

FANTOM enters 20th year: expansion of transcriptomic atlases and functional annotation of non-coding RNAs

Imad Abugessaisa¹, Jordan A. Ramilowski^{1,2}, Marina Lizio¹, Jesicca Severin¹, Akira Hasegawa¹, Jayson Harshbarger¹, Atsushi Kondo¹, Shuhei Noguchi¹, Chi Wai Yip¹, Jasmine Li Ching Ooi¹, Michihira Tagami¹, Fumi Hori¹, Saumya Agrawal¹, Chung Chau Hon¹, Melissa Cardon¹, Shuya Ikeda³, Hiromasa Ono³, Hidemasa Bono^{3,4}, Masaki Kato¹, Kosuke Hashimoto¹, Alessandro Bonetti^{1,5}, Masaki Kato⁶, Norio Kobayashi⁶, Jay Shin¹, Michiel de Hoon¹, Yoshihide Hayashizaki¹, Piero Carninci¹, Hideya Kawaji^{1,7,8,*} and Takeya Kasukawa^{1,9,*}

¹RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa, Japan, ²Advanced Medical Research Center, Yokohama City University, Kanagawa, Japan, ³Database Center for Life Science, Research Organization of Information and Systems, Mishima, Shizuoka, Japan, ⁴Program of Biomedical Science, Graduate School of Integrated Sciences for Life, Hiroshima University, ⁵Karolinska Institutet, Stockholm, Sweden, ⁶RIKEN Head Office for Information Systems and Cybersecurity, Wako, Saitama, Japan, ⁷RIKEN Preventive Medicine & Diagnosis Innovation Program, Wako, Saitama, Japan, ⁸Research Center for Genome & Medical Sciences, Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan and ⁹Institute for Protein Research, Osaka University, Suita, Osaka, Japan

Received September 15, 2020; Revised October 16, 2020; Editorial Decision October 19, 2020; Accepted October 21, 2020

ABSTRACT

The Functional ANnoTation Of the Mammalian genome (FANTOM) Consortium has continued to provide extensive resources in the pursuit of understanding the transcriptome, and transcriptional regulation, of mammalian genomes for the last 20 years. To share these resources with the research community, the FANTOM web-interfaces and databases are being regularly updated, enhanced and expanded with new data types. In recent years, the FANTOM Consortium's efforts have been mainly focused on creating new non-coding RNA datasets and resources. The existing FANTOM5 human and mouse miRNA atlas was supplemented with rat, dog, and chicken datasets. The sixth (latest) edition of the FANTOM project was launched to assess the function of human long non-coding RNAs (lncRNAs). From its creation until 2020, FANTOM6 has contributed to the research community a large dataset generated from the knock-down of 285 lncRNAs in human dermal fibroblasts; this is followed with extensive expression profiling and cellular phenotyping. Other

updates to the FANTOM resource includes the re-processing of the miRNA and promoter atlases of human, mouse and chicken with the latest reference genome assemblies. To facilitate the use and accessibility of all above resources we further enhanced FANTOM data viewers and web interfaces. The updated FANTOM web resource is publicly available at <https://fantom.gsc.riken.jp/>.

INTRODUCTION

This year marks the 20th year from the first publication by the FANTOM Consortium, where biology, genomics and bioinformatics experts from all over the world collaborated in providing the first collection and functional annotation of more than one million full-length mouse cDNAs (1). During the 20 years following the initial publication, the FANTOM consortium have not only contributed to, but also evolved genomics research. This was achieved by developing and applying novel technologies to create big datasets in each subsequent edition of the project (2). In the fifth edition of FANTOM (FANTOM5), we constructed 'atlases' of promoters (3,4), enhancers (5), miRNAs (6) and long non-coding RNAs (lncRNAs) (7) in several mammalian genomes. Recently, the sixth edition of

*To whom correspondence should be addressed. Tel: +81 45 503 9222; Fax: +81 45 503 9219; Email: takeya.kasukawa@riken.jp
Correspondence may also be addressed to Hideya Kawaji. Email: hideya.kawaji@riken.jp

