OXFORD

Briefings in Bioinformatics, 2023, **24(2)**, 1–7
https://doi.org/10.1093/bib/bbad037
Advance access publication date 9 February 2023
Problem Solving Protocol

eccDNA Atlas: a comprehensive resource of eccDNA catalog

Tengwei Zhong[†], Wenqing Wang[†], Houyan Liu, Maolin Zeng, Xinyu Zhao and Zhiyun Guo

Corresponding author: Zhiyun Guo, School of Life Science and Engineering, Southwest Jiaotong University, Chengdu 610031, P.R. China. Tel: +86 13438192585; Fax: +86 28 87603202. E-mail: zhiyunguo@swjtu.edu.cn

[†]Tengwei Zhong and Wenqing Wang contributed equally.

Abstract

Extrachromosomal circular DNA (eccDNA) represents a large category of non-mitochondrial and non-plasmid circular extrachromosomal DNA, playing an indispensable role in various aspects such as tumorigenesis, immune responses. However, the information of characteristics and functions about eccDNA is fragmented, hiding behind abundant literatures and massive whole-genome sequencing (WGS) data, which has not been sufficiently used for the identification of eccDNAs. Therefore, establishing an integrated repository portal is essential for identifying and analyzing eccDNAs. Here, we developed eccDNA Atlas (http://lcbb.swjtu.edu.cn/eccDNAatlas), a user-friendly database of eccDNAs that aims to provide a high-quality and integrated resource for browsing, searching and analyzing eccDNAs from multiple species. eccDNA Atlas currently contains 629 987 eccDNAs and 8221 ecDNAs manually curated from literatures and 1105 ecDNAs predicted by AmpliconArchitect based on WGS data involved in 66 diseases, 57 tissues and 319 cell lines. The content of each eccDNA entry includes multiple aspects such as sequence, disease, function, characteristic, validation strategies. Furthermore, abundant annotations and analyzing utilities were provided to explore existing eccDNAs in eccDNA Atlas or user-defined eccDNAs including oncogenes, typical enhancers, super enhancers, CTCF-binding sites, SNPs, chromatin accessibility, eQTLs, gene expression, survival and genome visualization. Overall, eccDNA Atlas provides an integrated eccDNA data warehouse and serves as an important tool for future research.

Keywords: extrachromosomal circular DNA, extrachromosomal DNA, database, enhancer

Introduction

Extrachromosomal circular DNA (eccDNA) is a non-mitochondrial and non-plasmid circular structured DNA that is widely found in eukaryotes [1]. Up to now, eccDNAs with different nomenclatures have been discovered in a variety of species such as ecDNAs (previously referred to as double minutes) [2], small polydispersed circular DNAs (spcDNAs) [3], microDNAs [4], telomeric circles or t-circles [5, 6], extrachromosomal ribosomal circles (ERC) [7]. Among them, ecDNA has been considered as the most crucial type of eccDNA for wide study. EcDNA (extrachromosomal DNA), which is also referred to as the long eccDNA (168 Kb ∼ 5 Mb, with an average of 1.26 Mb), mainly exists in tumor and can be observed in the metaphase of mitosis [8]. EcDNA appears as monomers or clusters [9], and usually carries a variety of oncogenes such as *DHFR*, *EGFR*, *MYC* [2, 8]. EcDNAs are vehicles for oncogene amplification and randomly segregated to daughter cells. They can promote drug resistance, tumor heterogeneity and even drive poor outcome for patients [2, 10–12]. In addition, the other types of eccDNA except ecDNA (hereinafter referred to as eccDNA) have also been confirmed to have crucial effects on eukaryotes. Unlike ecDNA, which is only present in tumors, eccDNA can be widely found in all tested cell types including healthy human cells and diseased cells [10]. Although eccDNA is generally smaller than ecDNA, studies have shown that eccDNAs also sometimes carry and express genes or miRNAs to promote tumorigenesis [13]. Previous studies showed that eccDNAs could be released into circulation such as plasma or serum, suggesting that they may serve as biomarkers for diseases [14]. Furthermore, eccDNAs with smaller size can function as potent innate immunostimulants in a manner that is independent of the eccDNA sequence but dependent on eccDNA circularity and the cytosolic DNA sensor sting [15].

