and monosomy X, which are typical for chromosome aneuploidy. In terms of omics datasets, the BDdb only contains those related to humans and mice at present. After submitting the search request, relevant results are displayed.

View module

Data from diverse sequencing types are displayed with different result modules (Fig. 3a). For example, in the RNA-Seq

datasets, the resulting interface contains five sections: (1) “Basic Information”, shows the information on karyotype, disease, organism, tissue, and cell line, which are the sample’s features, as well as the GEO title, literature, and searching link, which can help users to trace the origin of the data; (2) “DE Genes”, enables users to search or download the gene expression matrix as well as up- and down-regulated gene tables; (3) GO and KEGG enrichment, enables users to explore functional/pathway enrichment, with the bubble charts, bar charts, and cnetplots provided; (4) Network analysis of DEGs; (5) “Genome Browser”, can intuitively display expression patterns with a graphical interface. Apart from the RNA-Seq datasets, the database also provides fundamental analysis for other datasets. Particularly,

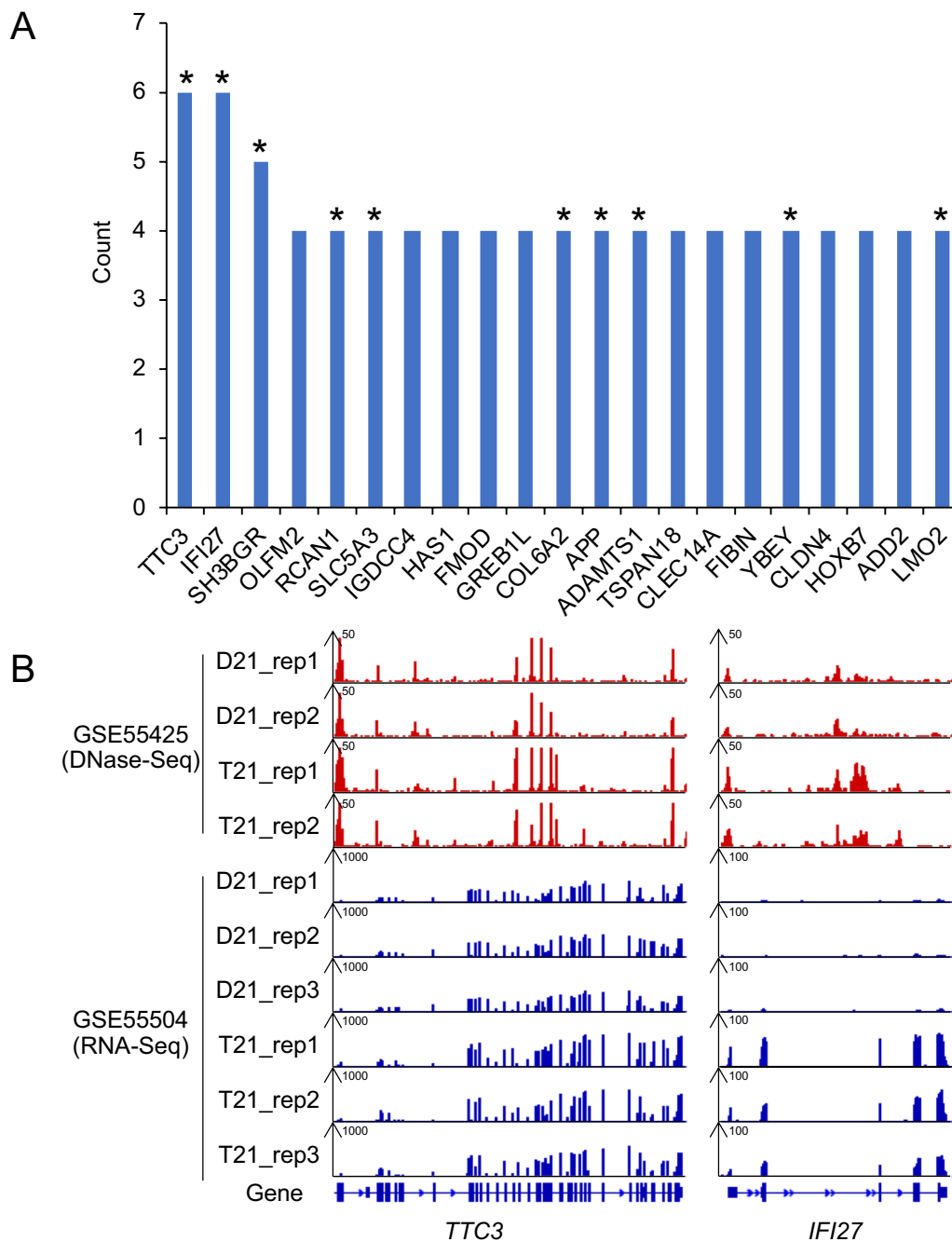


Fig. 4 Prospective biomarkers in down syndrome. **a** A total of 21 up-regulated genes counts more than 3 in fibroblasts from 8 GSE Series records, and half of which, marked with asterisk, have been reported as prospective biomarkers for trisomy 21. **b** Gene expression patterns of *TTC3* and *IFI27* gene observed from DNase-Seq data in GSE55425 and RNA-Seq data in GSE55504. T21 and D21 represent trisomy 21 and euploid controls, respectively

omics datasets from different studies can be displayed in the “Genome Browser” as per the user’s requirements to further mine for useful information (Fig. 3b).

Birth-defect diseases biomarker mapper

Biomarkers are stored in the “Biomarker” module. Users can search and view markers of interest by selecting species, diseases, and tissues from the pull-down menu.

Users can also download all analysis and biomarker results via the ‘Download’ function. The BDdb also provides detailed tutorials and answers to common questions on the “Help” page.

Case study: exploring biomarkers for diseases diagnosis using BDdb

To discover useful clues for diseases using the BDdb database, we targeted Down syndrome in humans, which has drawn considerable attention worldwide over many years. Taking fibroblasts as an example, a total of 13 GSE Series records were linked to trisomy 21, including various sequencing types such as RNA-Seq and DNase-Seq. We consolidated the up-regulated DEGs from eight GSE Series records obtained by RNA-Seq and DNA microarray, and then