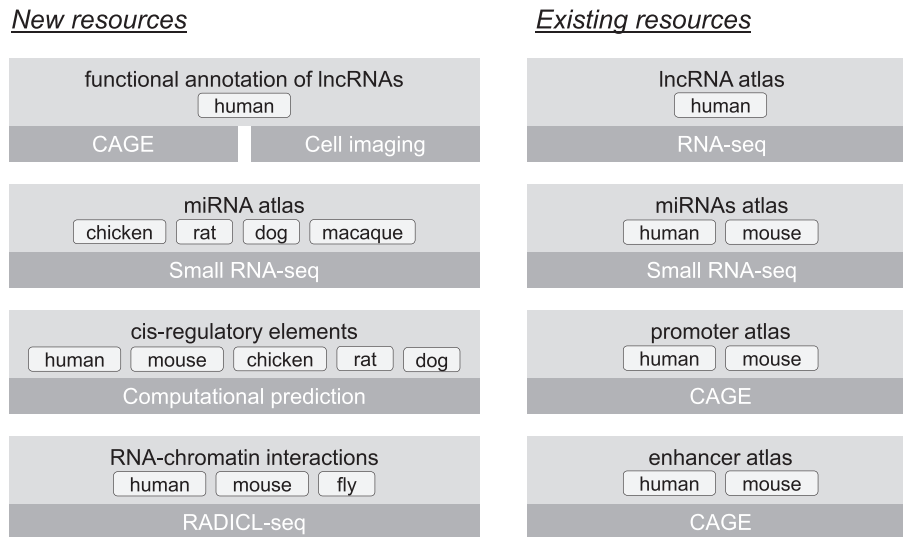


Figure 1. Overview of FANTOM5 and FANTOM6 web resources: new and existing resources generated in FANTOM5 and FANTOM6 projects. The light-gray boxes show the atlas or dataset and the dark-gray boxes show the method used to generate them.

FANTOM (FANTOM6) was launched to elucidate the biological functions of lncRNAs (8). lncRNAs constitute a major fraction of the transcriptome and can be broadly categorized into sense-intronic, antisense, intergenic, divergent and enhancer RNA regions (2). Although some of the lncRNAs characteristics such as their rapid evolution were already known (9,10), conservation of lncRNA sequences and genomic position are suggestive of functional potential (11–13). Currently, only a relatively small number of lncRNAs have any experimentally confirmed functions, compared to the tens of thousands of lncRNAs reported in the literature and available in databases such as Ensemble (14) or GENCODE (15). In the pilot phase of FANTOM6, we used a high-throughput screening (HTS) platform to systematically knockdown 285 lncRNAs in human dermal fibroblast cells and profiled the molecular phenotypes by cap analysis of gene expression (CAGE) (16) protocol; followed then by downstream bioinformatics analyses.

Here, we report the latest updates of the FANTOM web resource. Figure 1 shows the summary of the FANTOM web resource from FANTOM5 and FANTOM6 projects including previously reported and currently updated resources (17,18), as well as newly-released datasets (8,16,17). Mainly, we provide new non-coding RNA (ncRNA) resources including the functional annotation of 285 human lncRNAs from the FANTOM6 pilot study (8), along with the miRNA atlases of rat, dog, chicken and macaque (19). Each miRNA atlas was also supplemented with computationally-predicted motif activities, genome-wide transcription factor binding sites (TFBS) and multiple genomic alignments of human, mouse, rat, chicken and macaque. Another newly-added resource, is a dataset generated using RADICL-seq, a recently published method to identify genome-wide RNA-chromatin interactions (20), which can potentially make contributions to inferring the functions of lncRNAs. Moreover, the FANTOM5 sample metadata was further curated and improved upon to include

additional ontology terms. In parallel, the existing miRNA and CAGE datasets for several species (human, mouse and chicken) were remapped to their latest genome assemblies. Finally, we enhanced the web resources and upgraded user interfaces for better performance and usability.

RESULTS

Functional annotation of 285 lncRNAs in human fibroblast cells

The major goal of FANTOM6 is to systematically elucidate the function of lncRNAs in the human genome. For this, we used LNA antisense GapmeR technology to suppress the expression levels of a set of known and novel lncRNAs (7) in different cell types. In the pilot project (8) we selectively targeted 285 lncRNAs in human fibroblast cells using 5–15 non-overlapping antisense oligos (ASOs) in an automated high-throughput cell culture platform. We then evaluated the cellular phenotypic changes using Incucyte® live imaging and found that over 25% of the targeted lncRNAs affected cell growth and morphology, as well as cell migration (8). For the subset of 154 lncRNAs with high knockdown efficiency (generally >50%), we used deep-CAGE sequencing and profiled transcriptomes of 970 knock-down and control libraries. We then assessed the functional implications of the 340 targeted individual ASOs by performing differential gene expression, Motif Activity Response Analysis (MARA) (21), and Gene Set Enrichment Analysis (GSEA) (22). The resulting molecular phenotype recapitulated the observed cellular phenotypes while providing additional insights into the affected genes and pathways. In summary, we can now disseminate the largest (to-date) lncRNA knockdown dataset with molecular phenotyping (Figure 2A–D); this can be further explored using our interactive portal (Figure 2E). The interfaces and dataset download can be found at <https://fantom.gsc.riken.jp/6/>.

