

# NucMap: a database of genome-wide nucleosome positioning map across species

Yongbing Zhao<sup>1,†</sup>, Jinyue Wang<sup>2,3,4,†</sup>, Fang Liang<sup>2</sup>, Yanxia Liu<sup>2,3,4</sup>, Qi Wang<sup>2,3,4</sup>, Hao Zhang<sup>2,3,4</sup>, Meiye Jiang<sup>2,3</sup>, Zhewen Zhang<sup>2,3</sup>, Wenming Zhao<sup>2,3</sup>, Yiming Bao<sup>2,3</sup>, Zhang Zhang<sup>2,3,5</sup>, Jiayan Wu<sup>3</sup>, Yan W. Asmann<sup>1,\*</sup>, Rujiao Li<sup>2,\*</sup> and Jingfa Xiao<sup>2,3,4,5,\*</sup>

<sup>1</sup>Department of Health Sciences Research, Mayo Clinic, Jacksonville, FL 32224, USA, <sup>2</sup>BIG Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China, <sup>3</sup>CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China, <sup>4</sup>College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China and <sup>5</sup>Collaborative Innovation Center of Genetics and Development, Fudan University, Shanghai 200438, China

Received August 14, 2018; Revised September 21, 2018; Editorial Decision October 07, 2018; Accepted October 10, 2018

## ABSTRACT

**Dynamics of nucleosome positioning affects chromatin state, transcription and all other biological processes occurring on genomic DNA. While MNase-Seq has been used to depict nucleosome positioning map in eukaryote in the past years, nucleosome positioning data is increasing dramatically. To facilitate the usage of published data across studies, we developed a database named nucleosome positioning map (NucMap, <http://bigd.big.ac.cn/nucmap>). NucMap includes 798 experimental data from 477 samples across 15 species. With a series of functional modules, users can search profile of nucleosome positioning at the promoter region of each gene across all samples and make enrichment analysis on nucleosome positioning data in all genomic regions. Nucleosome browser was built to visualize the profiles of nucleosome positioning. Users can also visualize multiple sources of omics data with the nucleosome browser and make side-by-side comparisons. All processed data in the database are freely available. NucMap is the first comprehensive nucleosome positioning platform and it will serve as an important resource to facilitate the understanding of chromatin regulation.**

## INTRODUCTION

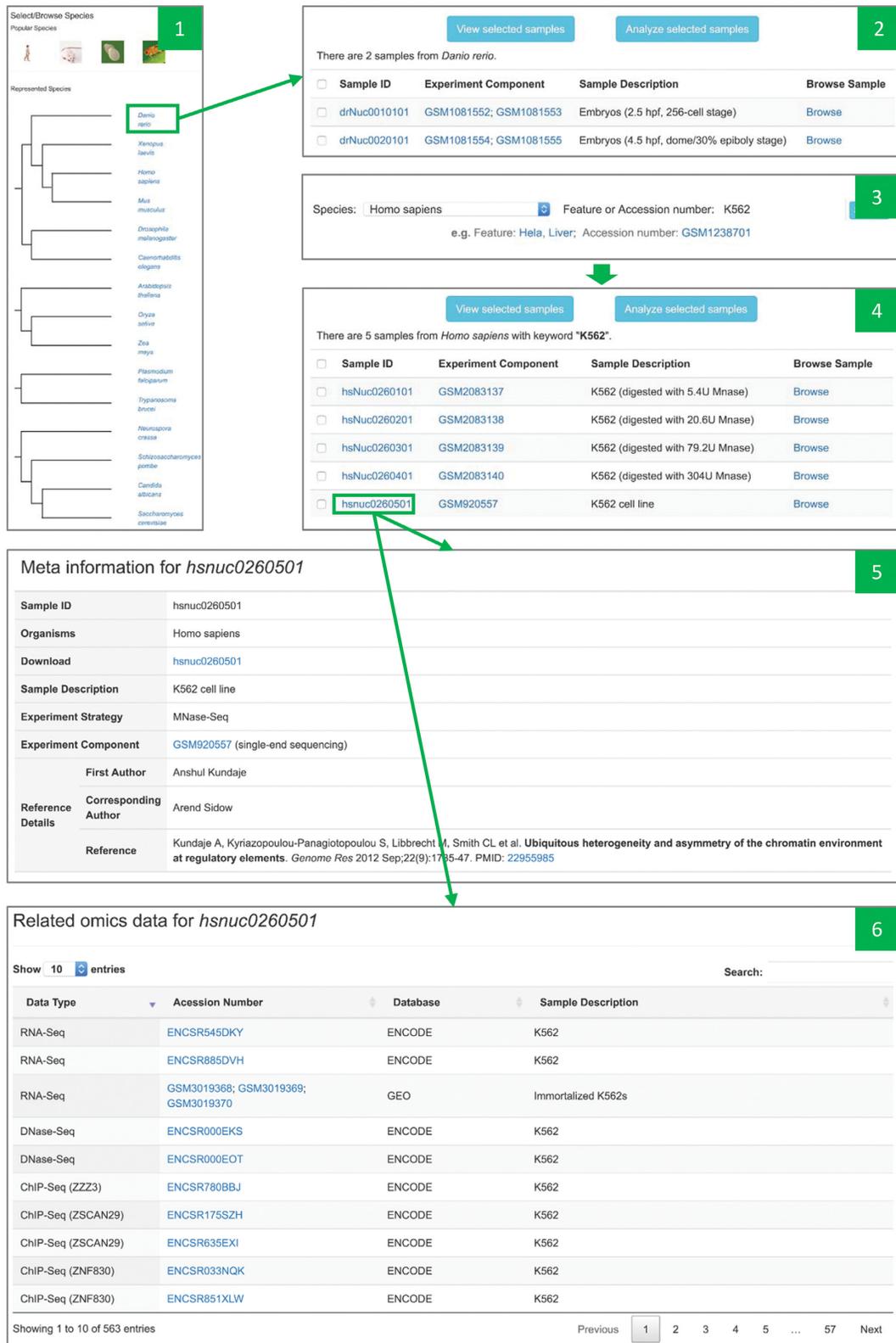
Eukaryotic genomic DNA is tightly packaged into compacted nucleosome arrays, which are the fundamental units of chromatin structure (1). The term ‘nucleosome position-