How to accurately identify and characterize eccDNAs/ecDNAs is the key point to study their functions. Previously, diverse methods have been developed to perform relevant researches including sedimentation in a sucrose gradient, electron microscopy [16], light microscopy [17], southern blotting [18], probe hybridization, florescence *in situ* hybridization (FISH) [19] and 2D gel electrophoresis [20]. In recent years, next-generation sequencing (NGS) technologies have been developed, especially the improvement of whole-genome sequencing (WGS) technology leads to the emergence of new methods for the identification of eccDNAs/ecDNAs. For example, AmpliconArchitect is a powerful tool that can be applied for predicting ecDNA focally amplified

Tengwei Zhong is a master student at School of Life Science and Engineering, Southwest Jiaotong University, Chengdu, China.
Wenqing Wang is a master student at School of Life Science and Engineering, Southwest Jiaotong University, Chengdu, China.
Houyan Liu is a master student at School of Life Science and Engineering, Southwest Jiaotong University, Chengdu, China.
Maolin Zeng is an undergraduate at School of Life Science and Engineering, Southwest Jiaotong University, Chengdu, China.
Xinyu Zhao is an undergraduate at School of Life Science and Engineering, Southwest Jiaotong University, Chengdu, China.
Zhiyun Guo is an associate professor at School of Life Science and Engineering, Southwest Jiaotong University, Chengdu, China. His research interests are focused on transcriptomics and genomics.
Received: August 24, 2022. **Revised:** January 13, 2023. **Accepted:** January 18, 2023

regions based on WGS data [21]. Moreover, another widely used method called Circle-Seq has been developed for the isolation and sequencing of eccDNA on a genomic scale [22]. The emergence of these methods mentioned above has generated vast amounts of eccDNA data and greatly promoted the research of eccDNA.

Nowadays, lots of wet and dry experiments have been performed to identify and clarify the functions of eccDNAs/ecDNAs across multiple species, especially after the completion of the TCGA and pan-genome projects [23]. The information of characteristics and functions about eccDNAs/ecDNAs is fragmented and hides behind almost countless literatures. Besides, a large number of WGS data was released by numerous studies. Therefore, integrating the above resources to build a database to store eccDNAs/ecDNAs data is very important for the follow-up study of eccDNAs/ecDNAs. To date, only two eccDNA databases, CircleBase and eccDNAdb, have been developed to explore this information. CircleBase [24] extracted eccDNAs from 13 literatures and performed six annotations based on the genomic and epigenetics data of eccDNAs. eccDNAdb [25] identified human eccDNAs by AmpliconArchitect from WGS data (131 samples from Turner *et al.* study and 60 tumor samples from SRA). However, these two databases only included eccDNAs from human species. Therefore, it is necessary to integrate existing eccDNA resources from multiple species in literatures as much as possible and to predict new eccDNAs based on the available WGS data for the subsequent research of eccDNA.

In this study, we developed a database called eccDNA Atlas, to extend resources of eccDNAs/ecDNAs. We extracted eccDNAs/ecDNAs from 3636 literatures using eccDNAs/ecDNAs keywords and finally obtained 638 208 eccDNAs (8221 ecDNAs) from seven species. Furthermore, we downloaded WGS data of 319 tumor samples for identifying ecDNAs by AmpliconArchitect. When the literatures have already predicted eccDNAs/ecDNAs by WGS, we collected these eccDNAs/ecDNAs from these literatures directly instead of making our own predictions (e.g. 131 samples with WGS data from Turner *et al.* study used by the eccDNAdb database). Finally, 1105 ecDNAs were identified. Besides, we obtained oncogenes, typical enhancers, super enhancers, CTCF-binding sites, SNPs, chromatin accessibility, eQTLs on these eccDNAs/ecDNAs and provided utilities for BLAST analysis and genome visualization. In addition, a user-friendly web interface was built and split into seven main pages: (i) Browse, (ii) Search, (iii) Analysis, (iv) Download, (v) Statistics, (vi) Submit and (vii) Help.

Materials and methods

Data collection

Data collection from literature

Due to the confusion of eccDNA nomenclatures, the eccDNA names appearing in literatures are very ambiguous. Therefore, we manually reviewed all eccDNA aliases and types names (full name or abbreviation) of eccDNA in literatures as follows: 'ecDNA', 'extrachromosomal DNA', 'eccDNA', 'extrachromosomal circular DNA', 'microDNA', 'spcDNA', 'extrachromosomal ribosomal circle' and 'telomeric circles/t-circles'. To ensure the highest quality in data collection process, all eccDNA entries were manually collected by the following steps: (i) a list of keywords with restrictions were searched in the PubMed. For example, the search keywords for ecDNA was '("extrachromosomal DNA" * [Title/Abstract], "ecDNA", "extrachromosomal DNA")', and the search keywords for eccDNA was '("eccDNA" * [Title/Abstract], "eccDNA", "extrachromosomal circular DNA", "microDNA", "spcDNA", "t-circles", "telomeric circles", "extrachromosomal ribosomal circles")'. Therefore, 7212 literatures were obtained through the above search strategy. (ii) After manually removing the duplicates and review literatures, 3636 literatures as the final literatures were extracted for further eccDNA collection. (iii) By manually curating from 3636 literatures, we found that there are a large number of ecDNAs and eccDNAs classified into plasmid DNA/telomeres. Finally, 814 optional literatures were selected for eccDNA data extraction according to the following information catalog: (i) species and eccDNA type; (ii) chromosomal localization of eccDNAs/ecDNAs; (iii) sample condition and disease; (iv) validation strategies and sequencing library types; (v) the function/characteristic of eccDNAs/ecDNAs; (vi) PMID and publish date; and (vii) Other information. All information was double checked and the names of cell lines were also checked according to the nomenclature of cell lines.