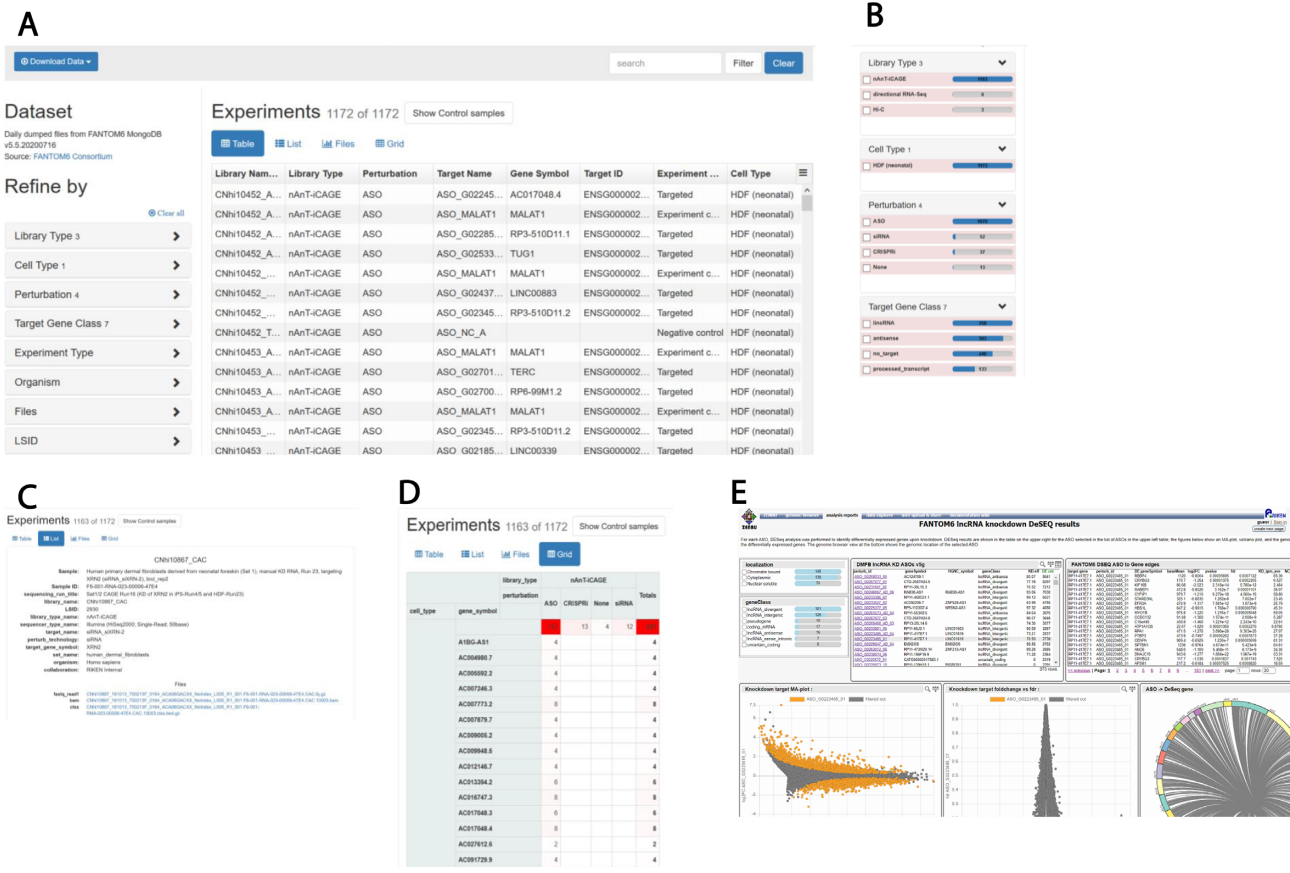


Figure 2. Access to the FANTOM6 data web resources: (A) web resources interface for FANTOM6 experiments in tabular view for each CAGE library with associated attributes; (B) web resource experiments view that can be refined by the check-box and adjusted; (C and D) provide a list, files and a grid view, respectively, of selected experiments as in (B); (E) analysis results of FANTOM6 pilot project using ZENBU analysis report view.

miRNAs atlas of chicken, rat, dog and macaque

miRNAs are a class of small non-coding RNAs that regulate mRNAs, post-transcriptionally, by either inhibiting their translation or by inducing their degradation (23). To profile the expression of miRNAs in different cell types, FANTOM5 generated short RNA (sRNA) sequencing data for 386 human and 42 murine RNA samples, most of which were obtained from primary cells (6). Additional sRNA sequencing data, produced in rat, dog and chicken for aortic smooth muscle primary cells; this allowed for an evolutionary analysis of miRNA expression patterns in this cell type (19). The data, together with FANTOM5 sRNA sequencing data of universal RNA samples, were further used to predict candidate novel miRNAs in these three species (19). All sequencing data mapped to genome assemblies rn6 (rat), canFam3 (dog) and galGal5 (chicken) are available as BAM files from <https://fantom.gsc.riken.jp/5/data/>. The FANTOM5 miRNA expression atlas can be accessed through a newly created visualization system at https://fantom.gsc.riken.jp/zenbu/reports/#FANTOM_miRNA_atlas. In addition to the expression profiles and associated annotations of each miRNA, the atlas also provides the putative promoter of the corresponding primary miRNA transcript, as identified by analysing FANTOM5 CAGE data, in each of five species (Figure 3). A list of samples used in this atlas and

their associated metadata was made available via SSTAR (Semantic catalog of Samples, Transcription initiation And Regulators), <https://fantom.gsc.riken.jp/5/sstar/> a semantic database of FANTOM5 samples (24).

