

# LincSNP 3.0: an updated database for linking functional variants to human long non-coding RNAs, circular RNAs and their regulatory elements

Yue Gao<sup>†</sup>, Xin Li<sup>†</sup>, Shipeng Shang<sup>†</sup>, Shuang Guo<sup>†</sup>, Peng Wang<sup>ORCID</sup>, Dailin Sun, Jing Gan, Jie Sun, Yakun Zhang, Junwei Wang, Xinyue Wang, Xia Li<sup>ORCID\*</sup>, Yunpeng Zhang\* and Shangwei Ning<sup>ORCID\*</sup>

College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, China

Received September 15, 2020; Revised October 15, 2020; Editorial Decision October 16, 2020; Accepted November 11, 2020

## ABSTRACT

We describe an updated comprehensive database, LincSNP 3.0 (<http://bioinfo.hrbmu.edu.cn/LincSNP>), which aims to document and annotate disease or phenotype-associated variants in human long non-coding RNAs (lncRNAs) and circular RNAs (circRNAs) or their regulatory elements. LincSNP 3.0 has updated with several novel features, including (i) more types of variants including single nucleotide polymorphisms (SNPs), linkage disequilibrium SNPs (LD SNPs), somatic mutations and RNA editing sites have been expanded; (ii) more regulatory elements including transcription factor binding sites (TFBSs), enhancers, DNase I hypersensitive sites (DHSs), topologically associated domains (TADs), footprints, methylations and open chromatin regions have been added; (iii) the associations among circRNAs, regulatory elements and variants have been identified; (iv) more experimentally supported variant-lncRNA/circRNA-disease/phenotype associations have been manually collected; (v) the sources of lncRNAs, circRNAs, SNPs, somatic mutations and RNA editing sites have been updated. Moreover, four flexible online tools including Genome Browser, Variant Mapper, Circos Plotter and Functional Annotation have been developed to retrieve, visualize and analyze the data. Collectively, LincSNP 3.0 provides associations among functional variants, regulatory elements, lncRNAs and circRNAs in diseases. It will serve as an important and continually updated resource for investigating functions and mechanisms of lncRNAs and circRNAs in diseases.

## INTRODUCTION

Emerging evidences have shown the importance of long non-coding RNAs (lncRNAs) as novel regulators of many physical or pathological processes (1). lncRNAs are a highly versatile class of transcripts that have sparked new lines of research in nearly all fields of life sciences (2). More recently, accumulating evidence has indicated that lncRNAs are major players in many types of disease processes. Circular RNAs (circRNAs) are considered as a kind of special lncRNAs and represent a recent research hotspot in the field of RNA. Unlike linear RNA molecules, circRNAs are closed circular molecules with a covalently closed loop structure that lack 5'-3' polarity or a polyadenylated tail (3). CircRNAs play a critical role in biological processes and are reported to participate in multiple disease processes, and thus, these molecules offer new potential opportunities for therapeutic intervention (4).

With the in-depth research for lncRNA, many researchers continue focusing on the influences of genetic variants on lncRNA function. In order to facilitate the study of lncRNA-related genetic variants, we reported the first and second versions of the LincSNP database (LincSNP 1.0 and 2.0) that allows users to search all known disease-associated single nucleotide polymorphisms (SNPs) to human lncRNAs and their transcription factor binding sites (TFBSs). In fact, there are many other kinds of functional variants besides SNPs. For example, somatic mutations, considered as genomic variation phenomena, can directly or indirectly alter lncRNA expression, protein activities, and signaling pathways (5). RNA editing is discovered as a process of variant at RNA level and represents a widespread and prominent post-transcriptional mechanism (6). Its deregulation has been implicated in the pathogenesis of many functions and diseases (7).

\*To whom correspondence should be addressed. Tel: +86 45186615922; Fax: +86 45186615922; Email: ningsw@ems.hrbmu.edu.cn  
Correspondence may also be addressed to Yunpeng Zhang. Email: zhangyp@hrbmu.edu.cn  
Correspondence may also be addressed to Xia Li. Email: lixia@hrbmu.edu.cn

<sup>†</sup>The authors wish it to be known that, in their opinion, the first four authors should be regarded as joint First Authors.

Encyclopedia of DNA elements (ENCODE) project announces >1.2 million candidate functional regulatory elements in human and mouse. This project shows that focusing on variants in regulatory elements within the non-coding regions can help to discover the real pathogenic variations (8). The ENCODE project has put forward a lot of novel insights on the regulated ways at the genome level, and identified many new types of regulatory elements such as TFBSs, enhancers, DNase I hypersensitive sites (DHSs), topologically associated domains (TADs), footprints and open chromatin regions (9). These functional elements could directly or indirectly influence the lncRNAs and circRNAs (10–12). In addition, SNPs can contribute to disease progression by influencing DNA methylation. DNA methylation quantitative trait loci (meQTL) have been identified in physiological and pathological contexts (13). Exploring the functional variants in these regulatory elements of lncRNAs and circRNAs could provide novel insights for investigating the functions and mechanisms of lncRNAs and circRNAs in human diseases.

