

Xenbase: deep integration of GEO & SRA RNA-seq and ChIP-seq data in a model organism database

Joshua D. Fortriede^{1,†}, Troy J. Pells^{2,†}, Stanley Chu², Praneet Chaturvedi¹, DongZhuo Wang², Malcom E. Fisher¹, Christina James-Zorn², Ying Wang¹, Mardi J. Nenni¹, Kevin A. Burns¹, Vaneet S. Lotay², Virgilio G. Ponferrada¹, Kamran Karimi^{2,*}, Aaron M. Zorn¹ and Peter D. Vize²

¹Cincinnati Children's Hospital, Division of Developmental Biology, 3333 Burnet Avenue, Cincinnati, OH 45229, USA and ²Departments of Biological Sciences and Computer Science, University of Calgary, Calgary, Alberta T2N1N4, Canada

Received September 14, 2019; Revised October 04, 2019; Editorial Decision October 06, 2019; Accepted November 12, 2019

ABSTRACT

Xenbase (www.xenbase.org) is a knowledge base for researchers and biomedical scientists that employ the amphibian *Xenopus* as a model organism in biomedical research to gain a deeper understanding of developmental and disease processes. Through expert curation and automated data provisioning from various sources Xenbase strives to integrate the body of knowledge on *Xenopus* genomics and biology together with the visualization of biologically significant interactions. Most current studies utilize next generation sequencing (NGS) but until now the results of different experiments were difficult to compare and not integrated with other Xenbase content. Xenbase has developed a suite of tools, interfaces and data processing pipelines that transforms NCBI Gene Expression Omnibus (GEO) NGS content into deeply integrated gene expression and chromatin data, mapping all aligned reads to the most recent genome builds. This content can be queried and visualized via multiple tools and also provides the basis for future automated 'gene expression as a phenotype' and gene regulatory network analyses.

INTRODUCTION

Frogs in the genus *Xenopus* have been used for biomedical research for many decades, initially due to the ability to induce egg laying on demand in these animals. Two species are used in the majority of experiments; the allotetraploid *Xenopus laevis* and the diploid *Xenopus tropicalis*. Each has its experimental advantages, a sequenced genome (1,2)

and extensive transcriptomics and chromatin state next generation sequencing (NGS) data. Xenbase (3,4) serves as the model organism database for researchers using these species, or biomedical researchers in other systems that seek to leverage the uniquely powerful data that these abundant, large volume embryos with rapid external development can provide. Xenbase supports disease research by mapping human diseases, via Disease Ontology (DO) terms, to frog genes and publications (5). Using *Xenopus*, many experiments, such as morpholino knockdowns or CRISPR/Cas9 based genome editing can be performed in F0 embryos without the need to perform long time scale genetic crosses. Two revolutionary and widely used technologies, RNA-seq and ChIP-seq, allow the assessment of these experimental manipulations on genome-wide transcription and chromatin states. However, the full potential of these rich datasets is limited due to differing bioinformatic pipelines and different genome assemblies used in between studies, insufficient metadata using controlled vocabularies, and the lack of a centralized resource to search, visualize, compare and analyze the data. To resolve these problems and make maximal value of this extraordinarily rich and powerful data source, Xenbase has developed an integrated system that loads, maps and processes all publicly available *Xenopus* RNA-seq and ChIP-seq data from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) (6). Beginning with raw FASTQ sequence files all data are mapped against the most recent genome builds and processed by a standardized bioinformatics pipeline based on The Encyclopedia of DNA elements (ENCODE) best practices (7). The processed results are loaded directly into Xenbase database tables (described below) and allow for deep integration with other Xenbase content and powerful new data visualizations.