ing’ is widely used to indicate where nucleosomes occupy on genomic DNA sequence (2–4). In the nucleus, nucleosomes dynamically transform between depletion and de novo occupation on genomic DNA, affecting all biological processes occurring on genomic DNA (5–7). It has been further reported that nucleosome positioning affects transcription initiation and elongation (8). Transcriptional machinery must access to chromatin to trigger sequential gene transcription (7), while nucleosome organization can influence gene activity by controlling the accessibility of transcriptional factor binding sites (9). Some studies suggested that nucleosome positioning influences the evolution of DNA sequence (10–12) since DNA repair machinery has different preferential access between linker DNA and nucleosomal DNA (13).

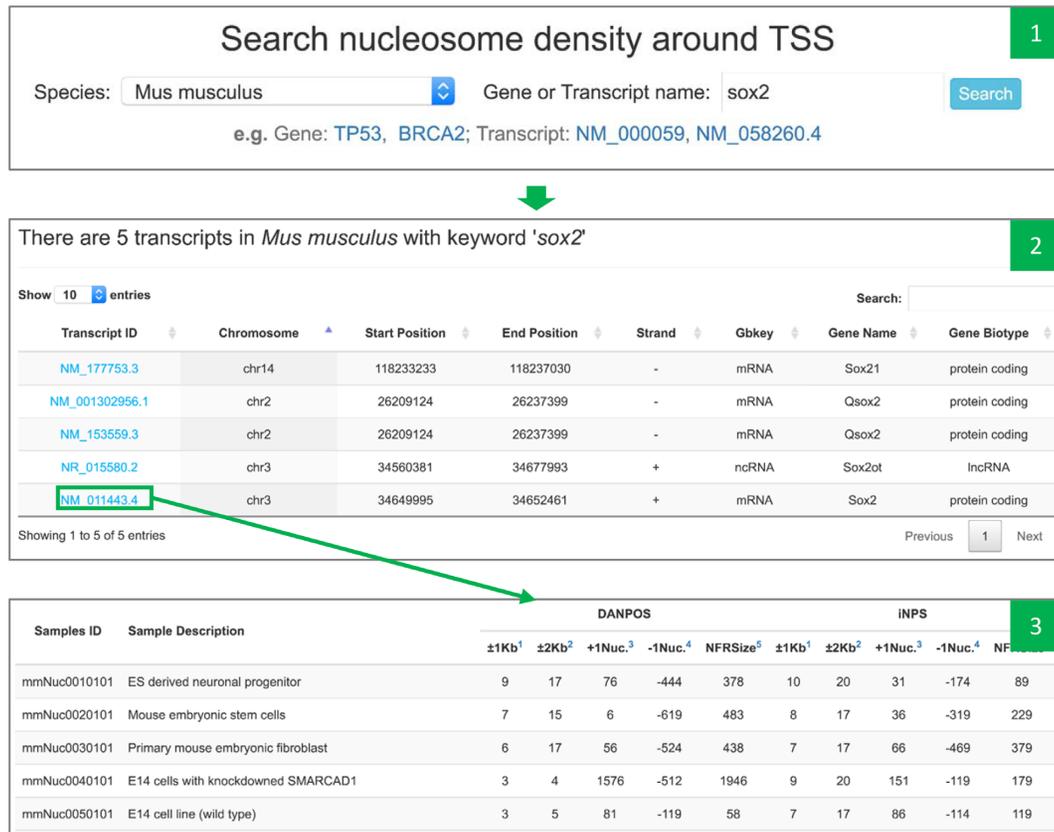
Until now, many different methods have been developed to mapping nucleosomes, such as predicting nucleosome positioning based on DNA sequence features (14,15), histone ChIP-Seq (16), or chromatin accessibility profiles (17). However, all these methods have limitation on either resolution or genome-wide coverage. MNase-Seq is another prevalent experimental approach in nucleosome mapping. In this approach, chromatin is digested with micrococcal nuclease and then followed by deep sequencing (18,19). Based on MNase-Seq, many computational tools have been developed to facilitate the application of this technology (20). Several programs have been published to identify nucleosome peaks, such as DANPOS (21) and iNPS (22). To better understand the role of nucleosome, it is very important to compare nucleosome profiles across different conditions or cell types. Multiple tools were developed to identify differential nucleosome regions, such as DANPOS (21), DiNuP (23), and Dimnp (24). In the past years,