To bridge this gap, we have updated LincSNP (14,15) to version 3.0 (LincSNP 3.0) (Figure 1 and Table 1). The database aims specifically to store and annotate disease or phenotype-associated variants including SNPs, linkage disequilibrium SNPs (LD SNPs), somatic mutations and RNA editing in human lncRNAs and circRNAs or their regulatory elements including TFBS, enhancer, DHS, TAD, footprint and open chromatin region. In addition, the effects of SNPs on methylation for lncRNAs and circRNAs are also included. Specially, four useful web-based tools are also provided to facilitate data analysis, extraction and visualization. LincSNP 3.0 bridges the gap between functional variants and human lncRNAs or circRNAs to enhance our understanding of lncRNA and circRNA function, particularly their potential roles in human disease.

## MATERIALS AND METHODS

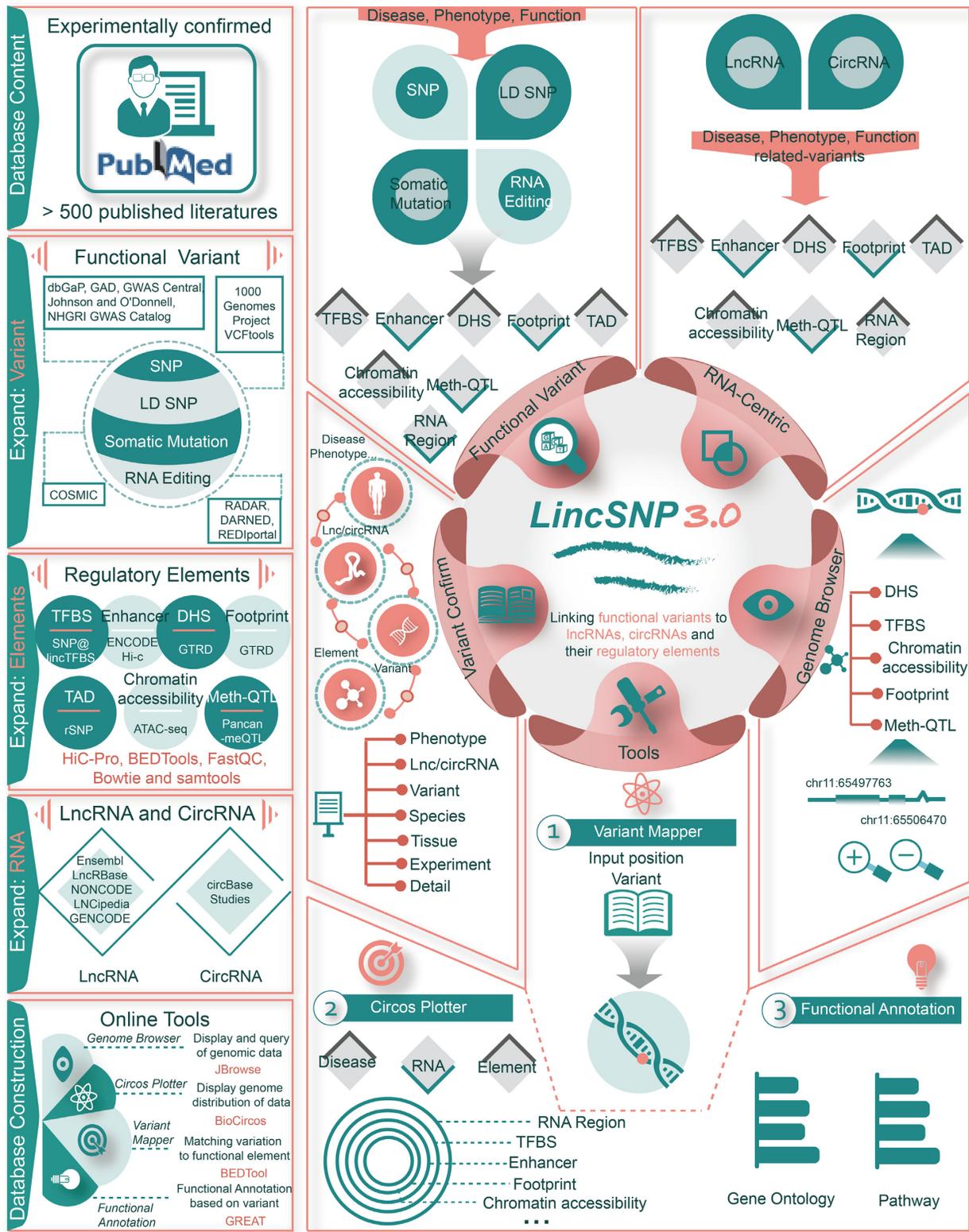
### Data collection and processing

*Experimentally verified functional variants in lncRNAs, circRNAs and their regulatory elements.* All experimentally supported variant–lncRNA/circRNA–disease/trait associations were manually collected through several steps, as previously described (16): (i) We downloaded all published literatures through the PubMed database (17) with a list of keywords (before March 2020), such as ‘lncRNA/circRNA SNP disease,’ ‘lncRNA/circRNA mutation disease,’ ‘lncRNA/circRNA variant disease,’ ‘lncRNA/circRNA polymorphism disease’ and ‘lncRNA/circRNA variant trait.’ (ii) Based on above published papers, experimentally supported variant–lncRNA/circRNA–disease/trait associations were manually curated by at least two researchers. We retrieved the lncRNAs, circRNAs, variants, phenotype and disease names, experimental samples and methods, PubMed IDs, paper titles and a brief description from the original studies. Only high-quality associations with multiple lines of strong experimental evidence were collected, including those confirmed by genotyping, western blot, qRT-PCR or luciferase reporter assays. (iii) All selected studies were rechecked for the lncRNAs, circRNAs, variants and disease names. In addition, some names were

replaced with their official or recommended name respectively, which acquired from Ensembl(18), circBase(19) and dbGAP(20) databases.

*Functional variants.* (i) *SNP*, the functional SNP information was obtained from nine high-quality databases including dbGaP (20), GAD (21), GWAS Central (22), Johnson and O’Donnell (23), the NHGRI GWAS Catalog (24), PharmGKB (25), GWASdb (Version 2) (26), GRASP (Version 2) (27) and LnCeVar (28). As the integrated strategy in LincSNP 2.0, functional SNPs were selected from original publications with moderate thresholds ( $P$ -values  $< 1.0 \times 10^{-3}$ ) and only the most significant SNP was selected in cases where the same SNP could be obtained from different publications. In total, 1 489 332 unique functional SNPs were collected. (ii) *LD SNP*, the SNPs that had