*To whom correspondence should be addressed. Tel: +403 210 8136; Fax: +203 289 9311; Email: kkarimi@ucalgary.ca

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

GEO METADATA AND THE CURATION PROCESS

The first step in our process is identifying new *Xenopus* data sets in GEO, followed by curating the associated metadata within Xenbase. An overview of the workflow is presented in Figure 1. GEO data is indexed by the GEO series number (GSE) and each GSE is made up of individual samples or replicates (GSMs) and their associated metadata. GSMs are in turn linked to their corresponding raw sequence files in the Sequence Read Archive (SRA; <https://www.ncbi.nlm.nih.gov/sra>). NCBI eUtilities is used to identify new *Xenopus* datasets and fetch the GEO MINiML metadata file together with the SRA Run Table for each GSE, which are subsequently parsed and loaded into database tables. While some metadata can be captured from GEO by automated processes, a major challenge is the considerable variations in author-supplied metadata describing the samples and experiments. To overcome this we use various automated and manual curation steps to curate the metadata with controlled vocabularies and ontologies, including the *Xenopus* Anatomical Ontology (XAO) (8). This includes reviewing the metadata consistency and documenting details described in the associated publication such as the experimental manipulation (e.g. CRISPR mutant, morpholino or mRNA injection), developmental stage, tissue or reagents (e.g. ChIP antibody) used. Replicate GSMs are grouped together and linked to their corresponding control GSMs. This later step is essential to perform the bioinformatics analysis for obtaining the differentially expressed genes (DEGs).

BIOINFORMATICS PIPELINE

Metadata from the automated and manual curations are used to generate a run file needed for the bioinformatic pipeline. The run file contains species, assay type (RNA-seq, ChIP-seq), library construction (single versus paired end reads), the grouping of biological replicates and controls as well as SRA accession numbers used to download the raw read files. A public NGS data analysis pipeline from Computational Suite for Bioinformaticians and Biologists (CSBB) (<https://github.com/csbbcompbio/CSBB-v3.0>) was deployed. CSBB has a standardized analysis pipeline that leverages peer-reviewed open source tools to download and process the raw data (Figure 1). Full documentation of the Xenbase GEO pipeline can be found at GitLab (<https://gitlab.com/Xenbase/bioinformatics/pipeline>)

In brief, SRA files are downloaded from the SRA, converted to FASTQ when necessary and quality checked with FASTQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). ChIP-seq data is mapped to the newest *X. laevis* v9.2 (2) or *X. tropicalis* v9.1 (1) genomes using Bowtie2 (9), and duplicate and multi-mapped reads are removed with Samtools (10). MACS2 (11) is used for peak calling. RNA-seq data are mapped and quantified using RSEM (12). RUVSeq (13) is used to perform differential gene expression analyses for control/treatment experiments, as defined during the curation process, using a >2 fold change and <5% FDR (false discovery rate) cut-offs and control or treatment TPM >1. RUVSeq was chosen for DEG analyses because of its ability to generate accurate

DEG lists in varying setups, including datasets that have no replicates. DeepTools (14) is used to convert both RNA-seq and ChIP-seq BAM files to bigWigs (Figure 1). Technical replicates (multiple FASTQ files for a given GSM) are merged prior to mapping. Biological replicates, multiple GSMs, are mapped and quantified individually, to provide statistical power in the subsequent differential expression analysis. However, to provide a consensus track for a given condition, BAM files for individual biological replicates are merged. The resulting data files include bigWig tracks for the JBrowse genome browser (15) for both RNA-Seq and ChIP-seq data, ChIP peak calls and read count files reporting mRNA expression levels normalized by transcripts per million reads (TPM). All files are available via the Xenbase FTP (File Transfer protocol) server in addition to their use in various visualizations.

One of the major features of this system is that all data are aligned to the most recent genome builds (1,2). To maximize the value of the GEO content, all datasets will be re-processed when stable new genome assemblies are available. Raw sequence reads are not stored in Xenbase as they are available from the SRA, so reprocessing will include re-downloading all of the original FASTQ files.