\*To whom correspondence should be addressed. Tel: +86 10 84097638; Fax: +86 10 84097845; E-mail: lirj@big.ac.cn  
Correspondence may also be addressed to Jingfa Xiao. Tel: +86 10 84097443; Fax: +86 10 84097720; Email: xiaojingfa@big.ac.cn  
Correspondence may also be addressed to Yan W. Asmann. Tel: +1 904 953 6717; Email: asmann.yan@mayo.edu

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



**Figure 1.** Screenshots of sample-based search module. Panel 1: interface of shortcut links for all species and popular species. Panel 2: searching result by clicking the shortcut link for '*Danio rerio*'. Panel 3: the interface for keyword-based search. Panel 4: result by searching with 'K562' in '*Homo sapiens*'. Panel 5: the snapshot of sample meta information for hsnuc0260501. Panel 6: the snapshot of related omics data for hsnuc0260501 (only partially datasets were shown).



**Figure 2.** Screenshots of gene-based search module. Panel 1: interface for gene/transcript searching. Panel 2: all available transcripts in *Mus musculus* with the keyword 'sox2'. Panel 3: peak number and nucleosome positioning profile at the transcription start site of NM\_011443.4.

a large number of studies has employed MNase-Seq to depict nucleosome positioning map in eukaryote ranging from yeast to human (5,18,25,26). Consequently, MNase-Seq data is rapidly growing across a wide variety of organisms. It is imperative to build a platform to collect and integrate all published data and make datasets from different studies reusable and comparable, which will largely help biologist to further understand the biology behind nucleosome positioning. However, no such a database or platform was reported. To fill this gap, here we present NucMap, a database of genome-wide nucleosome positioning map across species. Based on a large collection of raw sequence data from published studies, NucMap is dedicated to integrating, analyzing, and visualizing nucleosome positioning data across species.

## DATABASE IMPLEMENTATION

All raw MNase-Seq data were downloaded from GEO and ENCODE, processed by in-house pipeline, and then imported into the NucMap database. The main framework of NucMap was developed based on PHP and MySQL, which are a popular and open source script language and a relational database management system for web development, respectively. JQuery and Bootstrap were used to design the front-end web interface. AJAX (Asynchronous JavaScript And XML), a set of Web development techniques, was used to create asynchronous bioinformatics application running

in the back-end. Back-end bioinformatics applications were implemented with Python and Bash. JBrowser (27) was integrated to visualize nucleosome positioning data.