sorted them by counts. In total, 21 genes had counts ≥ 4 (Fig. 4a). Among them, the *TTC3* (*tetratricopeptide repeat domain 3*) and *IFI27* (*interferon α -inducible protein 27*) genes ranked first, with counts of six. *TTC3* is located on 21q22.2 within the Down syndrome critical region (DSCR) and plays an essential role in neural development. *TTC3* is commonly regarded as a candidate gene for Down syndrome and Alzheimer’s disease [37, 38]. In addition, *IFI27* is involved in the interferon response in trisomy 21 [39]. We found that both *TTC3* and *IFI27* had higher expression levels in trisomy 21 than in euploid controls in GSE55504. This was in accordance with the chromatin accessibility pattern in GSE55425 as assessed by DNase-Seq (Fig. 4b), implying that the extra copy of chromosome 21 or other transcript regulators in Down syndrome may confer this difference. In addition to *TTC3* and *IFI37*, other eight genes (marked with asterisks in Fig. 4a) such as *SH3BGR* [40, 41] and *APP* [42] are also typical biomarkers for Down syndrome. As these well-studied trisomy 21 marker genes can be captured by the BDdb, the rest that has not been reported yet, such as *OLFM2* and *HAS1*, may be prospective biomarkers for trisomy 21. Taken together, we can theoretically seek additional biomarkers associated with a particular disease using the BDdb.

Future perspectives

To assist clinicians and researchers, we developed the BDdb, which consists of multi-omics data and potential biomarkers of birth defects. The database will be updated constantly according to the frequency of publications associated with chromosomal aberrations. Aside from existing data, we will also add proteomics data to expand the repository. Moreover, other species and diseases will be added to provide more information to users. Ultimately, we hope that the BDdb, serving as an auxiliary tool, can provide clues for studies on birth defects, and hopefully, accelerate research progress.

Conclusions

The BDdb is a practical tool for those in the research community committed to studying birth defects. To the best of our knowledge, the BDdb is the first comprehensive database focusing on the collection of birth-defect multi-omics data, which were curated and re-analyzed before inclusion. Users can select data of interest to investigate related diseases, which can then be retrieved and downloaded. In addition, information on DEGs, enriched GO terms and KEGG pathways, and interaction networks are provided. Notably, users can also utilize the genome browser to compare diverse samples as well as the data from various sequencing methods, e.g., RNA-Seq, ChIP-Seq, and DNA methylation, and can further explore their correlations.