TPM AND DIFFERENTIAL EXPRESSION DATA

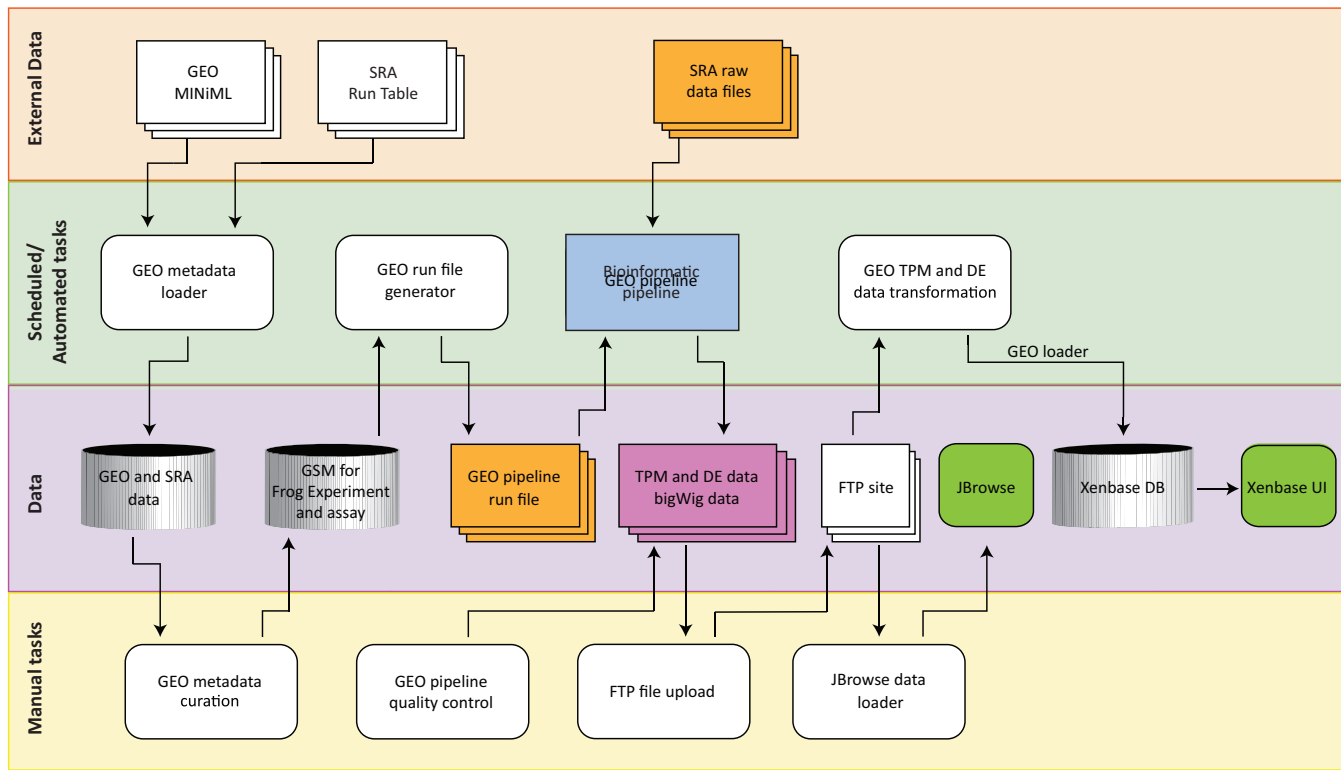
The GEO Loader is a Java application developed to consume the TPM read count and DEG files for loading into the Xenbase database (Figure 1). The imported data are mapped so that the correct links are made back to the GEO metadata that has been curated, and also to the Xenbase gene models, thereby linking the data to all other content including proteins, gene expression, GO annotations, publications etc. Many of the curation steps utilize the XAO to record developmental stages and tissues that are manipulated or tested. As the XAO includes many relationships between terms, such as what organ system or developmental origins a tissue may have, these relationships also exist between curated GEO data.

There are ~89 processed GSEs available, consisting of 1827 individual GSMs (342 ChIP-seq and 1485 RNA-seq). The volume of data output by this system varies by GSE. A single GSE can provide 500 000 averaged TPM records. A star schema with fact and dimension tables is used for storing these records directly in the database and fully integrated into the Xenbase schema, so queries can be used to pull any combination of results directly from the TPM and DEG tables in addition to any linked Xenbase content. On the database side SQL can be used to perform complex queries of any Xenbase content to return and sort TPM values in a flexible manner. For example, a list of DEGs can be sorted via GO terms/co-expressed genes/stage ranges or any other Xenbase content and then TPM values for any combination of GSMs within a GSE displayed.