Prediction of ecDNA based on WGS data

To supplement the ecDNAs from existing literatures, we predicted the ecDNAs from available WGS data. WGS data of 319 tumor samples were downloaded from the SRA database (Supplementary Table S1). AmpliconArchitect was employed for identification of eccDNAs/ecDNAs amplicons according following criteria: genomic segments >10 kb with copy numbers (CNs) >4. Finally, the predicted amplicons were classified using an AA classifier. Figure 1 shows the flowchart of the database construction process.

ecDNA/eccDNA annotation and visualization

All the eccDNA/ecDNA sequences collected by eccDNA Atlas were converted into default unified version using the UCSC LiftOver tool [26]. The R package (biomartR) was used to extract genomic location of oncogene, and a total of 803 oncogenes were downloaded from ONGene [27] (http://www.ongene.bioinfominzhao.org/index.html). A total of 116 941 typical enhancers were downloaded from ENdb [28] and EnhancerDB [29] databases, and 336 763 super enhancers were downloaded from dbSUPER [30], SEA [31] and SEdb [32] databases. A total of 11 499 254 CTCF-binding sites and 211 948 SNPs were downloaded from GREAP [33]. A total of 15 132 567 chromatin accessibility data were downloaded from ATACdb [34] and GREAP [33] databases, 341 316 eQTL data were downloaded from the OncoBase [35] database. If a regulatory element was entirely located within the eccDNA/ecDNA region, it was annotated as an eccDNA/ecDNA regulatory element. For the convenience of users, in addition to providing the ecDNA annotations already in the database, we also allow users to input customized chromosome positions for annotation. UCSC browser was used to provide visualization annotation of current eccDNA/ecDNA region and user-defined eccDNA/ecDNA region, such as histone modification signals, mutations, sequence conservation. The external links of GEPIA [36], GeneCard and String [37] databases were provided for gene expression, function, survival, regulatory network exploring, etc. BLAST was built based on the ViroBLAST [38] online tool. MakeBlastdb was used to construct nucleic acid sequence libraries for five species of eccDNA Atlas.

System design and implementation

The eccDNA Atlas website runs on an Nginx Web server (https://nginx.org/). The database was developed using MySQL 5.7.27 (http://www.mysql.com). PHP7.2.30 (http://www.php.net) was used for server-side scripting. The eccDNA Atlas web interface was built using Bootstrap v3.3.7 (https://v3.bootcss.com) and JQuery v2.1.1 (http://jquery.com). ECharts(http://echarts.baidu.com) was