Putative promoters in rat, dog and chicken

In the FANTOM5 project, promoters were revealed by identifying peaks in transcription initiation signals captured by CAGE. For quantitative analyses, a strict (*robust*) threshold was applied to expression levels of actively transcribed regions, whereas a more relaxed (*permissive*) threshold was used for qualitative analysis aimed at covering lowly expressed regions. Here, we release new sets of CAGE peaks for rat, dog, and chicken, based on the permissive threshold and following the same approach and parameters as in our previous study (4). These new data sets were utilized for identification of enhancers and miRNA promoters (16) in the three species, and made available for further analysis at <https://zenodo.org/record/3856413>.

Multiple genome alignments, genome-wide TFBS predictions and motif activity analysis

Cis-regulatory elements, in the promoter region of each gene, are a key contributor to the variation in gene expres-

A

Browse samples Chicken

Details of profiled samples (incl. tissue, primary cells)

Filter by sample category
 primary cells tissues

Copy Excel CSV PDF Show 25 entries Search:

FF ontology id	Name	Sample category
10004-101A9	chicken embryo whole body, day10 (HH37)	tissues
10210-103G3	chicken embryo whole body, 01hr30min (HH1)	tissues
10211-103G4	chicken embryo whole body,05hr30min (HH1)	tissues
10212-103G5	chicken embryo whole body,06hr (HH1-2)	tissues
10213-103G6	chicken embryo whole body,14hr (HH3)	tissues
10214-103G7	chicken embryo whole body,18hr (HH4)	tissues
10215-103G8	chicken embryo whole body, 22hr (HH5-6)	tissues
10216-103G9	chicken embryo whole body, day01 (HH7 1-3 somites)	tissues
10217-103H1	chicken embryo whole body, day01 (HH7 2-3 somites)	tissues

B

ZENBU genome browser analysis reports data explorer user upload & share documentation wiki

miRNA expression atlas - chicken

Chicken miRNA promoters

miRNA	premiRNA	promoter
gga-let-7a-2-3p*	gga-let-7a-2	chr24:3228348..3228362,+
gga-let-7a-3p*	gga-let-7a-3	Unknown
gga-let-7a-5p	gga-let-7a-2	chr24:3228348..3228362,+
gga-let-7a-5p	gga-let-7a-1	chr12:6100319..6100348,-
gga-let-7a-5p	gga-let-7a-3	Unknown
gga-let-7b	gga-let-7b	Unknown
gga-let-7c-3p*	gga-let-7c	chr1:98950082..98950097,+
gga-let-7c-5p*	gga-let-7c	chr1:98950082..98950097,+
gga-let-7d	gga-let-7d	chr12:6100319..6100348,-
gga-let-7f-3p*	gga-let-7f	chr12:6100319..6100348,-
gga-let-7f-5p	gga-let-7f	chr12:6100319..6100348,-
gga-let-7g-3p*	gga-let-7g	chr12:2978283..2978315,-
gga-let-7g-5p	gga-let-7g	chr12:2978283..2978315,-
gga-let-7i	gga-let-7i	Unknown
gga-let-7j-3p*	gga-let-7j	chr26:1644722..1644743,-
gga-let-7j-5p	gga-let-7j	chr26:1644722..1644743,-
gga-let-7k-3p*	gga-let-7k	chr26:1644722..1644743,-
gga-let-7k-5p	gga-let-7k	chr26:1644722..1644743,-
gga-miR-100-3p*	gga-mir-100	chr24:3228348..3228362,+
gga-miR-100-5p	gga-mir-100	chr24:3228348..3228362,+
gga-miR-101-1-5p*	gga-mir-101-1	chr2:27481270..27481318,+
gga-miR-101-2-5p*	gga-mir-101-2	chr8:28363254..28363290,-
gga-miR-101-3p	gga-mir-101-2	chr8:28363254..28363290,-
gga-miR-101-3p	gga-mir-101-1	chr2:27481270..27481318,+
gga-miR-103-1-5p*	gga-mir-103-1	chr13:4682767..4682805,+
gga-miR-103-2-5p*	gga-mir-103-2	chr4:89103235..89103280,-
gga-miR-103-3p	gga-mir-103-1	chr13:4682767..4682805,+
gga-miR-103-3p	gga-mir-103-2	chr4:89103235..89103280,-
gga-miR-106-3p*	gga-mir-106	chr4:3997212..3997226,-
gga-miR-106-5p	gga-mir-106	chr4:3997212..3997226,-