linkage disequilibrium (LD-SNP,  $r^2 \geq 0.8$ ) relationships with functional SNPs from the 1000 Genomes Project (Phase I, version 3) were extracted. After LD analysis by VCFtools (29), 4 315 056 LD-SNPs were collected in LincSNP 3.0. (iii) *Somatic mutation*, the somatic mutations were collected from COSMIC (Catalogue Of Somatic Mutations In Cancer) database (30). Furthermore, the FATHMM scores in COSMIC were used to screen functional somatic mutations. Following the suggestion of COSMIC, score  $>0.7$  was considered as ‘Pathogenic’. Thus, we extracted 2 492 320 functional somatic mutations. (iv) *RNA editing site*, RNA editing sites were collected and integrated from DARNED (31) and RADAR (32). The tissue information of RNA editing was obtained from REDIPortal (33). Lastly, 3 881 664 RNA editing sites were extracted.

*Regulatory elements.* (i) *TFBS*, the TFBS information was downloaded from GTRD (34) and rSNPBase 3.0 (35). In total, we identified 222 740 257 and 108 488 967 TFBSs in the defined promoter regions of human lncRNAs and circRNAs. (ii) *Enhancer*, first, the enhancer annotation information was obtained from SELER (36) and SEDb (37). Second, we obtained chromatin interaction region data detected by Hi-C (High-throughput/resolution chromosome conformation capture) seq from ENCODE based on HiC-Pro (38). Then, genome intersection between enhancer and chromatin interaction region was considered as a specific enhancer for this cell line. (iii) *DHS*, DHSs were identified by DNase-seq which can be downloaded from UCSC genome browser (39). Only those identified in more than three samples were extracted. Totally, 1 511 655 DHSs were contained for following analysis. (iv) *Footprint*, similar to DHS, footprints were also screened by DNase-seq in more than three samples. 1 791 645 footprints were obtained based on above steps. (v) *TAD*, TADs of lncRNAs and circRNAs were download from database rSNPBase 3.0 (35). 10 223 498 and 5 617 695 TADs of lncRNAs and circRNAs were obtained. (vi) *Open chromatin region*, the open chromatin regions were obtained from a previous study which detected by ATAC-seq (Assay for Transposase-Accessible Chromatin with high throughput sequencing) of 410 tumor samples spanning 23 cancer types (40). Only the open chromatin regions identified in multiple kinds of cancer types were extracted. 562 709 open chromatin regions were ob-



**Figure 1.** Database content and user interface of LincSNP 3.0. The left area contains the database content, which includes collection of confirmed functional variants from literatures, identification of functional variants, regulatory elements and RNAs and construction of web tools. The methods of analyzing raw data and constructing tools are also provided. The right area contain the user interface of LincSNP 3.0. A panel of basic functions and tools have been developed to infer associations among functional variants, regulatory elements, lncRNAs and circRNAs.