SEARCHING GEO DATA

Xenbase provides both a simple 'search everything' tool that queries GEO and SRA identifiers as well as metadata including experiment titles (e.g. placode), reagents (e.g. a ChIP antibody), authors, gene symbols etc. (Figure 2). A

A The Xenbase SRA/GEO processing system



B The Xenbase bioinformatic pipeline

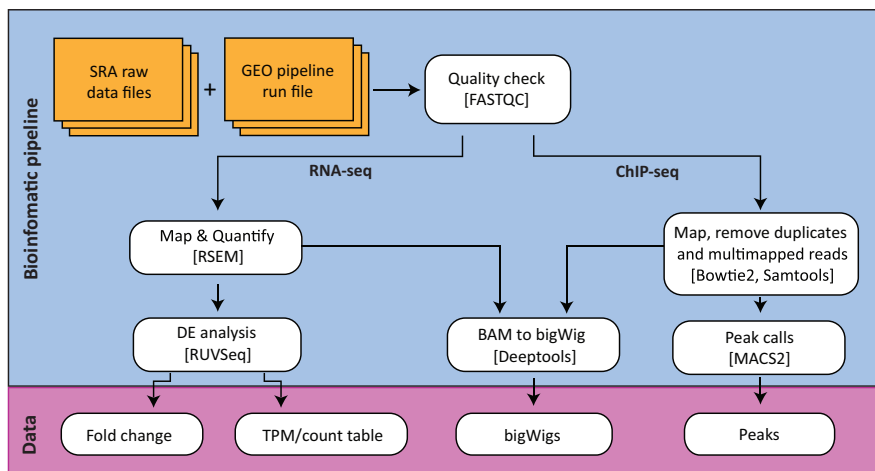


Figure 1. GEO and SRA data collection and processing. (A) Automated systems detect new data and load metadata for curators to use in manual annotation processes. The results are then used to generate a run file that feeds the raw sequencing data from the SRA in the appropriate manner into the GEO bioinformatics pipeline. The pipeline output then supplies files to various Xenbase resources, such as the FTP repository, the JBrowse genome browser and the Xenbase database. (B) Details of the RNA-seq and ChIP-seq data processing pipelines utilizing CSBB (see <https://github.com/csbbcompbio/CSBB-v3.0>).

more controlled advanced search is also available that allows users to select combinations of specific stages, assays, target tissues etc. Both are available via the main horizontal navigation bar on every page under ‘Expression/GEO data at Xenbase’. For example, a Simple Search for ‘wnt8a’ returns two GSEs (Figure 2). The results provide a brief synopsis of the GSE datasets, including the study title, assay types, a link to the article, a description of the sample and a

link to load the sample bigWig track directly into JBrowse. Each GSE record links to a Xenbase GSE page with a summary of the series, the associated metadata (experiment and reagents), links to the publication and the GEO record at NCBI and a link to the Xenbase FTP where users can download all the processed data files. The GSE page also includes a table of the individual GSM samples that make up the GSE series listing the sample name and experimental con-

A

Search High Throughput Sequencing Data (RNA-Seq, ChIP-Seq, etc.)

Keyword/ID:

Species: Any tropicalis laevis

GEO Links

GEO data on FTP
GEO Tutorial Videos
NCBI's GEO site

GSE Search Results:

Results 1 - 2 of 2 results
Page(s): 1

GSE	Title	Assay Types	Article	Sample	Load in JBrowse
GSE72657	Tissue- and stage-specific cellular context regulates Wnt target gene expression subsequent to β -catenin recruitment	RNA-Seq ChIP-Seq	XB-ART-52077	input WE - NF10.25 beta Catenin WE - NF10.25 WE - NF10.25 [±3]	Xtr <input type="checkbox"/> Xtr <input type="checkbox"/> Select All Xtr <input type="checkbox"/>
GSE93195	Spemann organizer transcriptome induction by early β -Catenin, Wnt, Nodal and Siamois signals in <i>Xenopus laevis</i>	RNA-Seq	XB-ART-53589	WE - NF10.5 WE + LiCl - NF10.5 WE + cer1 - NF10.5 [±11]	Xla <input type="checkbox"/> Xla <input type="checkbox"/> Xla <input type="checkbox"/> Select All Xla <input type="checkbox"/>