Chicken CAGE **Chicken shortRNA**

CAGE expression :

miRNA	tpm	sample description
gga-let-7c-5p	9.501	chicken embryo whole body, day07 (HH32)
gga-let-7c-5p	8.484	Aortic Smooth Muscle cells, donor3
gga-let-7c-5p	5.609	chicken embryo whole body, day10 (HH37)
gga-let-7c-5p	4.813	Aortic Smooth Muscle cells, donor1
gga-let-7c-5p	4.792	Aortic Smooth Muscle cells, donor2
gga-let-7c-5p	4.292	chicken embryo whole body, day05 (HH25-27)
gga-let-7c-5p	1.409	chicken embryo whole body, day15 (HH41)
gga-let-7c-5p	1.354	chicken embryo whole body, day20 (HH45)
gga-let-7c-5p	1.020	Mesenchymal stem cells - bone marrow derived, donor1
gga-let-7c-5p	0.8322	chicken embryo whole body, day04 (HH23)
gga-let-7c-5p	0.3582	chicken embryo whole body, day03 (HH19)
gga-let-7c-5p	0.2654	chicken embryo whole body, 14hr (HH3)
gga-let-7c-5p	0.1229	chicken leg buds, day03 (HH20)
gga-let-7c-5p	0.07854	chicken embryo whole body, 01hr30min (HH1)
gga-let-7c-5p	0	chicken embryo whole body, 05hr30min (HH1)
gga-let-7c-5p	0	chicken embryo whole body, 06hr (HH1-2)

32 rows

CAGE expression :

Figure3. Access to the miRNA atlas sample metadata and analysis results: (A) the list of chicken, rat and dog samples and their metadata are provided in FANTOM5 SSTAR resource; (B) ZENBU report for miRNA atlas allows for interactive browsing of miRNAs detected in given samples and for exploring their promoter(s) expression. The raw reads and alignments are available through the FANTOM5 data files repository.

sion observed between different cell types (4) and species (19). We used MotEvo (25) to generate genome-wide predictions of transcription factor binding sites (TFBSs), using position-weight matrix models obtained from SwissRegulon (26). The evolutionary model used by MotEvo requires a multiple genome alignment, which was available from existing sources for human and mouse, but was generated *ab initio* for rat, dog, and chicken, each of these against 29 vertebrate species (19). These multiple genome alignments, as well as the genome-wide TFBS predictions, for human (genome assembly hg19 and hg38), mouse (mm9

and mm10), rat (rn6), dog (canFam3) and chicken (galGal5 and galGal6) (19) are available at <https://fantom.gsc.riken.jp/5/suppl/Alam.et.al.2020/>. The genome-wide TFBS predictions can be used in a wide variety of bioinformatics analyses, including motif activity analysis; a methodology to infer the regulatory activity of transcription factors (TFs) by decomposing CAGE expression levels of promoters based on the presence of TF binding sites in each promoter (21). Scripts to perform motif activity analysis, using the provided genome-wide TFBS predictions, are available at <https://fantom.gsc.riken.jp/5/suppl/Alam.et.al.2020/>.

Human, mouse and *Drosophila* dataset for genome-wide RNA-chromatin interactions by RADICL-seq

FANTOM projects identified a large number of lncRNAs in the mammalian genome (7,11). Many of these RNAs are found in the cell nucleus and are bound to the chromatin. However, it is not yet understood which RNAs and what DNA genomic regions interact specifically with each other in different cell types. To better understand these interactions, we developed a new technology called RNA And DNA Interacting Complexes Ligated and sequenced (RADICL-seq) (17), which maps genome-wide RNA-chromatin interactions in intact nuclei. For the development of the technology, we generated and sequenced mouse ES cell and OPC (oligodendrocyte progenitor cells) libraries. Additionally, RADICL-seq sequencing data were produced in human iPSC and *Drosophila* S2 DRSC cells as a pilot experiment to expand the project and to ultimately uncover RNA-chromatin interactomes in multiple cell types. All sequencing data, mapped to genome assemblies hg38 (human), mm10 (mouse), and dm6 (*Drosophila*) are available as BEDPE files from <https://fantom.gsc.riken.jp/6/datafiles/RADICL-Seq/>.