**Table 1.** Data content of LincSNP 3.0 and comparison of the current and previous versions

Features	LincSNP 2.0	LincSNP 3.0	Fold increase
<b>Functional variant</b>			
SNP	809 451	1 489 332	1.84
LD SNP	~1 160 000	4 315 056	~3.72
Somatic mutation	–	2 492 320	New
RNA editing site	–	3 881 664	New
<b>Regulatory element</b>			
TFBS for lncRNA/circRNA	593 492/–	222 740 257/108 488 967	375.30/New
Enhancer for lncRNA/circRNA	–/–	209 968/173 438	New/New
DHS for lncRNA/circRNA	–/–	5 560 190/2 868 132	New/New
Footprint for lncRNA/circRNA	–/–	9 084 051/3 258 258	New/New
TAD for lncRNA/circRNA	–/–	10 223 498/5 617 695	New/New
Open chromatin region for lncRNA/circRNA	–/–	1 806 843/882 305	New/New
Meth-QTL for lncRNA/circRNA	–/–	974 853/408 967	New/New
<b>RNA</b>			
LncRNA	244 545	287 313	1.17
CircRNA	–	173 207	New

tained. (vii) *Meth-QTL*, we obtained DNA methylation quantitative trait loci from database Pancan-meQTL (41).

**LncRNA and circRNA.** (i) *lncRNA*, five databases including Ensembl (18), LncRBase (42), NONCODE (43), LNCipedia (44) and GENCODE (45) were integrated for getting lncRNA information. To provide a universal lncRNA annotation for users, lncRNA transcripts downloaded from different sources were considered to be the same transcripts if they have the same positions. Lastly, 287 313 human lncRNAs were extracted. (ii) *CircRNA*, to construct comprehensive and systematic circRNA annotations, we collected predicted circRNAs supported by back-spliced junction sites from the circBase database (19) and another five studies (46–50). Combining all these studies, a total of 173 207 unique circRNAs were included for subsequent analyses. All the annotations of functional variants, regulatory elements and RNAs were transformed to GRCh38 (The Genome Reference Consortium Human Genome Build 38).

### Linking functional variants to human lncRNAs, circRNAs and their regulatory elements

In order to construct associations among functional variants, lncRNAs, circRNAs and their regulatory elements, BEDTools (51) was applied to match genomic locations. The associations between regulatory elements and lncRNAs or circRNAs were based on following two sections: (i) *trans*-regulatory elements were obtained from current high-quality databases; (ii) *cis*-regulatory elements were located in the promoter regions of human lncRNAs or circRNAs (5 kb upstream to 1 kb downstream region of the transcription start site for each lncRNA or circRNA). The functional variants must be located at regions of lncRNAs, circRNAs or their regulatory elements. After the above steps, functional variants to lncRNAs, circRNAs and their regulatory elements were linked.

### Database construction and improved user interface

All data in LincSNP 3.0 were stored and managed using MySQL (version 5.5.58). The web interfaces were upgraded by applying Linux, Apache, MySQL and JSP (pre-

hypertext processor) (LAMP) technologies. LincSNP 3.0 is freely available to the research community at <http://bioinfo.hrbmu.edu.cn/LincSNP>. The LincSNP 3.0 online web server was developed using Java Server Pages within Tomcat software (v6). Result tables and visualization of data were implemented by using jQuery (v1.11.3), data table (1.10.10) and Highcharts (V4.0) plugin software. All statistical analyses were performed on R framework (V3.6.3). The LincSNP 3.0 website is well supported by several popular web browsers, such as Microsoft Edge, Google Chrome, Firefox and Safari. In addition, for the convenience of users who have used LincSNP 1.0 and 2.0, the old versions are still in service. Researchers can enter them by clicking on the gateways in the LincSNP 3.0 homepage.