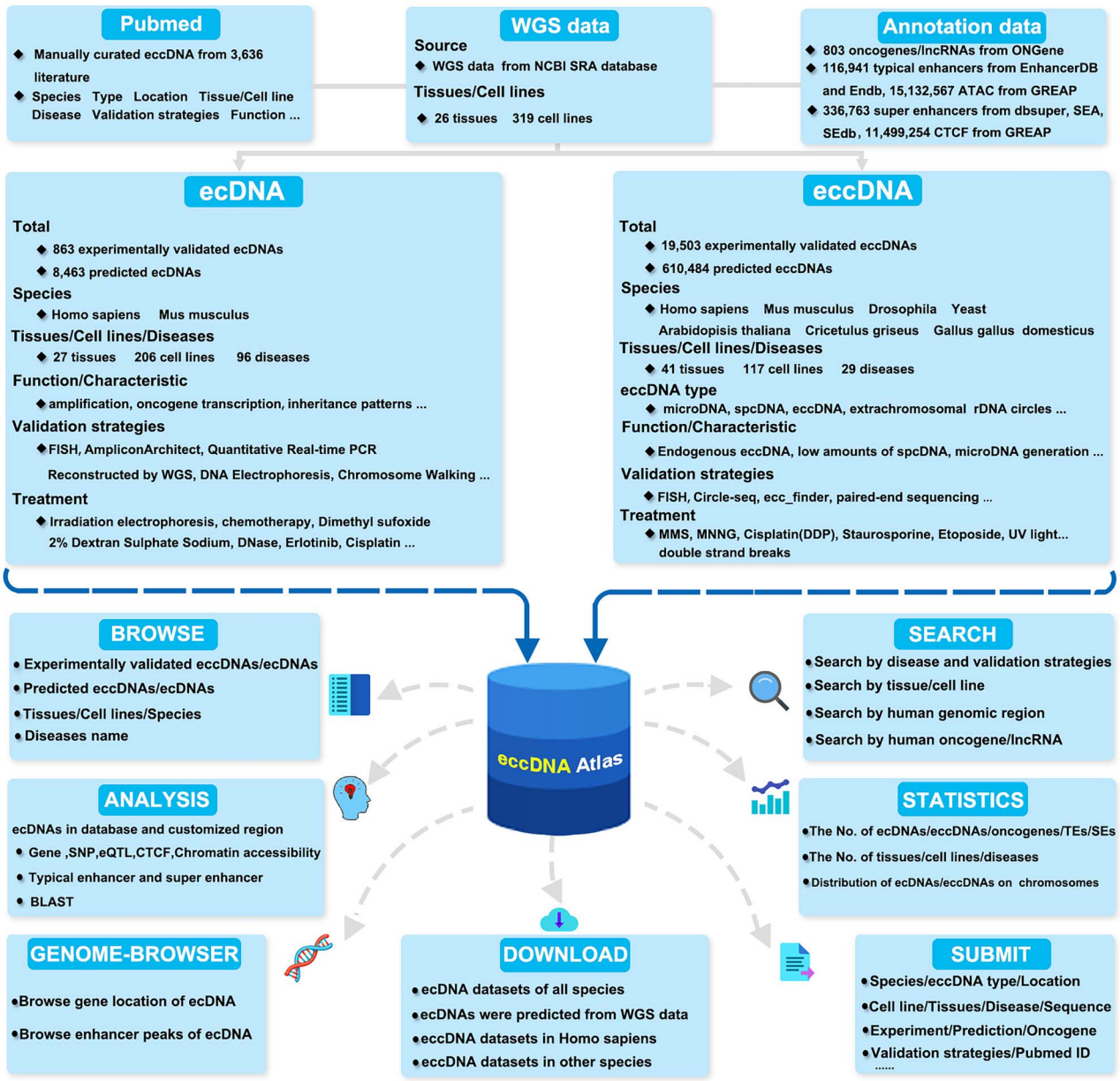

Figure 1. A schematic workflow of eccDNA Atlas.

used as a graphical visualization framework. We recommend using the latest versions of Firefox and Google Chrome for the best experience.

Results

Database statistics

Currently, there are two types of data involved in eccDNA Atlas: (i) 629 987 eccDNAs and 8221 ecDNAs manually curated from literatures; and (ii) 1105 ecDNAs identified by AmpliconArchitect based on public WGS data. The detailed statistics were listed in the Table 1.

Web interface and usage

We developed a user-friendly web interface to help users to browse, search, analysis, download and submit the eccDNA/ecDNA data. The web interface was split into the following seven main pages: (i) Browse, (ii) Search, (iii) Analysis, (iv) Download, (v) Statistics, (vi) Submit and (vii) Help (Figure 2A).

Browse

The browse page was divided into two menus: 'browse by ecDNA' and 'browse by eccDNA' (Figure 2B). The 'browse by ecDNA' page includes 'manually curated from literatures' and 'predicted from WGS data'. To make browsing easier for users, we classified the cell lines according to affiliated tissue source. Users can browse the ecDNAs by clicking the labels matching them up with the tissue type on schematic diagram or by choosing from the drop-down menu of tissues in the corresponding species catalog. In the eccDNA part, users can preferentially select species (e.g. Drosophila, Yeast and *Mus musculus*) to browse eccDNA data, or choose the corresponding tissue to browse.

Search

The search page was divided into two major parts: 'search by ecDNA' and 'search by eccDNA'. Our search page of database is composed of four entries (Figure 2C–E): (i) 'Search by disease and validation strategies' for searching eccDNA/ecDNA of

Table 1. Statistics of eccDNAs/ecDNAs in eccDNA Atlas

Type	eccDNA	ecDNA
Curated from literatures	629 987	8221
Identified by WGS	414 795	8293
Species	7	2
Normal sample	191 138	–
Disease sample	438 849	9326
Cell line	571 078 (from 117 cell lines)	1980 (from 206 cell lines)
Tissue	58 909 (from 41 tissue)	7346 (from 27 tissue)
Disease	29	96
Experimentally validated	19 503	863
Prediction	610 484	8463
Oncogene	680	645

disease from human and mouse; (ii) 'Search by cell line/tissue' for searching the eccDNA/ecDNA from disease/health sample of all species collected by eccDNA Atlas; (iii) 'Search by human genomic region' for searching chromosome location of interest by user of all species we collected; (iv) 'Search by human oncogene/lncRNA' for searching human oncogene/lncRNA of ecDNA. For each search section, there is a step-by-step guide to ensure that the search process is user-friendly. For example, for 'search by disease and validation strategies', the first step is to choose a species type. Secondly, a disease needs to be selected to refine the further search. Based on the search results of steps 1 and 2, experiment/prediction and validation strategies can be selected as an option (steps 3 and 4, respectively) for a more advanced search. Furthermore, we also provided genome visualization utilities for ecDNA based on UCSC Genome Browser. Users can select an ecDNA list of eccDNA Atlas or enter the user-defined ecDNA chromosomal coordinate range of ecDNA in the genome browse section, and get a general overview of the information contained in the ecDNA region such as histone modification signals, mutations and sequence conservation (Figure 2F).