Reprocessing of the miRNA atlases with the latest reference genome assemblies

The original FANTOM5 human and mouse miRNA atlases (6) were based on the GRCh37 assembly for human and the GRCm37 for mouse. Just as we processed our human and mouse promoter atlases (27), we remapped all human and mouse short RNA-seq libraries to the latest GRCh38 and GRCm38 reference genome assemblies; provided by the Genome Reference Consortium (<https://www.ncbi.nlm.nih.gov/grc>). The CAGE and short RNA-seq alignments for chicken, which were previously based on the galGal5 reference genome, were remapped to the latest galGal6 genome assembly. All reprocessed datasets are available in our 'reprocessed' section at (<https://fantom.gsc.riken.jp/5/datafiles/reprocessed/>).

Expansion of the ontology annotations in sample metadata

We revised the metadata for FANTOM5 human and mouse samples to better facilitate the usage of the corresponding datasets. The update includes additional annotations of cancer cell line samples based on NCI Thesaurus (<https://ncit.nci.nih.gov/>), as well as refined annotations of cell and tissue samples based on Cell Ontology (28) and UBERON (29) respectively. These updates also improve the interconnectivity to the RefEx database (<https://refex.dbcls.jp/>) (30). The revised metadata files are available on our web site at (https://fantom.gsc.riken.jp/5/datafiles/reprocessed/metadata_latest/).

Updates of viewers and interfaces

To serve the research community via FANTOM web portals and databases (31), the FANTOM web-based interfaces are continually updated to provide reliable and user-friendly resources. In the latest update, the metadata for the rat, dog, chicken and macaque libraries used in a report by Alam, T

et al. (16) were added to the FANTOM5 SSTAR at <https://fantom.gsc.riken.jp/5/sstar/>. ZENBU genome browser, an interactive tool for visualization and analysis of large-scale sequencing datasets (32), provides a genomic view of several FANTOM atlases; it was updated to accommodate recently published data for macaque, see <http://fantom.gsc.riken.jp/zenbu/gLyphs/#config=TJD69JeXM5ylAJTRKiJ45D>.

FANTOM5 utilized the UCSC Genome Browser data hub (33) and setup the FANTOM5 data hub, which provided the TSS activities in individual biological states, and the corresponding identified regions in human and mouse genomes <https://fantom.gsc.riken.jp/5/datahub/description.html>.

Recently, the FANTOM5 data hub was updated by adding individual tracks for chicken, dog, rat and macaque using aforementioned dataset updates, as described in (16).

The FANTOM5 metadata are exported to the RIKEN MetaDatabase (<https://metadb.riken.jp/>) (34). Here, the metadata for RIKEN's life-science databases are integrated and openly provided to researchers world-wide with the semantic web technology. The FANTOM5 metadata in the RIKEN MetaDatabase is available at <https://metadb.riken.jp/metadb/db/fantom5>.

Conclusion and future updates

The latest updates of the FANTOM web resource consist mainly of the enhancement and expansion of ncRNA resources. These updates cover the latest published functional annotations of 285 human lncRNAs; the miRNAs atlas of rat, dog, chicken and macaque; and preliminary results of genome-wide RNA-chromatin interaction with the new RADICAL-seq protocol; and aim to contribute to uncovering the functions of ncRNAs. Moreover, the FANTOM web resource has also been updated by (i) adding computationally predicted resources of motif activity; genome-wide TFBS and multiple genome alignments for human, mouse, chicken, rat and macaque; (ii) reprocessing of the existing atlases and datasets based on the latest-available genome assemblies; (iii) further curating metadata for FANTOM5 sample ontologies; and (iv) enhancing existing web applications and user interfaces. These updates will help researchers to better understand transcriptomes of these profiled species, facilitate cross-species studies and improve our understanding of the biological functions of mammalian ncRNAs. The contribution of the FANTOM consortium to genomics research will further continue by providing additional datasets and resources to enhance our understanding of mammalian genomes and the function of lncRNAs, including those from the ongoing FANTOM6 projects.

ACKNOWLEDGEMENTS

The authors would like to thank all the FANTOM consortium members for their fruitful collaboration, Prof. A.R.R. Forrest for scientific coordination of the FANTOM5 project, and the RIKEN GeNAS facility (ceased in 2018) for the library construction, sequencing and subsequent processing of the FANTOM data. We greatly appreciate efforts of Nobuyu Takeda, Teruaki Kitakura and Akira Furukawa in providing FANTOM projects with technical

support and the IT infrastructure. We are thankful for the English proof reading by Scott Walker.