B

Sample	View	Compare	GSMs	Assay Type	Load in JBrowse
input WE - NF10.25			GSM1867401	ChIP-Seq	[±]
beta Catenin WE - NF10.25			GSM1867400	ChIP-Seq	Xtr <input type="checkbox"/> [±]
WE - NF10.25			GSM1867402 GSM1867406 GSM1867410	RNA-Seq	Xtr <input type="checkbox"/> [±]
WE + hbg1 MO - NF10.25	DEG	<input type="checkbox"/>	GSM1867403 GSM1867407 GSM1867411	RNA-Seq	Xtr <input type="checkbox"/> [±]
WE + wnt8a MO + wnt8a - NF10.25	DEG	<input type="checkbox"/>	GSM1867405 GSM1867409 GSM1867413	RNA-Seq	Xtr <input type="checkbox"/> [±]
WE + wnt8a MO - NF10.25	DEG	<input type="checkbox"/>	GSM1867404 GSM1867408 GSM1867412	RNA-Seq	Xtr <input type="checkbox"/> [±]

Select All DEG Select All Xtr

Figure 2. The GEO and SRA simple search interface and search results. (A) The Simple Search Interface. Any common search term, such as a GEO ID, gene symbol, tissue, author or reagent can be queried. In this example the query was 'wnt8a', returned two GSE results. (B) Within the GSE page opened by selecting the search result (A) of interest, in addition to the metadata and other information on the GSE a table is displayed providing access to the processed data. Results of the DEG analysis between samples can be viewed in Table format by selecting the DEG link (red arrow) or by selecting multiple check boxes under 'Compare' (blue arrow) then selecting this button. Alternatively, bigWig tracks can be viewed by choosing those of interest and selecting the 'Load in JBrowse' button.

ditions with biological replicates grouped together. Importantly the table also provides a link to the DEG data comparing experimental and control samples (discussed below).

VIEWING RNA-SEQ AND CHIP-SEQ DATA

A key goal was to enable users to aggregate and compare results from different functional genomic studies and to this end we have made the processed NGS data available in a number of different ways. The first utilizes the JBrowse genome browser (15), where RNA-seq and ChIP-seq data from many different studies can be directly compared. BigWig files of all the RNA-seq and ChIP-seq tracks can be loaded into JBrowse in a user defined manner. Tracks can be selected from GSE search results, the GSE page and from within JBrowse itself using either the hierarchical or faceted track selectors. The hierarchical selector groups similar tracks first by RNA-seq and ChIP-seq (epigenetic and transcription factor) assay type, and then by antibody type (for ChIP-seq) or tissue and stage for RNA-seq. The

faceted browser allows filtering based on a variety of metadata, including assay type, stage, tissue, manipulation or ChIP-antibody. Xenbase also provides many functional genomics tracks to help users interpret the NGS data including transcription factor binding site predictions, genome segmentation, and species conservation. This is the first time that the results from different studies aligned to various older genome builds can be directly compared using the same data processing system.

RNA-seq expression data are available as simple TPM matrices for all genes in the genome, or as a set of DEGs comparing experimental to control samples. Experiments are matched to controls by curators as described above, and pairs typically reflect sets along the lines of a morpholino injected sample versus a control injected sample set or a region-specific explant versus a stage matched whole embryo. DEG sets are pre-computed using these combinations- they cannot be generated on the fly for custom comparisons. However, these DEG tables are available for download on the FTP and are accessed from the GSE