## DATABASE CONTENT AND USAGE

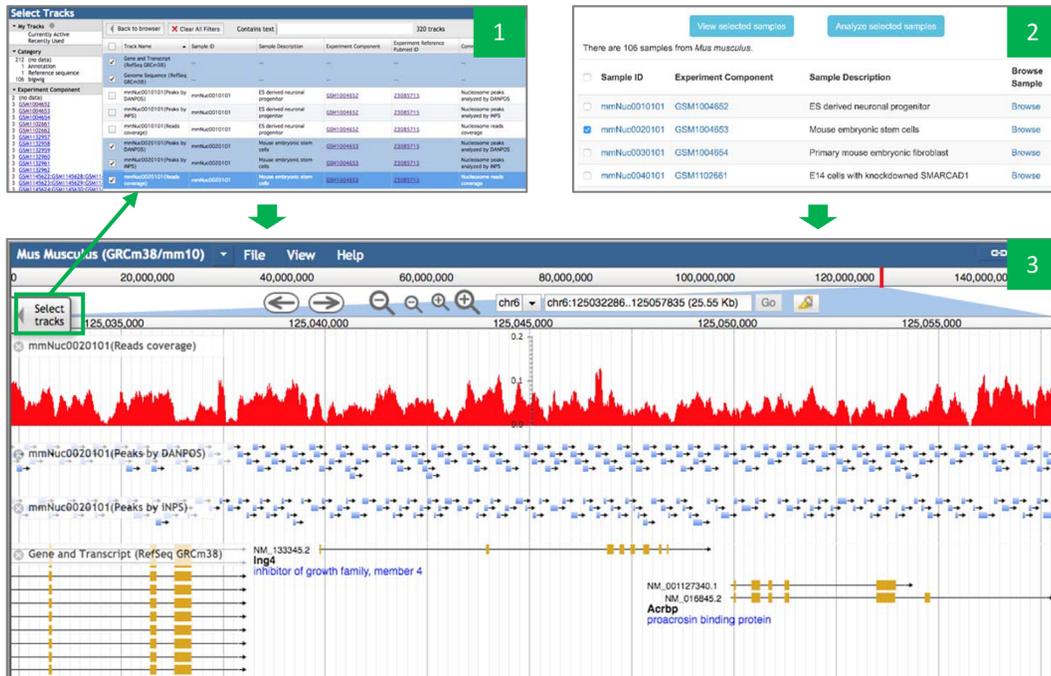
### Overview of NucMap

Currently, we have collected and processed 798 experimental datasets from 477 samples across 15 species. All functionalities in NucMap are organized into four modules: search, browse, analysis and download.

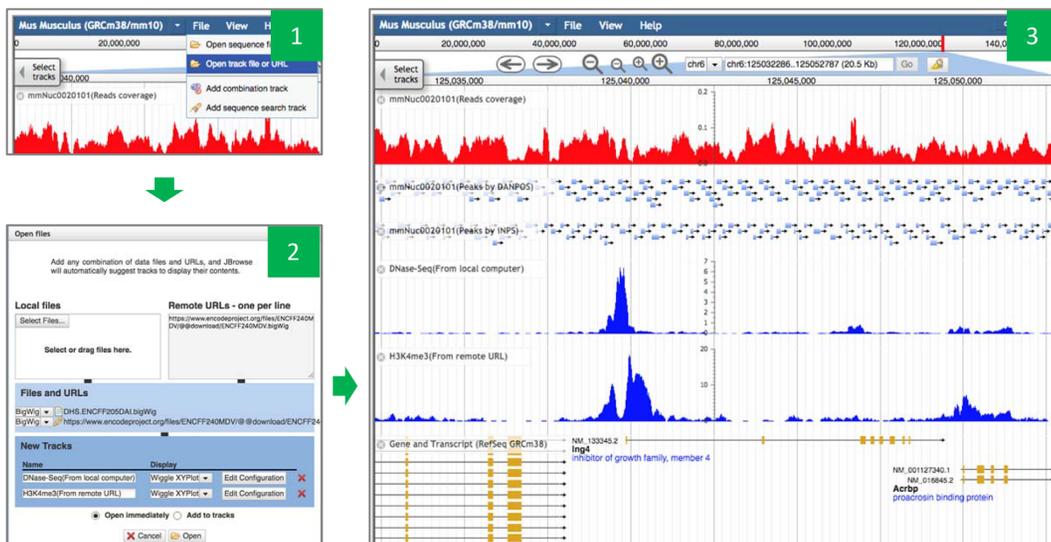
### Searching NucMap

We have developed two types of search modules in NucMap, which are sample-based and gene-based. Sample-based search mainly focuses on helping users to find the samples they are interested in (Figure 1). Shortcut links can be used to obtain all samples for specific species. Users can also search sample of interest with accession number in GEO and ENCODE or sample feature. With hyper-linked sample ID, users can access more specific information for each sample, including original data source, original publication, all downloadable data for this sample and other related omics samples. On the page of search result, the buttons 'View selected samples' and 'Analyzed selected samples' connect searching results to other modules