Analysis

Users can perform annotation analysis on ecDNA regions included in eccDNA Atlas or custom chromatin regions including oncogenes/lncRNAs, typical enhancers, super enhancers, CTCF-binding sites, SNPs, chromatin accessibility and eQTLs. Furthermore, we provided some external links for gene expressions, functions, survival and regulatory network exploring analysis, including GEPIA, GeneCard and String databases. In addition, in order to help users to find regions of local similarity between user interested sequences and eccDNA sequences of eccDNA Atlas, we provided the BLAST tool in the 'ANALYSIS' module of the web page. Users can perform BLAST against local eccDNA database of an assigned species and set filtering parameters for an advanced search (Figure 2G).

Download, submit and statistics

Users can download all eccDNAs/ecDNAs and annotation data from eccDNA Atlas (Figure 2H) or share their research data (e.g. location, sequence) through submitting interface (Figure 2I). It should be noted that the data submitted by users will not be online until strict manual check possess is completed. On the statistics page, we mainly classified and counted the eccDNAs/ecDNAs in each chromosome, tissue, disease, etc. A chromosome distribution diagram was provided to help users to browse the ecDNAs/eccDNAs in a particular chromosome that they are interested in (Figure 2J).

A case study

Take ec_hsa_6592 from human prostate cancer as an example to introduce the use of eccDNA Atlas. In the browse page, users can click the prostate label to get the data containing ec_hsa_6592. In the search page, for 'search by disease and validation strategies', users can first select the species type as '*Homo sapiens*' and secondly select the disease type as 'Prostate Cancer' to get the data containing ec_hsa_6592. The result page contains the information including eccDNA ID, species, eccDNA type, location, health/disease, tissue/cell line, disease name, experiment/prediction, validation strategies and pubmed ID. Users can sort the results on the header of each row. In addition, filter function is also available in the 'Search' box on the upper left corner of the result page. Then users can click the eccDNA ID for details including eccDNA length, eccDNA sequence, eccDNA coordinate, publication date, treatment, function/characteristic, remarks and data source. When an ecDNA consisted of fragments from multiple chromosomes, we added the remarks on the detail page to declare that this ecDNA is made up of multiple chromosome segments. Two frequently amplified oncogenes (*MYC* and *PVT1*) in human prostate cancer were annotated on ec_hsa_6592 by performing an analysis of eccDNA Atlas, which are consistent with the previous study [13]. The results showed that *PVT1* had a significant effect on patient survival in prostate cancer, and there was a relatively significant positive correlation between *MYC* and *PVT1* ($R = 0.32$, $P = 1.4e\text{-}12$). In addition, 10 typical enhancers and two super enhancers were annotated in this region, and most of the enhancers had strong H3K27Ac signal peaks by genome-browser visualization of eccDNA Atlas. In summary, this information obtained from the database has important implications for the subsequent studies of the potential functions and regulatory ability of eccDNAs/ecDNAs.

Discussion

Although EccDNA was discovered nearly 60 years ago, the eccDNA/ecDNA sequence information is insufficient in earlier research literatures due to technical limitations. By the application of high-throughput DNA sequencing techniques, especially NGS, it becomes possible to identify more and more eccDNAs. We collected existing literatures containing eccDNA/ecDNA information as much as possible by manually curated 3636 literatures. Some publications only found eccDNA (e.g. FISH) in cell/tissue but they did not perform DNA sequencing, therefore, there is no corresponding eccDNA location information. For such data, we did not deposit the location information of eccDNA but still remained the other existing information such