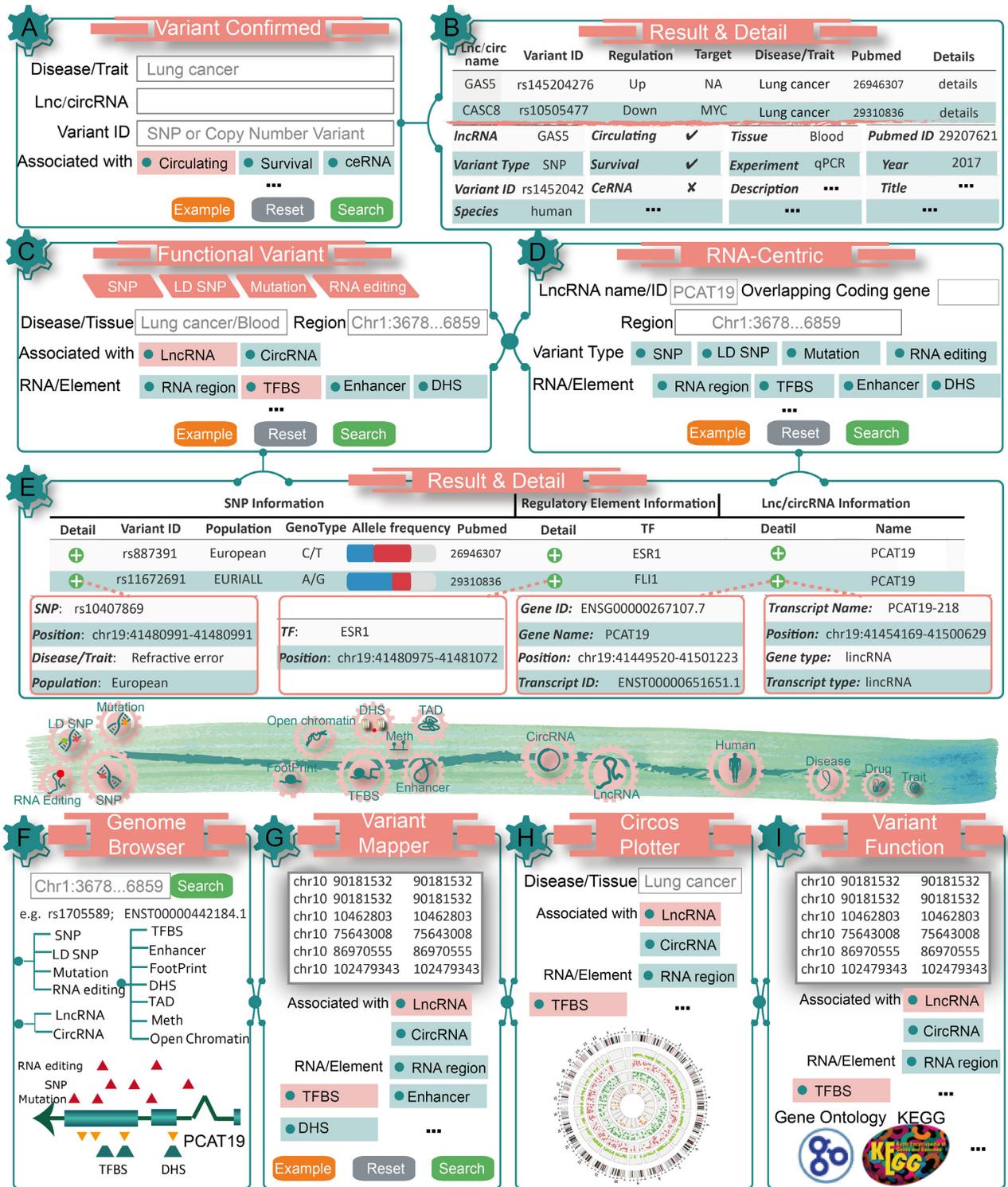
## RESULTS

### Newly designed and more humanized user-friendly interface

LincSNP 3.0 provides a user-friendly web interface that enables users to search, browse, visualize and download data in a few easy steps (Figure 2) including (i) **Variant Confirm** (Figure 2A, B), a retrieval module for experimentally verified lncRNA/circRNA-variants associations. (ii) **Functional Variant-Centric** (Figure 2C), a retrieval module for searching diverse types of functional variants located on lncRNAs, circRNAs or their regulatory elements. (iii) **RNA-centric** (Figure 2D, E), a retrieval module for searching diverse types of RNAs associated with their regulatory elements and different kinds of functional variants. (iv) **Download**, a module for downloading the associations among variants and lncRNAs, circRNAs and (v) **Help**, a module with detailed documentation of user tutorials.

### Web-based online analysis tools of LincSNP 3.0

In order to allow users to explore the features and functions of variants on lncRNAs and circRNAs, four newly designed web-based online tools were developed. These tools could provide assistant for users conveniently to browse, retrieve, visualize and analyze the data. (i) **Genome Browser** (Figure 2F), provides a user-friendly interface for navigating the transcript structures and visualizing the variant sites of the lncRNA or circRNA transcripts in specific diseases.



**Figure 2.** Workflow and examples of using LincSNP 3.0. (A) The variant confirmed search interface of LincSNP 3.0 by querying an example of lung cancer. (B) The query results and detailed information of variants, RNAs, disease, biomarkers and literatures in LincSNP 3.0. (C) The functional variant search interface of LincSNP 3.0 by querying an example of lung cancer and blood. (D) The RNA-centric search interface of LincSNP 3.0 by querying an example of lncRNA PCAT19. (E) The query results and detailed information of variants, RNAs and regulatory elements in LincSNP 3.0. (F) The variants and regulatory elements landscape of PCAT19 in LincSNP 3.0 Genome Browser. (G) Interested variants map to lncRNAs and circRNAs in LincSNP 3.0 Variant Mapper. (H) The circos plot diagram including variants, regulatory and lncRNAs for lung cancer in LincSNP 3.0 Circos Plotter. (I) The functional enrichment analysis for variants in LincSNP 3.0 Functional Annotation.

Users could browse the loci of the lncRNAs, circRNAs and the variants by submitting the genomic interval. Tracks for types of variants and regulatory elements could be selected and added into the 'Search' page. **(ii) Variant Mapper** (Figure 2G), allows users to give interested genomic locations of variants. This tool could implement an integrated pipeline to match the variants on lncRNAs, circRNAs and their regulatory elements. **(iii) Circos Plotter** (Figure 2H), is a data visualizer which could provide a genome circos plot for variants, lncRNAs, circRNAs and their regulatory elements in a specific disease or trait. It visualizes data in a circular layout for exploring relationships between variants, lncRNAs, circRNAs or their regulatory elements. **(iv) Functional Annotation** (Figure 2I), provides functional annotations from Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways based on GREAT (52).