A



B

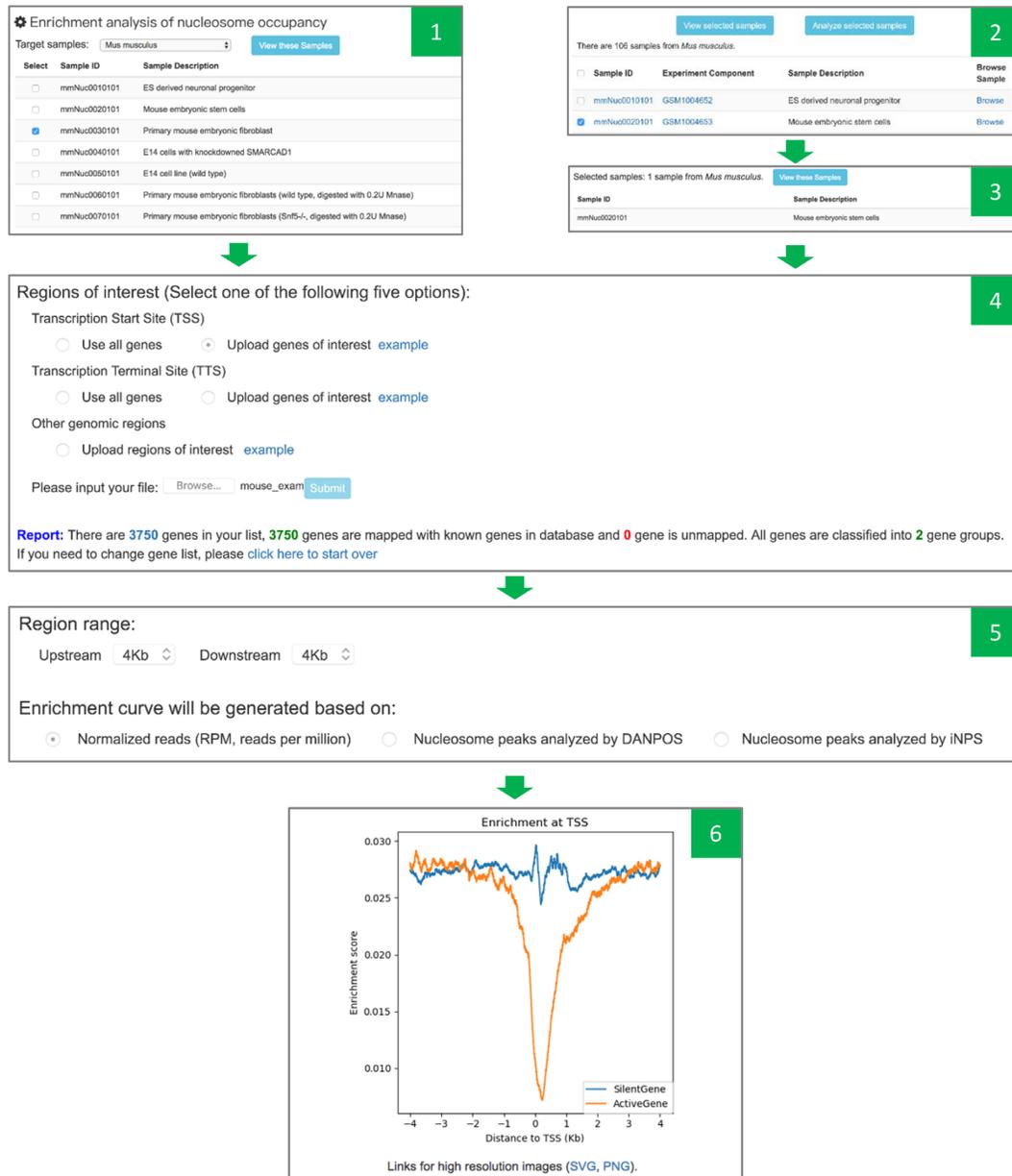


**Figure 3.** Screenshots of nucleosome browser. (A) Browsing tracks deposited the NucMap database. There are two routes to choose and browse sample tracks of interest, which are shown as Panel 1 and 2 respectively. Panel 1: choose tracks of interest with the track selection table in the nucleosome browser. Panel 2: click ‘Browse’ link associated to each sample or click ‘View selected sample’ button after selecting sample of interest in sample-based search module. Panel 3: visualizing tracks from the sample ‘mmNuc0020101’, in which green box is the button to show track selection table. (B) Browsing tracks from local computer or remote URL. Panel 1: menu for opening a track file or remote URL. Panel 2: interface for loading and configuring tracks from local computer and remote URL. Panel 3: visualizing tracks from local computer and remote URL (blue) together with tracks from the sample ‘mmNuc0020101’.

in NucMap. With these two buttons, users can visualize nucleosome positioning data from selected samples in the nucleosome browser or perform analysis on selected samples with the online analysis module, which will be introduced in the later sections.