FUNDING

The FANTOM5 and FANTOM6 projects were supported by a Research Grant for RIKEN Center for Life Science Technology, Division of Genomic Technologies (CLST DGT); RIKEN Center for Integrative Medical Sciences (IMS) from MEXT, Japan; the FANTOM5 was also supported by the following grants: Research Grant for RIKEN Omics Science Center from MEXT (to Y.H.); Grant of the Innovative Cell Biology by Innovative Technology (Cell Innovation Program) from the MEXT (to Y.H.); Research Grant to RIKEN Preventive Medicine and Diagnosis Innovation Program from MEXT (to Y.H.); JSPS Fellowship [S19058 to L.M.]; we gratefully acknowledge the computational resources of the HOKUSAI supercomputer system provided by RIKEN [Q18305/Q19305], which enabled us to perform the pairwise genome alignments as well as the genome-wide TFBS predictions by MotEvo; the FANTOM5 metadata update was supported by the ROIS-DS-JOINT 2019 grant [015RP2019]; the RIKEN Metadatabase export was supported by the grant by RIKEN Open life science platform. Funding for open access charge: MEXT Kofukin [to RIKEN].

Conflict of interest statement. None declared.

REFERENCES

- Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H. *et al.* (2001) Functional annotation of a full-length mouse cDNA collection. *Nature*, **409**, 685–690.
- de Hoon, M., Shin, J.W. and Carninci, P. (2015) Paradigm shifts in genomics through the FANTOM projects. *Mamm. Genome*, **26**, 391–402.
- Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmid, C., Suzuki, T. *et al.* (2014) An atlas of active enhancers across human cell types and tissues. *Nature*, **507**, 455–461.
- Forrest, A.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J., Haberle, V., Lassmann, T., Kulakovskiy, I.V., Lizio, M., Itoh, M. *et al.* (2014) A promoter-level mammalian expression atlas. *Nature*, **507**, 462–470.
- Arner, E., Daub, C.O., Vitting-Seerup, K., Andersson, R., Lilje, B., Drablos, F., Lennartsson, A., Rønnerblad, M., Hrydziuszko, O., Vitezic, M. *et al.* (2015) Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science*, **347**, 1010–1014.
- De Rie, D., Abugessaisa, I., Alam, T., Arner, E., Arner, P., Ashoor, H., Åström, G., Babina, M., Bertin, N., Burroughs, A.M. *et al.* (2017) An integrated expression atlas of miRNAs and their promoters in human and mouse. *Nat. Biotechnol.*, **35**, 872–878.
- Hon, C.C., Ramiłowski, J.A., Harshbarger, J., Bertin, N., Rackham, O.J., Gough, J., Denisenko, E., Schmeier, S., Poulsen, T.M., Severin, J. *et al.* (2017) An atlas of human long non-coding RNAs with accurate 5' ends. *Nature*, **543**, 199–204.
- Ramiłowski, J.A., Yip, C.W., Agrawal, S., Chang, J.C., Ciani, Y., Kulakovskiy, I.V., Mendez, M., Ooi, J.L.C., Ouyang, J.F., Parkinson, N. *et al.* (2020) Functional annotation of human long noncoding RNAs via molecular phenotyping. *Genome Res.*, **30**, 1060–1072.
- Sharma, H. and Carninci, P. (2020) The secret life of lncRNAs: conserved, yet not conserved. *Cell*, **181**, 512–514.
- Guo, C.J., Ma, X.K., Xing, Y.H., Zheng, C.C., Xu, Y.F., Shan, L., Zhang, J., Wang, S., Wang, Y., Carmichael, G.G. *et al.* (2020) Distinct processing of lncRNAs contributes to non-conserved functions in stem cells. *Cell*, **181**, 621–636.
- Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C. *et al.* (2005) The transcriptional landscape of the mammalian genome. *Science*, **309**, 1559–1563.
- Hezroni, H., Koppstein, D., Schwartz, M.G., Avrutin, A., Bartel, D.P. and Ulitsky, I. (2015) Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep.*, **11**, 1110–1122.
- Ulitsky, I. (2016) Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nat. Rev. Genet.*, **17**, 601–614.
- Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amode, M.R., Armean, I.M., Bennett, R., Bhai, J., Billis, K., Boddu, S. *et al.* (2019) Ensembl 2019. *Nucleic Acids Res.*, **47**, D745–D751.
- 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.
- Kanamori-Katayama, M., Itoh, M., Kawaji, H., Lassmann, T., Katayama, S., Kojima, M., Bertin, N., Kaiho, A., Ninomiya, N., Daub, C.O. *et al.* (2011) Unamplified cap analysis of gene expression on a single-molecule sequencer. *Genome Res.*, **21**, 1150–1159.
- Lizio, M., Harshbarger, J., Abugessaisa, I., Noguchi, S., Kondo, A., Severin, J., Mungall, C., Arenillas, D., Mathelier, A., Medvedeva, Y.A. *et al.* (2017) Update of the FANTOM web resource: high resolution transcriptome of diverse cell types in mammals. *Nucleic Acids Res.*, **45**, D737–D743.
- Lizio, M., Abugessaisa, I., Noguchi, S., Kondo, A., Hasegawa, A., Hon, C.C., de Hoon, M., Severin, J., Oki, S., Hayashizaki, Y. *et al.* (2018) Update of the FANTOM web resource: expansion to provide additional transcriptome atlases. *Nucleic Acids Res.*, **47**, D752–D758.
- Alam, T., Agrawal, S., Severin, J., Young, R.S., Andersson, R., Arner, E., Hasegawa, A., Lizio, M., Ramiłowski, J.A., Abugessaisa, I. *et al.* (2020) Comparative transcriptomics of primary cells in vertebrates. *Genome Res.*, **30**, 951–961.
- Bonetti, A., Agostini, F., Suzuki, A.M., Hashimoto, K., Pascarella, G., Gimenez, J., Roos, L., Nash, A.J., Ghilotti, M., Cameron, C.J.F. *et al.* (2020) RADICL-seq identifies general and cell type-specific principles of genome-wide RNA-chromatin interactions. *Nat. Commun.*, **11**, 1018.
- Suzuki, H., Forrest, A.R., van Nimwegen, E., Daub, C.O., Balwiercz, P.J., Irvine, K.M., Lassmann, T., Ravasi, T., Hasegawa, Y., de Hoon, M.J. *et al.* (2009) The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nat. Genet.*, **41**, 553–562.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. U.S.A.*, **102**, 15545–15550.
- Jonas, S. and Izaurralde, E. (2015) Towards a molecular understanding of microRNA-mediated gene silencing. *Nat. Rev. Genet.*, **16**, 421–433.
- Abugessaisa, I., Shimoji, H., Sahin, S., Kondo, A., Harshbarger, J., Lizio, M., Hayashizaki, Y., Carninci, P., Forrest, A., Kasukawa, T. *et al.* (2016) FANTOM5 transcriptome catalog of cellular states based on Semantic MediaWiki. *Database (Oxford)*, **2016**, baw105.
- Arnold, P., Erb, I., Pachkov, M., Molina, N. and van Nimwegen, E. (2012) MotEvo: integrated Bayesian probabilistic methods for inferring regulatory sites and motifs on multiple alignments of DNA sequences. *Bioinformatics*, **28**, 487–494.
- Pachkov, M., Balwiercz, P.J., Arnold, P., Ozonov, E. and van Nimwegen, E. (2013) SwissRegulon, a database of genome-wide annotations of regulatory sites: recent updates. *Nucleic Acids Res.*, **41**, D214–D220.
- Abugessaisa, I., Noguchi, S., Hasegawa, A., Harshbarger, J., Kondo, A., Lizio, M., Severin, J., Carninci, P., Kawaji, H. and Kasukawa, T. (2017) FANTOM5 CAGE profiles of human and mouse reprocessed for GRCh38 and GRCm38 genome assemblies. *Sci Data*, **4**, 170107.
- Diehl, A.D., Meehan, T.F., Bradford, Y.M., Brush, M.H., Dahdul, W.M., Dougall, D.S., He, Y., Osumi-Sutherland, D., Ruttnerberg, A., Sarntinvijai, S. *et al.* (2016) The Cell Ontology 2016: enhanced content, modularization, and ontology interoperability. *J. Biomed. Semantics*, **7**, 44.