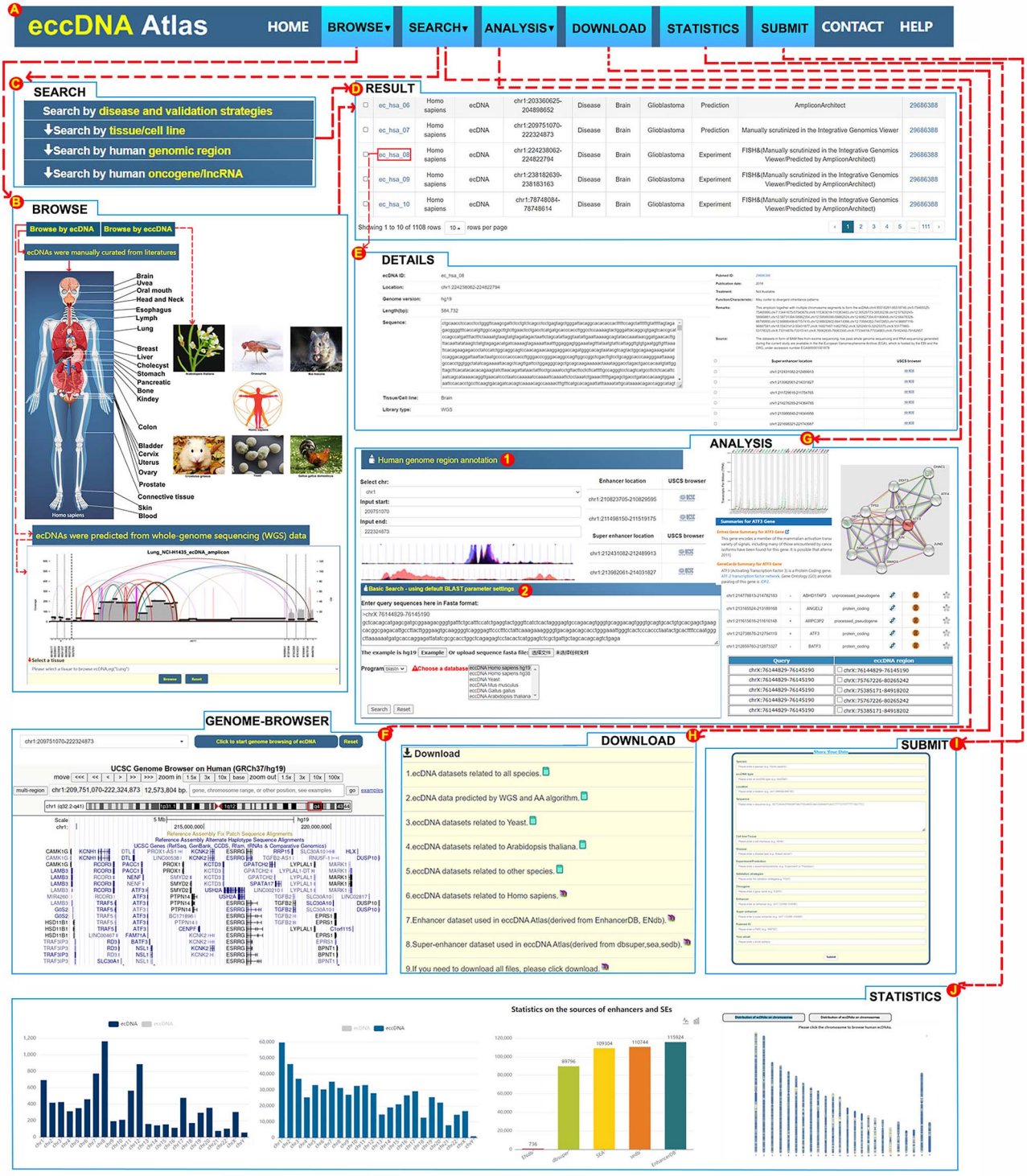

Figure 2. The web interface and usage of eccDNA Atlas. (**A**) The navigation bar of eccDNA Atlas. (**B**) Users can browse ecDNA/eccDNA data by species, tissue and data source (literature or WGS). (**C**) Users can query ecDNA/eccDNA data through four ways: 'Search by disease and validation strategies', 'Search by tissue/cell line', 'Search by human genomic region' and 'Search by human oncogene/lncRNA'. (**D**) Results table includes the following information: eccDNA ID, species, eccDNA type, location, health/disease, tissue/cell line, disease name, experiment/prediction, validation strategies and PubMed ID. (**E**) Detail page of each eccDNA/ecDNA includes the following information: publication date, treatment, function/characteristic, remarks and data source. (**F**) Visualization annotation of database/user-defined eccDNA/ecDNA region. (**G**) The analysis module of eccDNA Atlas including functional annotation, BLAST, gene expression, survival and regulatory network. (**H**) Data download page. (**I**) Data submit page. (**J**) Statistics of eccDNA Atlas.

as formation mechanism and function. As the most important class of eccDNAs, the mechanisms of ecDNA biogenesis are still unclear, and it is difficult to identify the high amplification region of ecDNA sequence with traditional experimental methods such as DNA electrophoresis, FISH. In order to help researchers to study ecDNA more deeply, we collected massive WGS data from 319 cell lines by NCBI SRA and predicted 1105 ecDNAs from 150 cell lines by AmpliconArchitect. Furthermore, we included a lot of annotation information for these ecDNAs such as oncogenes, lncRNAs, typical enhancers, super enhancers, SNPs. Nowadays,

there are several methods can distinguish eccDNAs from linear chromosomes such as FISH, Circle-Seq [22] and CRISPR-CATCH [39]. However, there are still many issues needed to be solved for accurate and comprehensive annotations of ecDNAs instead of linear chromosomes, as currently most available multi-omics sequencing data (e.g. RNA-seq and ATAC-seq) are produced from samples in which eccDNAs are mixed with linear chromosomes.