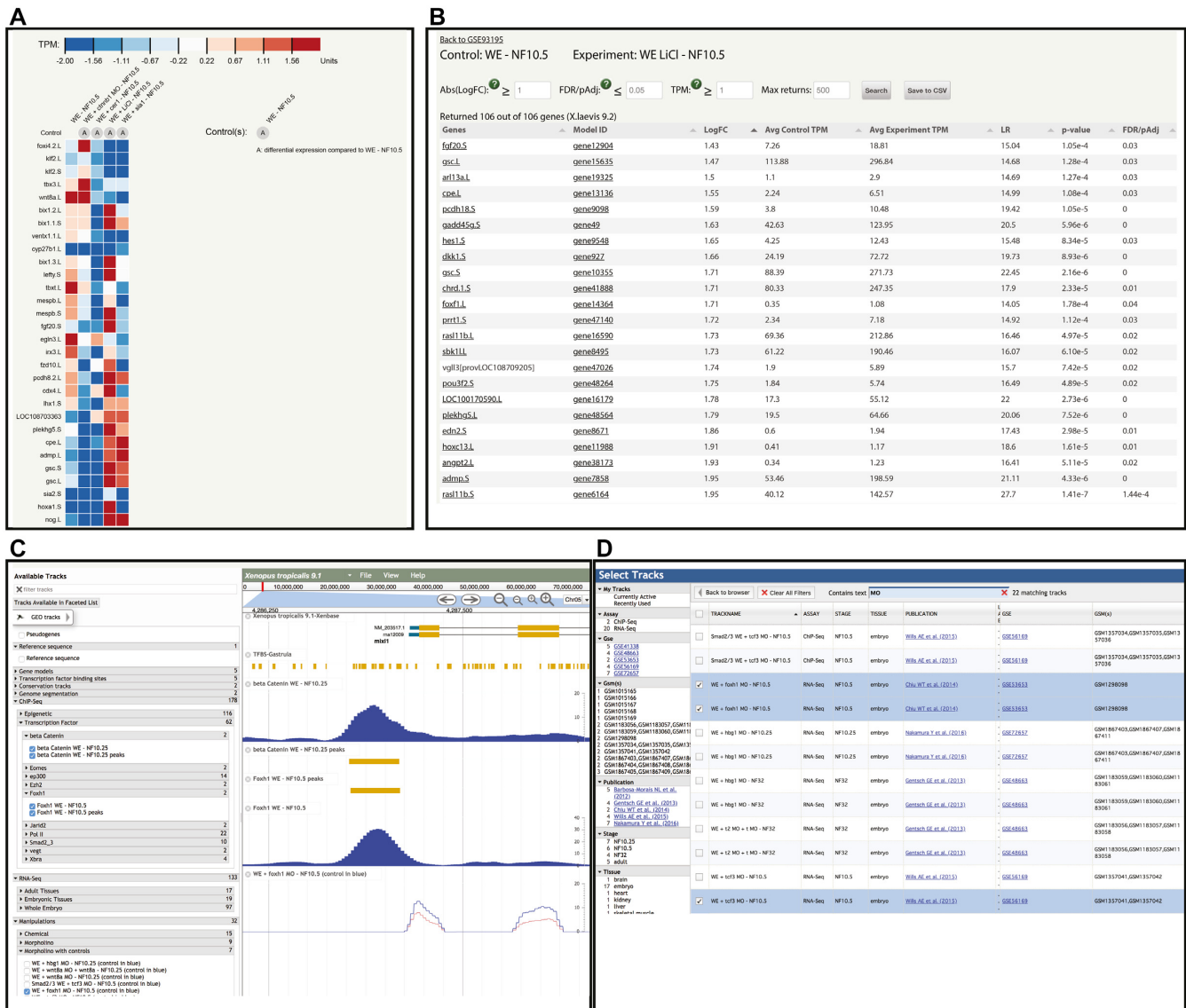


Figure 3. Visualization of GEO data. (A) An example of a heatmap displaying TPM values for a set of six experiments from within a single GSE. This view is generated using the check boxes described in Figure 2B. Users can change the color map, switch to sorting results by Log fold change, FDR and more using various tools in the interface. Cutoffs/thresholds can also be changed, as can the number of DEGs display. Mouse over will display the values for any individual tile. (B) The table view of a set of DEGs from a GSM, compared to its controls. Results can be sorted by any column (e.g. LogFC, FDR etc.) using the up or down arrowheads. (C) Results loaded into JBrowse displaying ChIP results as a bigWig, the corresponding peak call or RNA-Seq data illustrating the impact on a target gene of injecting a morpholino against foxH1. These data can be loaded using the ‘Load in JBrowse’ tool illustrated in Figure 2, or via the hierarchical menu on the left side of the JBrowse window. In this example results from multiple different studies are compared. (D) Alternatively tracks of interest can be loaded using the faceted track viewer available for both *Xenopus laevis* and *Xenopus tropicalis* genomes. In this example the ‘GEO tracks’ button in the top left of panel C was selected, then data filtered by entering MO into the ‘Contains text’ box to restricts tracks displayed to morpholino data. This method also allows users to combine results loaded via the GEO results interface to additional results from other studies (GSEs), which can also be loaded via the faceted track viewer. The selected tracks are viewed by clicking on the ‘Back to browser’ button. While different studies cannot be stringently compared, the same pipeline and thresholds were used across all dataset, so are useful for hypothesis generation and further experimentation.