## DISCUSSION AND FUTURE EXTENSIONS

With the completion of the latest research phase of ENCODE project, the researchers discover that 80% of the genome contains elements link to biochemical functions. It also shows that many variants previously correlated with certain diseases lie within or very near the non-coding functional DNA elements, providing new leads for linking genetic variations and diseases (53). The functions and mechanisms of lncRNAs, including circRNAs, need to be further explored and analyzed. Understanding and exploring the variants, regulatory elements, lncRNA and circRNA associations in disease or phenotype will greatly help to reveal the mystery of biological functions. Thus, we construct an updated database, LincSNP 3.0, which aims to integrate functional variants, lncRNAs, circRNAs and their regulatory elements. We believe that LincSNP 3.0 will be a useful resource for functional variants in lncRNAs and circRNAs.

More types of variants including SNPs, LD SNPs, somatic mutations and RNA editing were identified and annotated. We annotate RNA editing sites on DNA regulatory elements to show genome distribution for them. In future work, we will focus on if there are something transcribed from these regulatory regions and annotate the RNA editing sites on RNA level to improve our database. Many kinds of regulatory elements containing TFBSs, enhancers, DHSs, TADs, footprints and open chromatin regions were included. CircRNA information was also added and improved. To improve the functions of data processing and database access, four web-based tools, Genome Browser, Variant Mapper, Circos Plotter and Functional annotation, were developed. We expect that more disease and phenotype-associated variants are mapped to lncRNAs, circRNAs and their regulatory elements and to collect the newly identified regulatory elements for lncRNAs and circRNAs. We will continually maintain and update the LincSNP database and integrate more data sets. We expect that through continuous improvements, LincSNP can become an effective tool to analyze the functional variants of lncRNAs and circRNAs and even contribute to disease diagnosis and treatment.

## FUNDING

National Natural Science Foundation of China [32070672, 32070622, 61873075]; National Key R&D Program of China [2018YFC2000100]; Heilongjiang Touyan Innovation Team Program; Heilongjiang Provincial Natural Science Foundation [LH2020C057]. Funding for open access charge: National Key R&D Program of China [2018YFC2000100].

*Conflict of interest statement.* None declared.

## REFERENCES

- Marchese, F.P., Raimondi, I. and Huarte, M. (2017) The multidimensional mechanisms of long noncoding RNA function. *Genome Biol.*, **18**, 206.
- Grote, P. and Boon, R.A. (2018) LncRNAs coming of age. *Circ. Res.*, **123**, 535–537.
- Chen, L.L. and Yang, L. (2015) Regulation of circRNA biogenesis. *RNA Biol.*, **12**, 381–388.
- Han, B., Chao, J. and Yao, H. (2018) Circular RNA and its mechanisms in disease: from the bench to the clinic. *Pharmacol. Ther.*, **187**, 31–44.
- Gao, Y., Li, X., Zhi, H., Zhang, Y., Wang, P., Wang, Y., Shang, S., Fang, Y., Shen, W., Ning, S. *et al.* (2019) Comprehensive characterization of somatic mutations impacting lncRNA expression for Pan-Cancer. *Mol. Ther. Nucleic Acids*, **18**, 66–79.
- Lo Giudice, C., Tangaro, M.A., Pesole, G. and Picardi, E. (2020) Investigating RNA editing in deep transcriptome datasets with REDIttools and REDIportal. *Nat. Protoc.*, **15**, 1098–1131.
- Gallo, A., Vukic, D., Michalik, D., O'Connell, M.A. and Keegan, L.P. (2017) ADAR RNA editing in human disease; more to it than meets the I. *Hum. Genet.*, **136**, 1265–1278.
- Luo, Y., Hitz, B.C., Gabdank, I., Hilton, J.A., Kagda, M.S., Lam, B., Myers, Z., Sud, P., Jou, J., Lin, K. *et al.* (2020) New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res.*, **48**, D882–D889.
- Snyder, M.P., Gingeras, T.R., Moore, J.E., Weng, Z., Gerstein, M.B., Ren, B., Hardison, R.C., Stamatoiyannopoulos, J.A., Graveley, B.R., Feingold, E.A. *et al.* (2020) Perspectives on ENCODE. *Nature*, **583**, 693–698.
- Ding, L., Zhao, Y., Dang, S., Wang, Y., Li, X., Yu, X., Li, Z., Wei, J., Liu, M. and Li, G. (2019) Circular RNA circ-DONSON facilitates gastric cancer growth and invasion via NURF complex dependent activation of transcription factor SOX4. *Mol. Cancer*, **18**, 45.
- Huang, S., Li, X., Zheng, H., Si, X., Li, B., Wei, G., Li, C., Chen, Y., Liao, W., Liao, Y. *et al.* (2019) Loss of super-enhancer-regulated circRNA Nfix induces cardiac regeneration after myocardial infarction in adult mice. *Circulation*, **139**, 2857–2876.
- Yang, X.H., Nadadur, R.D., Hilvering, C.R., Bianchi, V., Werner, M., Mazurek, S.R., Gadek, M., Shen, K.M., Goldman, J.A., Tyan, L. *et al.* (2017) Transcription-factor-dependent enhancer transcription defines a gene regulatory network for cardiac rhythm. *Elife*, **6**, e31683.
- Heyn, H. (2016) Quantitative trait loci identify functional noncoding variation in cancer. *PLoS Genet.*, **12**, e1005826.
- Ning, S., Yue, M., Wang, P., Liu, Y., Zhi, H., Zhang, Y., Zhang, J., Gao, Y., Guo, M., Zhou, D. *et al.* (2017) LincSNP 2.0: an updated database for linking disease-associated SNPs to human long non-coding RNAs and their TFBSs. *Nucleic Acids Res.*, **45**, D74–D78.
- Ning, S., Zhao, Z., Ye, J., Wang, P., Zhi, H., Li, R., Wang, T. and Li, X. (2014) LincSNP: a database of linking disease-associated SNPs to human large intergenic non-coding RNAs. *BMC Bioinformatics*, **15**, 152.
- Gao, Y., Wang, P., Wang, Y., Ma, X., Zhi, H., Zhou, D., Li, X., Fang, Y., Shen, W., Xu, Y. *et al.* (2019) Lnc2Cancer v2.0: updated database of experimentally supported long non-coding RNAs in human cancers. *Nucleic Acids Res.*, **47**, D1028–D1033.
- NCBI, Resource Coordinators (2018) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **46**, D8–D13.