In the future, we will continue to follow up on eccDNAs/ecDNAs related literatures. With the further data yielded for eccDNAs/ecDNAs, we will continuously collect the latest datasets to keep our database up-to-date. We will develop new database analysis functions and enhance the interactivity of the database to give users a better experience. In addition, with the advancement of experimental methods, it is expected to collect more sequencing data performed for eccDNAs instead of linear chromosomes, such as genomic interaction data between eccDNAs or eccDNA hubs, survival data and epigenetic modification data. The database will be updated and backed up during the user's off-peak hours at regular intervals. During major maintenance, users will be given a temporary entry to continue using the eccDNA Atlas normally. Overall, eccDNA Atlas is a comprehensive data portal of eccDNAs/ecDNAs for health and diseases across multiple species.

Key points

- In all, 629 987 eccDNAs and 8221 ecDNAs were manually curated from 3636 literatures across multiple species.
- In all, 1105 ecDNAs were identified by AmpliconArchitect algorithm based on WGS data from 319 tumor samples.
- A user-friendly web interface was built to help users to search, browse, analyze, download and submit data.
- Customized eccDNA annotation, analysis and genome visualization were provided for oncogenes, enhancers, SNPs, chromatin accessibility, etc.

Availability

The data are freely accessible for research purposes at http://lcbb.swjtu.edu.cn/eccDNAatlas/index.html.

Supplementary data

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

Acknowledgements

We thank the Informatization and Network Management Office of Southwest Jiaotong University for providing the cloud-based technical support.

Funding

The Sichuan Science and Technology Program (Grant 2022NSFSC0779) and Basic Research Cultivation Support Program of Fundamental Research Funds for the Central Universities (2682021ZTPY016).

References

1. Noer JB, Hørsdal OK, Xiang X, *et al.* Extrachromosomal circular DNA in cancer: history, current knowledge, and methods. *Trends Genet TIG* 2022;**38**:766–81.
2. Turner KM, Deshpande V, Beyter D, *et al.* Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature* 2017;**543**:122–5.
3. Smith CA, Vinograd J. Small polydisperse circular DNA of HeLa cells. *J Mol Biol* 1972;**69**:163–78.
4. Shibata Y, Kumar P, Layer R, *et al.* Extrachromosomal microDNAs and chromosomal microdeletions in normal tissues. *Science* 2012;**336**:82–6.
5. Tomaska L, Nosek J, Makhov AM, *et al.* Extragenomic double-stranded DNA circles in yeast with linear mitochondrial genomes: potential involvement in telomere maintenance. *Nucleic Acids Res* 2000;**28**:4479–87.
6. Cesare AJ, Griffith JD. Telomeric DNA in ALT cells is characterized by free telomeric circles and heterogeneous t-loops. *Mol Cell Biol* 2004;**24**:9948–57.
7. Sinclair DA, Guarente L. Extrachromosomal rDNA circles—a cause of aging in yeast. *Cell* 1997;**91**:1033–42.
8. Verhaak RGW, Bafna V, Mischel PS. Extrachromosomal oncogene amplification in tumour pathogenesis and evolution. *Nat Rev Cancer* 2019;**19**:283–8.
9. Hung KL, Yost KE, Xie L, *et al.* ecDNA hubs drive cooperative intermolecular oncogene expression. *Nature* 2021;**600**:731–6.
10. Møller HD, Mohiyuddin M, Prada-Luengo I, *et al.* Circular DNA elements of chromosomal origin are common in healthy human somatic tissue. *Nat Commun* 2018;**9**:1069.
11. Vogt N, Gibaud A, Lemoine F, *et al.* Amplicon rearrangements during the extrachromosomal and intrachromosomal amplification process in a glioma. *Nucleic Acids Res* 2014;**42**:13194–205.
12. Koche RP, Rodriguez-Fos E, Helmsauer K, *et al.* Extrachromosomal circular DNA drives oncogenic genome remodeling in neuroblastoma. *Nat Genet* 2020;**52**:29–34.
13. Paulsen T, Shibata Y, Kumar P, *et al.* Small extrachromosomal circular DNAs, microDNA, produce short regulatory RNAs that suppress gene expression independent of canonical promoters. *Nucleic Acids Res* 2019;**47**:4586–96.
14. Kumar P, Dillon LW, Shibata Y, *et al.* Normal and cancerous tissues release extrachromosomal circular DNA (eccDNA) into the circulation. *Mol Cancer Res* 2017;**15**:1197–205.
15. Wang Y, Wang M, Djekidel MN, *et al.* eccDNAs are apoptotic products with high innate immunostimulatory activity. *Nature* 2021;**599**:308–14.
16. Hotta Y, Bassel A. Molecular size and circularity of DNA in cells of mammals and higher plants. *Proc Natl Acad Sci USA* 1965;**53**:356–62.
17. Cox D, Yuncken C, Spriggs AI. Minute chromatin bodies in malignant tumours of childhood. *Lancet Lond Engl* 1965;**1**:55–8.
18. Stanfield SW, Helinski DR. Cloning and characterization of small circular DNA from Chinese hamster ovary cells. *Mol Cell Biol* 1984;**4**:173–80.
19. Shapiro DN, Valentine MB, Rowe ST, *et al.* Detection of N-myc gene amplification by fluorescence in situ hybridization. Diagnostic utility for neuroblastoma. *Am J Pathol* 1993;**142**:1339–46.
20. Cohen S, Lavi S. Induction of circles of heterogeneous sizes in carcinogen-treated cells: two-dimensional gel analysis of circular DNA molecules. *Mol Cell Biol* 1996;**16**:2002–14.
21. Deshpande V, Luebeck J, Nguyen N-PD, *et al.* Exploring the landscape of focal amplifications in cancer using AmpliconArchitect. *Nat Commun* 2019;**10**:392.
22. Møller HD, Parsons L, Jørgensen TS, *et al.* Extrachromosomal circular DNA is common in yeast. *Proc Natl Acad Sci USA* 2015;**112**:E3114–22.

23. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature* 2020;**578**:82–93.
24. Zhao X, Shi L, Ruan S, *et al.* CircleBase: an integrated resource and analysis platform for human eccDNAs. *Nucleic Acids Res* 2022;**50**:D72–82.
25. Peng L, Zhou N, Zhang C-Y, *et al.* eccDNAdb: a database of extrachromosomal circular DNA profiles in human cancers. *Oncogene* 2022;**41**:2696–705.
26. Rosenbloom KR, Armstrong J, Barber GP, *et al.* The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res* 2015;**43**:D670–81.
27. Liu Y, Sun J, Zhao M. ONGene: a literature-based database for human oncogenes. *J Genet Genomics* 2017;**44**:119–21.
28. Bai X, Shi S, Ai B, *et al.* ENdb: a manually curated database of experimentally supported enhancers for human and mouse. *Nucleic Acids Res* 2020;**48**:D51–7.
29. Kang R, Zhang Y, Huang Q, *et al.* EnhancerDB: a resource of transcriptional regulation in the context of enhancers. *Database (Oxford)* 2019;**2019**.
30. Khan A, Zhang X. dbSUPER: a database of super-enhancers in mouse and human genome. *Nucleic Acids Res* 2016;**44**:D164–71.
31. Chen C, Zhou D, Gu Y, *et al.* SEA version 3.0: a comprehensive extension and update of the super-enhancer archive. *Nucleic Acids Res* 2020;**48**:D198–203.
32. Jiang Y, Qian F, Bai X, *et al.* SEdb: a comprehensive human super-enhancer database. *Nucleic Acids Res* 2019;**47**:D235–43.
33. Yang Y, Qian F, Li X, *et al.* GREAP: a comprehensive enrichment analysis software for human genomic regions. *Brief Bioinform* 2022;**23**:bbac329.
34. Wang F, Bai X, Wang Y, *et al.* ATACdb: a comprehensive human chromatin accessibility database. *Nucleic Acids Res* 2021;**49**:D55–64.
35. Li X, Shi L, Wang Y, *et al.* OncoBase: a platform for decoding regulatory somatic mutations in human cancers. *Nucleic Acids Res* 2019;**47**:D1044–55.
36. Tang Z, Li C, Kang B, *et al.* GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res* 2017;**45**:W98–102.
37. Szklarczyk D, Franceschini A, Kuhn M, *et al.* The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* 2011;**39**:D561–8.
38. Deng W, Nickle DC, Learn GH, *et al.* ViroBLAST: a stand-alone BLAST web server for flexible queries of multiple databases and user's datasets. *Bioinforma Oxf Engl* 2007;**23**:2334–6.
39. Hung KL, Luebeck J, Dehkordi SR, *et al.* Targeted profiling of human extrachromosomal DNA by CRISPR-CATCH. *Nat Genet* 2022;**54**:1746–54.