Promoter-associated nucleosome free region (NFR) is related to promoter-proximal pausing to enable precise gene regulation (28,29). Gene-based searching helps users to

checkup nucleosome positioning at the promoter region of an individual gene. Both gene name and transcript name are supported in searching. The number of nucleosome peaks at transcription start site were provided in different genomic ranges. The positions of +1 and -1 nucleosome are provided to check up the size of NFR at promoter. Moreover, nucleosome density information across all samples in the



**Figure 4.** Screenshots of online enrichment analysis. There are two routes to choose samples of interest: choosing samples on the analysis module page (Panel 1) and choosing samples on the sample searching page (shown as Panel 2 and Panel 3). Panel 4 is showing the interface for selecting regions of interest, uploading file for genes or regions of interest, and validity report of uploaded genes or genomic regions. Panel 5 is the interface for analysis parameter selection. Panel 6 is the figure automatically generated by the online analysis module and the links for high resolution figures.

same species will be shown in the same table, so that users can make side-by-side comparison (Figure 2; Table 1).

### Nucleosome browser

To facilitate browsing nucleosome profile at single-base resolution, NucMap has deployed a nucleosome browser based on the open source program JBrowser. In the nucleosome browser, each species has an independent browser instance and track selector. With track selector, users can load or unload the tracks for all processed genomic data, including raw reads density and nucleosome peaks analyzed

by different algorithms (Figure 3A). With interactive buttons and interfaces, users can choose tracks of interest, and zoom in/out and highlight any genomic region on the whole genome. This feature will help users to check every detail regarding nucleosome occupancy on each individual gene or genomic region. Meanwhile, users can also directly load track files from their local computer or a third-part database to the nucleosome browser without uploading data to our server (Figure 3B). Therefore, nucleosome browser can help users quickly make side-by-side comparison across multiple relevant genomic track files. For example, biologist can load DNA methylation data or histone ChIP-Seq data into the

**Table 1.** The number of experiments and samples in each species

Species	#Experiment	#Samples
<i>Arabidopsis thaliana</i>	19	12
<i>Caenorhabditis elegans</i>	21	11
<i>Candida albicans</i>	4	2
<i>Danio rerio</i>	4	2
<i>Drosophila melanogaster</i>	124	70
<i>Homo sapiens</i>	71	50
<i>Mus musculus</i>	215	106
<i>Neurospora crassa</i>	10	6
<i>Oryza sativa</i>	3	1
<i>Plasmodium falciparum</i>	9	9
<i>Saccharomyces cerevisiae</i>	284	186
<i>Schizosaccharomyces pombe</i>	18	14
<i>Trypanosoma brucei</i>	8	4
<i>Xenopus laevis</i>	6	2
<i>Zea mays</i>	2	2

same browser session and obtain a comprehensive overview of chromatin state around a gene of interest.

### Analysis

Genome-wide enrichment analysis is a popular method to understand global features in omics data. To help users make a global analysis on nucleosome positioning patterns, we have developed online analysis module. This module can characterize nucleosome occupancy profile at all genomic regions (Figure 4). Users can also classify the regions of interest into multiple groups according to the purpose of their studies, so that they can compare the difference of enrichment curves among groups. Both normalized raw reads and nucleosome peaks are supported in the enrichment analysis. Finally, publication-quality figures will be presented. All operations are based on web interface, and there is no requirement for prior knowledge regarding bioinformatics tools and programming.

### Download

All processed nucleosome positioning data are freely available. The data for each sample include three levels; (i) processed reads: bigwig track based on aligned reads and aligned reads after enhancing signal; (ii) nucleosome peaks: nucleosome peaks identified by iNPS and DANPOS; (iii) annotated peaks and reads: nucleosome peaks annotated to nearest TSS, the matrix of peak count around each TSS and the matrix of aligned reads count around each TSS, which were normalized to RPM (Reads Per Million). For each species, all these data were organized in two ways on the download page, by sample and by data type. Users can also visualize our data with our online links in their local browser or other online genome browsers.