page (Figure 2) as tab separated value files. As such they can be filterable based on TPMs levels, fold change or FDR using spreadsheet software or can be used in custom comparisons by users using their own software pipelines. The simplest method to view DEG sets is to select the ‘DEG’ link on a GSE page in a row corresponding to a specific experiment of interest. This will open a table-based listing of the complete DEG set using default cut-offs. The table is

arranged in columns, and users can select and sort the results using the up and down arrowheads in the header of each column. They can also change the thresholds applied to make the displayed results more or less stringent.

More dynamic and complex data displays are available via a set of JavaScript based visualization tools (Figure 3). Queries generated by the search interface pull the results from the TPM matrix and the associated statistics. D3

JavaScript code (D3.js; <https://d3js.org/>) then displays the results in dynamic displays that can be customized by users. Results for any combination of samples within the same GSE can be loaded and compared as TPM, log fold change or FDR heatmaps, parallel plots and hive plots. This option is available by selecting the check boxes of the samples of interest displayed for a GSE then clicking on the ‘Compare’ button. Once loaded, users can switch between the type of display plus alter the number of genes, selected thresholds etc. using the various drop downs and data windows provided. The order of the samples can be changed by selecting which dataset is displayed in the left column, and by representing this step the results can be setup to suit a user’s goals. All graphical tools allow users to download the results as a scalable vector graphics file for inclusion in presentations or publications, or a comma separated values file for offline analysis using a spreadsheet or other visualization/processing tools.

HELP WITH GEO CONTENT

To help users navigate the new GEO tools on Xenbase we have produced a number of video tutorials (<http://www.xenbase.org/other/static-xenbase/HowTo.jsp>). All the GEO pages have help tools visible on mouse-over to explain features and terms used. Throughout Xenbase, information icons, a lower case ‘i’ in a green circle, provide additional data upon mouse-over. At last, every Xenbase web page has a ‘Contact Us’ link in the top right corner that will lead to an email and on the lower navigation bar of every page a ‘Need Help?’ button links to a suite of documentation on how to use various Xenbase features.

THE FUTURE

Our plans for the immediate future are to leverage the extraordinarily rich data the DEG datasets from GEO provide. Each set of genes responding to an experimental intervention represent examples of ‘expression as a phenotype’—the experiment causing a change in gene expression that may help explain the phenotypic impact of the experiment. By utilizing stringent cut-offs Xenbase will produce a set of expression changes for every experiment processed from GEO, and integrate these results into gene and phenotype pages. These data will also be readily valuable for those performing gene regulatory network analyses. Our second immediate goal is to improve data visualization tools by allowing users to compare data from different GSEs, and more sophisticated options for on the fly clustering and sorting.

DATA AVAILABILITY

All tools described are accessible live on <https://www.xenbase.org/>.

The GEO processing bioinformatic pipeline CSBB is available in the GitHub repository at (<https://github.com/csbbcompbio/CSBB-v3.0>) and details available from the GitLab repository (<https://gitlab.com/Xenbase/bioinformatics/pipeline>). All processed data is available from the Xenbase FTP resource

at <http://xenbaseturbofrog.org/pub/genomics/GEO/> while individual GSE results can be viewed on GSE pages via direct links to the corresponding sub-directory.

Any additional code, for example the D3.js or the Xenbase schema, is available upon request.