29. Mungall,C.J., Torniai,C., Gkoutos,G.V., Lewis,S.E. and Haendel,M.A. (2012) Uberon, an integrative multi-species anatomy ontology. *Genome Biol.*, **13**, R5.
30. Ono,H., Ogasawara,O., Okubo,K. and Bono,H. (2017) RefEx, a reference gene expression dataset as a web tool for the functional analysis of genes. *Sci. Data*, **4**, 170105.
31. Abugessaisa,I., Noguchi,S., Carninci,P. and Kasukawa,T. (2017) The FANTOM5 computation ecosystem: genomic information hub for promoters and active enhancers. *Methods Mol. Biol.*, **1611**, 199–217.
32. Severin,J., Lizio,M., Harshbarger,J., Kawaji,H., Daub,C.O., Hayashizaki,Y., Bertin,N., Forrest,A.R. and Consortium,F. (2014) Interactive visualization and analysis of large-scale sequencing datasets using ZENBU. *Nat. Biotechnol.*, **32**, 217–219.
33. Casper,J., Zweig,A.S., Villarreal,C., Tyner,C., Speir,M.L., Rosenbloom,K.R., Raney,B.J., Lee,C.M., Lee,B.T., Karolchik,D. *et al.* (2018) The UCSC Genome Browser database: 2018 update. *Nucleic Acids Res.*, **46**, D762–D769.
34. Kobayashi,N., Kume,S., Lenz,K. and Masuya,H. (2018) RIKEN MetaDatabase: a database platform for health care and life sciences as a microcosm of linked open data cloud. *Int. J. Semant. Web Inf. Syst.*, **14**, 140–164.