Abbreviations

BDdb: Birth defect multi-omics database; GEO: Gene expression omnibus; SRA: Sequence read archive; ScRNA-Seq: Single-cell RNA sequencing; BS-Seq: Bisulfite sequencing; RRBS-Seq: Reduced-representation bisulfite sequencing; ChIP-Seq: Chromatin immunoprecipitation sequencing; DNase-Seq: DNase I hypersensitive sites sequencing; smRNA: Small RNA; PCGs: Protein-coding genes; lncRNAs: Long non-coding RNAs; PCR: Polymerase chain reaction; GO: Gene ontology; KEGG: Kyoto encyclopedia of genes and genomes; DEGs: Differentially expressed genes.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12920-021-01110-x>.

Additional file 1: Table S1. The basic information of collected datasets.

Additional file 2: Table S2. The summary of biomarkers related to birth defects.

Acknowledgements

We would like to thank China National GeneBank and BGI Research.

Authors’ contributions

DC, HS, YG2, YS and FC designed the research study; DZ, SZ, ZZ, JH, SQ and SY performed data collection; DZ, XJ1 and ZZ analyzed the data; XJ1, DC, SY and XJ2 constructed the website; DZ, SZ, ZZ, DC and HS wrote the paper; DZ, SZ, ZZ, DC, HS, JH, SQ, YG1 and XZ revised the paper; XJ2, YG2, YS and FC performed project administration and acquired funding; All authors provided intellectual input for the study. All authors read and approved the final manuscript.

Funding

This work was supported by Guangdong Provincial Key Laboratory of Genome Read and Write (No. 2017B030301011), Shenzhen Engineering Laboratory for Innovative Molecular Diagnostics (DRC-SZ[2016]884), Shenzhen Municipal Government of China (JCYJ20180703093402288), Science, Technology and Innovation Commission of Shenzhen Municipality under Grant Nos. JCYJ20170412153100794, Key-Area Research and Development Program of Guangdong Province (2019B020227001). The funding bodies have no roles in the design of the study, the collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

The BDdb is freely available to all users at <http://t21omics.cngb.org>. This web portal is accessible by web browser. The datasets used and analysed during the current study are publically available on GEO database (<https://www.ncbi>).

nlm.nih.gov/geo/). Detailed information can be found in Additional files 1 and 2: Tables S1 and S2.

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

¹College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China. ²BGI-Shenzhen, Shenzhen 518083, People's Republic of China. ³Guangdong Provincial Key Laboratory of Genome Read and Write, BGI-Shenzhen, Shenzhen 518120, People's Republic of China. ⁴Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, People's Republic of China. ⁵BGI Genomics, BGI-Shenzhen, Shenzhen 518083, People's Republic of China. ⁶China National Genebank, BGI-Shenzhen, Shenzhen 518120, People's Republic of China. ⁷National Institutes for Food and Drug Control (NIFDC), Beijing 100050, People's Republic of China. ⁸Guangdong Provincial Key Laboratory of Human Disease Genomics, Shenzhen Key Laboratory of Genomics, BGI-Shenzhen, Shenzhen 518083, People's Republic of China. ⁹Shenzhen Engineering Laboratory for Birth Defects Screening, BGI-Shenzhen, Shenzhen 518083, People's Republic of China. ¹⁰MGI, BGI-Shenzhen, Shenzhen 518083, People's Republic of China.