ACKNOWLEDGEMENTS

We would like to thank our external advisory board and the *Xenopus* research community for feedback and guidance. In particular users and external resources (EBI/IntAct and Facebase) who beta-tested and provided feedback on the Xenbase GEO tools. Mutant line data on Xenbase is provided by the National Xenopus Resource RRID:SCR_013731.

FUNDING

Eunice Kennedy Shriver National Institute of Child Health and Human Development [P41HD064556]; Wellcome Trust via the EXRC [RRID:SCR_007164]. Funding for open access charge: Eunice Kennedy Shriver National Institute of Child Health and Human Development award [P41HD064556].

Conflict of interest statement. None declared.

REFERENCES

- Mitros,T, Lyons,J.B., Session,A.M., Jenkins,J., Shu,S., Kwon,T., Lane,M., Ng,C., Grammer,T.C., Khokha,M.K. *et al.* (2019) A chromosome-scale genome assembly and dense genetic map for *Xenopus tropicalis*. *Dev. Biol.*, **452**, 8–20.
- Session,A.M., Uno,Y., Kwon,T., Chapman,J.A., Toyoda,A., Takahashi,S., Fukui,A., Hikosaka,A., Suzuki,A., Kondo,M. *et al.* (2016) Genome evolution in the allotetraploid frog *Xenopus laevis*. *Nature*, **538**, 336–343.
- Karimi,K., Fortriede,J.D., Lotay,V.S., Burns,K.A., Wang,D.Z., Fisher,M.E., Pells,T.J., James-Zorn,C., Wang,Y., Ponferrada,V.G. *et al.* (2018) Xenbase: a genomic, epigenomic and transcriptomic model organism database. *Nucleic Acids Res.*, **46**, D861–D868.
- James-Zorn,C., Ponferrada,V.G., Fisher,M.E., Burns,K.A., Fortriede,J.D., Segerdell,E., Karimi,K., Lotay,V.S., Wang,D.Z., Chu,S. *et al.* (2018) Navigating Xenbase: An Integrated *Xenopus* Genomics and Gene Expression Database. *Methods Mol. Biol.*, **1757**, 251–305.
- Nenni,M.J., Fisher,M.E., James-Zorn,C., Pells,T.J., Ponferrada,V.G., Chu,S., Fortriede,J.D., Burns,K.A., Wang,Y., Lotay,V.S. *et al.* (2019) Xenbase: facilitating the use of *Xenopus* to model human disease. *Front. Physiol.*, **10**, 154.
- Barrett,T., Wilhite,S.E., Ledoux,P., Evangelista,C., Kim,I.F., Tomashevsky,M., Marshall,K.A., Phillippy,K.H., Sherman,P.M., Holko,M. *et al.* (2013) NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.*, **41**, D991–D995.
- Davies,C.A., Hitz,B.C., Sloan,C.A., Chan,E.T., Davidson,J.M., Gabdank,I., Hilton,J.A., Jain,K., Baymuradov,U.K., Narayanan,A.k. *et al.* (2018). The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.*, **46**, D794–D801.
- Segerdell,E., Ponferrada,V.G., James-Zorn,C., Burns,K.A., Fortriede,J.D., Dahdul,W.M., Vize,P.D. and Zorn,A.M. (2013) Enhanced XAO: the ontology of *Xenopus* anatomy and development underpins more accurate annotation of gene expression and queries on Xenbase. *J. Biomed. Semantics*, **4**, 31.
- Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
- Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G., Durbin,R. and Genome Project Data Processing, S. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

11. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
12. Li, B. and Dewey, C.N. (2011) RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics*, **12**, 323.
13. Risso, D., Ngai, J., Speed, T.P. and Dudoit, S. (2014) Normalization of RNA-Seq data using factor analysis of control genes or samples. *Nat. Biotechnol.*, **32**, 896–902.
14. Ramirez, F., Dunder, F., Diehl, S., Gruning, B.A. and Manke, T. (2014) deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.*, **42**, W187–W191.
15. Buels, R., Yao, E., Diesh, C.M., Hayes, R.D., Munoz-Torres, M., Helt, G., Goodstein, D.M., Elsik, C.G., Lewis, S.E., Stein, L. *et al.* (2016) JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.*, **17**, 66.