18. Cunningham,F., Achuthan,P., Akanni,W., Allen,J., Amode,M.R., Armean,I.M., Bennett,R., Bhai,J., Billis,K., Boddur,S. *et al.* (2019) Ensembl 2019. *Nucleic Acids Res.*, **47**, D745–D751.
19. Glazar,P., Papavasiliou,P. and Rajewsky,N. (2014) circBase: a database for circular RNAs. *RNA*, **20**, 1666–1670.
20. Tryka,K.A., Hao,L., Starcke,A., Jin,Y., Wang,Z.Y., Ziyabari,L., Lee,M., Popova,N., Sharopova,N., Kimura,M. *et al.* (2014) NCBI's Database of Genotypes and Phenotypes: dbGaP. *Nucleic Acids Res.*, **42**, D975–D979.
21. Becker,K.G., Barnes,K.C., Bright,T.J. and Wang,S.A. (2004) The genetic association database. *Nat. Genet.*, **36**, 431–432.
22. Beck,T., Shorter,T. and Brookes,A.J. (2020) GWAS Central: a comprehensive resource for the discovery and comparison of genotype and phenotype data from genome-wide association studies. *Nucleic Acids Res.*, **48**, D933–D940.
23. Johnson,A.D. and O'Donnell,C.J. (2009) An open access database of genome-wide association results. *BMC Med. Genet.*, **10**, 6.
24. Welter,D., MacArthur,J., Morales,J., Burdett,T., Hall,P., Junkins,H., Klemm,A., Flicek,P., Manolio,T., Hindorf,L. *et al.* (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.*, **42**, D1001–D1006.
25. Altman,R.B. (2007) PharmGKB: a logical home for knowledge relating genotype to drug response phenotype. *Nat. Genet.*, **39**, 426.
26. Li,M.J., Liu,Z., Wang,P., Wong,M.P., Nelson,M.R., Kocher,J.P., Yeager,M., Sham,P.C., Chanock,S.J., Xia,Z. *et al.* (2016) GWASdb v2: an update database for human genetic variants identified by genome-wide association studies. *Nucleic Acids Res.*, **44**, D869–D876.
27. Eicher,J.D., Landowski,C., Stackhouse,B., Sloan,A., Chen,W., Jensen,N., Lien,J.P., Leslie,R. and Johnson,A.D. (2015) GRASP v2.0: an update on the Genome-Wide Repository of Associations between SNPs and phenotypes. *Nucleic Acids Res.*, **43**, D799–D804.
28. Wang,P., Li,X., Gao,Y., Guo,Q., Ning,S., Zhang,Y., Shang,S., Wang,J., Wang,Y., Zhi,H. *et al.* (2020) LnCeVar: a comprehensive database of genomic variations that disturb ceRNA network regulation. *Nucleic Acids Res.*, **48**, D111–D117.
29. Danecek,P., Auton,A., Abecasis,G., Albers,C.A., Banks,E., DePristo,M.A., Handsaker,R.E., Lunter,G., Marth,G.T., Sherry,S.T. *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
30. Forbes,S.A., Bindal,N., Bamford,S., Cole,C., Kok,C.Y., Beare,D., Jia,M., Shepherd,R., Leung,K., Menzies,A. *et al.* (2011) COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.*, **39**, D945–D950.
31. Kiran,A. and Baranov,P.V. (2010) DARNED: a DATABASE of RNA EDiting in humans. *Bioinformatics*, **26**, 1772–1776.
32. Ramaswami,G. and Li,J.B. (2014) RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic Acids Res.*, **42**, D109–D113.
33. Picardi,E., D'Erchia,A.M., Lo Giudice,C. and Pesole,G. (2017) REDIPortal: a comprehensive database of A-to-I RNA editing events in humans. *Nucleic Acids Res.*, **45**, D750–D757.
34. Yevshin,I., Sharipov,R., Kolmykov,S., Kondrakhin,Y. and Kolpakov,F. (2019) GTRD: a database on gene transcription regulation-2019 update. *Nucleic Acids Res.*, **47**, D100–D105.
35. Guo,L. and Wang,J. (2018) rSNPBase 3.0: an updated database of SNP-related regulatory elements, element-gene pairs and SNP-based gene regulatory networks. *Nucleic Acids Res.*, **46**, D1111–D1116.