Received: 20 September 2020 Accepted: 27 October 2021

Published online: 04 November 2021

References

- Czeizel AE. Birth defects are preventable. *Int J Med Sci*. 2005.
- Mannan M, Afroze S, Dey SK, Moni SC, Shabuj MKH, Jahan I, et al. Birth defect and its impact in neonatal health: a review. *Bangladesh J Child Heal*. 2019;43:49–58.
- Feldkamp ML, Carey JC, Byrne JLB, Krikov S, Botto LD. Etiology and clinical presentation of birth defects: population based study. *BMJ*. 2017;357:1–8.
- Xie D, Yang T, Liu Z, Wang H. Epidemiology of birth defects based on a birth defect surveillance system from 2005 to 2014 in Hunan Province, China. *PLoS ONE*. 2016.
- Shaffer L, Theisen. Disorders caused by chromosome abnormalities. *Appl Clin Genet*. 2010.
- Faas BHW, Cirigliano V, Bui TH. Rapid methods for targeted prenatal diagnosis of common chromosome aneuploidies. *Semin Fetal Neonatal Med*. 2011;16:81–7. <https://doi.org/10.1016/j.jsiny.2011.01.003>.
- Asim A, Kumar A, Muthuswamy S, Jain S, Agarwal S. "Down syndrome: an insight of the disease." *J Biomed Sci*. 2015.
- Rappaport N, Twik M, Plaschkes I, Nudel R, Stein TI, Levitt J, et al. MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. *Nucleic Acids Res*. 2017.
- Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, Ronzano F, Centeno E, Sanz F, et al. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res*. 2020.
- Peng K, Xu W, Zheng J, Huang K, Wang H, Tong J, et al. The disease and gene annotations (DGA): an annotation resource for human disease. *Nucleic Acids Res*. 2013.
- Badapanda C, Gupta H. RareDDB: an integrated catalog of rare disease database. *Clin Med Biochem Open Access*. 2016.
- Jiang Q, Wang Y, Hao Y, Juan L, Teng M, Zhang X, et al. miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res*. 2009.
- Wang Z, Zhang Q, Zhang W, Lin JR, Cai Y, Mitra J, et al. HEDD: human enhancer disease database. *Nucleic Acids Res*. 2018.
- Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: Archive for functional genomics data sets—update. *Nucleic Acids Res*. 2013.
- Yuan GC, Cai L, Elowitz M, Enver T, Fan G, Guo G, et al. Challenges and emerging directions in single-cell analysis. *Genome Biol*. 2017.
- Spektor R, Yang JW, Lee S, Soloway PD. Single cell ATAC-seq identifies broad changes in neuronal abundance and chromatin accessibility in down syndrome. *bioRxiv*. 2019.
- Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, et al. SOAPnuke: A MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience*. 2018.
- Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res*. 2019.
- Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015.
- Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014.
- Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2009.
- Yu G, Wang LG, Han Y, He QY. ClusterProfiler: An R package for comparing biological themes among gene clusters. *Omi A J Integr Biol*. 2012.
- Krueger F. Trim Galore!: a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. *Babraham Institute*. 2015.
- Krueger F, Andrews SR. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*. 2011.
- Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, et al. MethyKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol*. 2012.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012.
- Zhang Y, Liu T, Meyer CA, Eickhout J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008.
- Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. *Nat Protoc*. 2012.
- Yu G, Wang LG, He QY. ChIP seeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*. 2015.
- Stark R, Brown G. DiffBind: differential binding analysis of ChIP-Seq peak data. *Bioconductor*. 2011.
- Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: From microRNA sequences to function. *Nucleic Acids Res*. 2019.
- Gautier L, Cope L, Bolstad BM, Irizarry RA. Affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*. 2004.
- Wilson CL, Miller CJ. Simpleaffy: a bioconductor package for affymetrix quality control and data analysis. *Bioinformatics*. 2005.
- Brettschneider J, Bolstad B, Collin F, Speed T. Quality assessment for short oligonucleotide microarray data. *Technometrics*. 2008.
- Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019.
- Gong Y, Wang K, Xiao SP, Mi P, Li W, Shang Y, et al. Overexpressed TTC3 protein tends to be cleaved into fragments and form aggregates in the nucleus. *NeuroMolecular Med*. 2019;21:85–96. <https://doi.org/10.1007/s12017-018-8509-7>.
- Kohli MA, Cukier HN, Hamilton-Nelson KL, Rolati S, Kunkle BW, Whitehead PL, et al. Segregation of a rare TTC3 variant in an extended family with late-onset Alzheimer disease. *Neurol Genet*. 2016.
- Sullivan KD, Lewis HC, Hill AA, Pandey A, Jackson LP, Cabral JM, et al. Trisomy 21 consistently activates the interferon response. *Elife*. 2016;
- Sandri C, Di Lisi R, Picard A, Argentini C, Calabria E, Myklak K, et al. Heart morphogenesis is not affected by overexpression of the Sh3bgr gene mapping to the down syndrome heart critical region. *Hum Genet*. 2004;114:517–9.
- Vidal-Taboada JM, Bergoñon S, Scartezzini P, Egea A, Nizetic D, Oliva R. High-resolution physical map and identification of potentially regulatory

sequences of the human SH3BGR located in the Down syndrome chromosomal region. *Biochem Biophys Res Commun.* 1997;241:321–6.

42. Lott IT, Head E. Dementia in down syndrome: unique insights for Alzheimer disease research. *Nat Rev Neurol.* 2019;15:135–47. <https://doi.org/10.1038/s41582-018-0132-6>.

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