### FUTURE DIRECTIONS

MNase-Seq is an important approach to study the role of nucleosome in transcriptional regulation. With an increasing usage of MNase-Seq in eukaryotes, nucleosome positioning data is rapidly growing. NucMap is the first open resource and platform for nucleosome positioning data from MNase-Seq across species. All available MNase-Seq data in

GEO and ENCODE up to date are included in NucMap. As one of important database resources in BIG Data Center (30), NucMap will be continuously collecting and integrating published data.

Nowadays, biologists usually integrate and analyze multiple-scales omics data to study transcriptional regulation. Nucleosome positioning is one type of chromatin state information. To deeply understand chromatin biology, we will make NucMap compatible to other public epigenomics databases, such as MethBank (31), Cistrome (32) and ENCODE (33). Thus, datasets in other repositories, such as DNA methylation data, histone and transcription factor ChIP-Seq, can be directly loaded and compared with nucleosome positioning data in NucMap. Based on comprehensive analysis on cross-omics data, biologists will therefore learn more about chromatin regulation.

### ACKNOWLEDGEMENTS

We thank members of the BIG Data Center for reporting bugs and sending comments and Dr Jinfeng Shao and Dr Gjalt G. Wybenga for valuable discussion and suggestion on paper writing.

### FUNDING

National Key Research Program of China [2016YFB0201702 to J.X.]; Promoting Big Data Development Project, the National Development and Reform Commission of China [2016-999999-65-01-000696-07 to J.X.]; National Natural Science Foundation of China [31771465 and 31471248 to J.X.]; International Partnership Program of the Chinese Academy of Sciences [153F11KYSB20160008]; Key Program of the Chinese Academy of Sciences [KJZD-EW-L14 to J.X.]; The 13th Five-year Informatization Plan of Chinese Academy of Sciences [XXH13505-05 to J.X.]. Funding for open access charge: National Key Research Program of China [2016YFB0201702].

*Conflict of interest statement.* None declared.

### REFERENCES

- Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. and Richmond, T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, **389**, 251–260.
- Drew, H.R. and Travers, A.A. (1985) DNA bending and its relation to nucleosome positioning. *J. Mol. Biol.*, **186**, 773–790.
- Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thastrom, A., Field, Y., Moore, I.K., Wang, J.P. and Widom, J. (2006) A genomic code for nucleosome positioning. *Nature*, **442**, 772–778.
- Struhl, K. and Segal, E. (2013) Determinants of nucleosome positioning. *Nat. Struct. Mol. Biol.*, **20**, 267–273.
- Teif, V.B., Vainshtein, Y., Caudron-Herger, M., Mallm, J.P., Marth, C., Hofer, T. and Rippe, K. (2012) Genome-wide nucleosome positioning during embryonic stem cell development. *Nat. Struct. Mol. Biol.*, **19**, 1185–1192.
- Jiang, C. and Pugh, B.F. (2009) Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.*, **10**, 161–172.
- Bai, L. and Morozov, A.V. (2010) Gene regulation by nucleosome positioning. *Trends Genet.*, **26**, 476–483.
- Jin, Y.S., Heim, S., Mandahl, N., Biorlund, A., Wennerberg, J. and Mitelman, F. (1990) Unrelated clonal chromosomal aberrations in carcinomas of the oral cavity. *Genes Chromosomes Cancer*, **1**, 209–215.