36. Guo,Z.W., Xie,C., Li,K., Zhai,X.M., Cai,G.X., Yang,X.X. and Wu,Y.S. (2019) SELER: a database of super-enhancer-associated lncRNA-directed transcriptional regulation in human cancers. *Database (Oxford)*, **2019**, baz27.
37. Jiang,Y., Qian,F., Bai,X., Liu,Y., Wang,Q., Ai,B., Han,X., Shi,S., Zhang,J., Li,X. *et al.* (2019) SEDb: a comprehensive human super-enhancer database. *Nucleic Acids Res.*, **47**, D235–D243.
38. Servant,N., Varoquaux,N., Lajoie,B.R., Viara,E., Chen,C.J., Vert,J.P., Heard,E., Dekker,J. and Barillot,E. (2015) HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.*, **16**, 259.
39. Haeussler,M., Zweig,A.S., Tyner,C., Speir,M.L., Rosenbloom,K.R., Raney,B.J., Lee,C.M., Lee,B.T., Hinrichs,A.S., Gonzalez,J.N. *et al.* (2019) The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res.*, **47**, D853–D858.
40. Corces,M.R., Granja,J.M., Shams,S., Louie,B.H., Seoane,J.A., Zhou,W., Silva,T.C., Groeneveld,C., Wong,C.K., Cho,S.W. *et al.* (2018) The chromatin accessibility landscape of primary human cancers. *Science*, **362**, eaav1898.
41. Gong,J., Wan,H., Mei,S., Ruan,H., Zhang,Z., Liu,C., Guo,A.Y., Diao,L., Miao,X. and Han,L. (2019) Pancan-meQTL: a database to systematically evaluate the effects of genetic variants on methylation in human cancer. *Nucleic Acids Res.*, **47**, D1066–D1072.
42. Chakraborty,S., Deb,A., Maji,R.K., Saha,S. and Ghosh,Z. (2014) LncRBase: an enriched resource for lncRNA information. *PLoS One*, **9**, e108010.
43. Fang,S., Zhang,L., Guo,J., Niu,Y., Wu,Y., Li,H., Zhao,L., Li,X., Teng,X., Sun,X. *et al.* (2018) NONCODEV5: a comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Res.*, **46**, D308–D314.
44. Volders,P.J., Anckaert,J., Verheggen,K., Nuytens,J., Martens,L., Mestdagh,P. and Vandesompele,J. (2019) LNCipedia 5: towards a reference set of human long non-coding RNAs. *Nucleic Acids Res.*, **47**, D135–D139.
45. Frankish,A., Diekhans,M., Ferreira,A.M., Johnson,R., Jungreis,I., Loveland,J., Mudge,J.M., Sisu,C., Wright,J., Armstrong,J. *et al.* (2019) GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.*, **47**, D766–D773.
46. Kelly,S., Greenman,C., Cook,P.R. and Papanonis,A. (2015) Exon skipping is correlated with exon circularization. *J. Mol. Biol.*, **427**, 2414–2417.
47. Bachmayr-Heyda,A., Reiner,A.T., Auer,K., Sukhbaatar,N., Aust,S., Bachleitner-Hofmann,T., Mesteri,I., Grunt,T.W., Zeillinger,R. and Pils,D. (2015) Correlation of circular RNA abundance with proliferation—exemplified with colorectal and ovarian cancer, idiopathic lung fibrosis, and normal human tissues. *Sci. Rep.*, **5**, 8057.
48. Gao,Y., Wang,J. and Zhao,F. (2015) CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. *Genome Biol.*, **16**, 4.
49. Zhang,X.O., Wang,H.B., Zhang,Y., Lu,X., Chen,L.L. and Yang,L. (2014) Complementary sequence-mediated exon circularization. *Cell*, **159**, 134–147.
50. Guo,J.U., Agarwal,V., Guo,H. and Bartel,D.P. (2014) Expanded identification and characterization of mammalian circular RNAs. *Genome Biol.*, **15**, 409.
51. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
52. McLean,C.Y., Bristor,D., Hiller,M., Clarke,S.L., Schaar,B.T., Lowe,C.B., Wenger,A.M. and Bejerano,G. (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.*, **28**, 495–501.
53. Ecker,J.R., Bickmore,W.A., Barroso,I., Pritchard,J.K., Gilad,Y. and Segal,E. (2012) Genomics: ENCODE explained. *Nature*, **489**, 52–55.