9. Voong,L.N., Xi,L., Sebeson,A.C., Xiong,B., Wang,J.P. and Wang,X. (2016) Insights into nucleosome organization in mouse embryonic stem cells through chemical mapping. *Cell*, **167**, 1555–1570.
10. Dai,Z., Dai,X. and Xiang,Q. (2011) Genome-wide DNA sequence polymorphisms facilitate nucleosome positioning in yeast. *Bioinformatics*, **27**, 1758–1764.
11. Warnecke,T., Batada,N.N. and Hurst,L.D. (2008) The impact of the nucleosome code on protein-coding sequence evolution in yeast. *PLoS Genet.*, **4**, e1000250.
12. Washietl,S., Machne,R. and Goldman,N. (2008) Evolutionary footprints of nucleosome positions in yeast. *Trends Genet.*, **24**, 583–587.
13. Shim,E.Y., Hong,S.J., Oum,J.H., Yanez,Y., Zhang,Y. and Lee,S.E. (2007) RSC mobilizes nucleosomes to improve accessibility of repair machinery to the damaged chromatin. *Mol. Cell Biol.*, **27**, 1602–1613.
14. Chen,K., Wang,L., Yang,M., Liu,J., Xin,C., Hu,S. and Yu,J. (2010) Sequence signatures of nucleosome positioning in *Caenorhabditis elegans*. *Genomics Proteomics Bioinformatics*, **8**, 92–102.
15. Zhang,J., Peng,W. and Wang,L. (2018) LeNup: learning nucleosome positioning from DNA sequences with improved convolutional neural networks. *Bioinformatics*, **34**, 1705–1712.
16. Zhang,Y., Shin,H., Song,J.S., Lei,Y. and Liu,X.S. (2008) Identifying positioned nucleosomes with epigenetic marks in human from ChIP-Seq. *BMC Genomics*, **9**, 537.
17. Zhong,J., Luo,K., Winter,P.S., Crawford,G.E., Iversen,E.S. and Hartemink,A.J. (2016) Mapping nucleosome positions using DNase-seq. *Genome Res.*, **26**, 351–364.
18. Cui,K. and Zhao,K. (2012) Genome-wide approaches to determining nucleosome occupancy in metazoans using MNase-Seq. *Methods Mol. Biol.*, **833**, 413–419.
19. Schones,D.E., Cui,K., Cuddapah,S., Roh,T.Y., Barski,A., Wang,Z., Wei,G. and Zhao,K. (2008) Dynamic regulation of nucleosome positioning in the human genome. *Cell*, **132**, 887–898.
20. Teif,V.B. (2016) Nucleosome positioning: resources and tools online. *Brief. Bioinform.*, **17**, 745–757.
21. Chen,K., Xi,Y., Pan,X., Li,Z., Kaestner,K., Tyler,J., Dent,S., He,X. and Li,W. (2013) DANPOS: dynamic analysis of nucleosome position and occupancy by sequencing. *Genome Res.*, **23**, 341–351.
22. Chen,W., Liu,Y., Zhu,S., Green,C.D., Wei,G. and Han,J.D. (2014) Improved nucleosome-positioning algorithm iNPS for accurate nucleosome positioning from sequencing data. *Nat. Commun.*, **5**, 4909.
23. Fu,K., Tang,Q., Feng,J., Liu,X.S. and Zhang,Y. (2012) DiNuP: a systematic approach to identify regions of differential nucleosome positioning. *Bioinformatics*, **28**, 1965–1971.
24. Liu,L., Xie,J., Sun,X., Luo,K., Qin,Z.S. and Liu,H. (2017) An approach of identifying differential nucleosome regions in multiple samples. *BMC Genomics*, **18**, 135.
25. Valouev,A., Ichikawa,J., Tonthat,T., Stuart,J., Ranade,S., Peckham,H., Zeng,K., Malek,J.A., Costa,G., McKernan,K. *et al.* (2008) A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Res.*, **18**, 1051–1063.
26. Weiner,A., Hughes,A., Yassour,M., Rando,O.J. and Friedman,N. (2010) High-resolution, nucleosome mapping reveals transcription-dependent promoter packaging. *Genome Res.*, **20**, 90–100.
27. Skinner,M.E., Uzilov,A.V., Stein,L.D., Mungall,C.J. and Holmes,I.H. (2009) JBrowse: a next-generation genome browser. *Genome Res.*, **19**, 1630–1638.
28. Gilchrist,D.A., Dos Santos,G., Fargo,D.C., Xie,B., Gao,Y., Li,L. and Adelman,K. (2010) Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. *Cell*, **143**, 540–551.
29. Jimeno-Gonzalez,S., Ceballos-Chavez,M. and Reyes,J.C. (2015) A positioned +1 nucleosome enhances promoter-proximal pausing. *Nucleic Acids Res.*, **43**, 3068–3078.
30. Members, B.I.G.D.C. (2018) Database resources of the BIG Data Center in 2018. *Nucleic Acids Res.*, **46**, D14–D20.
31. Li,R., Liang,F., Li,M., Zou,D., Sun,S., Zhao,Y., Zhao,W., Bao,Y., Xiao,J. and Zhang,Z. (2018) MethBank 3.0: a database of DNA methylomes across a variety of species. *Nucleic Acids Res.*, **46**, D288–D295.
32. Liu,T., Ortiz,J.A., Taing,L., Meyer,C.A., Lee,B., Zhang,Y., Shin,H., Wong,S.S., Ma,J., Lei,Y. *et al.* (2011) Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol.*, **12**, R83.
33. Consortium,E.P. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.