Publicly open databases, software and protocols contributed by RIKEN IMS researchers (Summary) (2023/8)

Category	Title	Description (English)	URL	Related reference	Resources, related information, or comments
Database, dataset	FANTOM5	Atlases of mammalian promoters, enhancers, lncRNAs and miRNAs in FANTOM5	https://fantom.gsc.riken.jp/5/	https://doi.org/10.1186/s13059-014-0560-6 https://doi.org/10.1093/nar/gkaa1054	FANTOM5 Data resources: https://fantom.gsc.riken.jp/5/datafiles/
Database, dataset	FANTOM5 web resource	A suite of web interfaces and databases for accessing to the FANTOM5 dataset (SSTAR, FANTOM CAT Browser, FANTOM miRNA atlas, ZENBU (FANTOM5), Cell Connectome Visualization, TET)	https://fantom.gsc.riken.jp/5/	https://doi.org/10.1186/s13059-014-0560-6 https://doi.org/10.1093/nar/gkaa1054	FANTOM views: https://fantom.gsc.riken.jp/views/
Database, dataset	FANTOM6	Expression profiles with large-scale lncRNA KD experiments in FANTOM6	https://fantom.gsc.riken.jp/6/	https://doi.org/10.1093/nar/gkaa1054	FANTOM6 Data resources: https://fantom.gsc.riken.jp/6/datafiles/
Database, dataset	FANTOM6 web resource	A suite of web interfaces and databases for accessing to the FANTOM6 dataset (FANTOM6 Experimental Index, ZENBU-Reports (FANTOM6))	https://fantom.gsc.riken.jp/6/	https://doi.org/10.1093/nar/gkaa1054	FANTOM views: https://fantom.gsc.riken.jp/views/
Database, dataset	refTSS	Reference data set of transcriptoin start sites in human and mouse	https://reftss.riken.jp/	https://doi.org/10.1016/j.jmb.2019.04.045	refTSS data files: https://reftss.riken.jp/datafiles/
Database, dataset	SCPortalen	A cell-centric single-cell RNA-seq data repository leveraging cutting-edge QC techniques and primary processing	https://single-cell.riken.jp/	https://doi.org/10.1093/nar/gkx949	SCPortalen web site: http://single-cell.riken.jp/
Database, dataset	UCSC Genome Browser (asian mirror)	An office asian mirror of UCSC Genome Browser	https://genome-asia.ucsc.edu/		UCSC Genome Browser mirro sites: https://genome.ucsc.edu/mirror.html
Database, dataset	Single cell RNA-seq of Japanese COVID-19	Large-scale single cell RNA-seq resource of the Japanese COVID-19 patients	https://humandbs.biosciencedbc.jp/en/hum0197-v15#JGAS000543	https://doi.org/10.1038/s41588-023-01375-1 https://doi.org/10.1038/s41586-022-05163-5	
Database, dataset	Pheweb.jp	GWAS database of Biobank Japan and cross-population studies	https://pheweb.jp/	https://doi.org/10.1038/s41588-021-00931-x https://doi.org/10.1038/s41586-022-05163-5 https://doi.org/10.1038/s41588-022-01213-w https://doi.org/10.1038/s41467-022-32005-9 https://doi.org/10.1136/annrheumdis-2022-222460 https://doi.org/10.1038/s41467-021-21011-y https://doi.org/10.1038/s41562-019-0805-1 https://doi.org/10.1093/hmg/ddab361 https://doi.org/10.1038/nature12873 https://doi.org/10.1101/2021.09.03.21262975	
Database, dataset	JMAG (Japanese Metagenome Assembled Genomes Platform) JVD (Japanese Virus Database)	Gut metagenome and virome catalogue of the Japanese population	https://humandbs.biosciencedbc.jp/en/hum0197-v12#micro-mag https://humandbs.biosciencedbc.jp/en/hum0197-v12#micro-virus	https://doi.org/10.1016/j.xgen.2022.100241	
Database, dataset	Serum metabolome of Japanese	Serum metabolome data of the type 2 diabetes patients of Japanese	https://humandbs.biosciencedbc.jp/en/hum0372-v1 - jgas000572	https://doi.org/10.1038/s43856-022-00231-3	

Database, dataset	eQTL & sQTL databases of peripheral blood of Japanese COVID-19	eQTL & sQTL databases of peripheral blood obtained from the Japanese COVID-19	https://humandbs.biosciencedbc.jp/en/hum0343-v2 - qtl	https://doi.org/10.1038/s41467-022-32276-2	
Database, dataset	microRNA eQTL database of peripheral blood of Japanese	microRNA eQTL dabase of peripheral blood obtained from Japanese	https://humandbs.biosciencedbc.jp/en/hum0197-v15 - hum0197.v6.eqtl.v1	https://doi.org/10.1093/hmg/ddab361	
Database, dataset	KIR imputation reference panel	KIR gene variant imputation reference panel	https://humandbs.biosciencedbc.jp/en/hum0114-v3	https://doi.org/10.1016/j.xgen.2022.100101	
Software	Human reads in metagenome	Human read analysis tool in metagenome data	https://github.com/ytomofuji/Human_reads_in_metagenome		
Software	KIRAP (Killer Immunoglobulin-like Receptor variant Analytical Platform)	KIR gene variant calling pipeline	https://github.com/saorisakaue/KIR_project	https://doi.org/10.1016/j.xgen.2022.100101	
Software	OMARU (Omnibus Metagenome-wide Association study with RobUstness)	Metagenome-wide association study (MWAS) pipeline	https://github.com/toshi-kishikawa/OMARU	https://doi.org/10.1093/nargab/lqac019	
Software	Trans-Phar (Integration of TWAS and pharmacological database)	Drug discovery tool utilizing transcriptome-wide association study (TWAS)	https://github.com/konumat/Trans-Phar	https://doi.org/10.1093/hmg/ddab049	
Software	DEEP*HLA (DEEP learning for HLA allelic imputation)	Deep learning-based HLA imputation software	https://github.com/tatsuhikonaito/DEEP-HLA	https://doi.org/10.1038/s41467-021-21975-x	
Software	Obelisc (Observational linkage scan)	Non-parametric linkage analysis utilizing IBD stretch	https://github.com/qsonehara/Obelisc	https://doi.org/10.1093/bioinformatics/btaa940	
Software	GREP (Genome for REPositioning drugs)	Drug discovery tool utilizing disease risks genes	https://github.com/saorisakaue/GREP	https://doi.org/10.1093/bioinformatics/btz166	
Software	MIGWAS (miRNA-target gene networks enrichment on GWAS)	Biomarker microRNA screening tool based on GWAS results	https://github.com/saorisakaue/MIGWAS	https://doi.org/10.1093/nar/gky1066	
Software	eLD (entropy-based Linkage Disequilibrium index between multiallelic sites)	Entropy-based linkage disequilibrium (LD) index	www.sg.med.osaka-u.ac.jp/files/eLD.v1.0.zip	https://doi.org/10.1038/s41439-018-0030-x	
Software	Grimon (Graphical interface to visualize multi-layer omics networks)	Graphical interface to visualize multi-layer omics networks	https://github.com/mkanai/grimon	https://doi.org/10.1093/bioinformatics/bty488	
Software	MENTR	MENTR is a machine-learning model which reliably links genome sequence and ncRNA expression at the cell type level	https://github.com/koido/MENTR	https://doi.org/10.1038/s41551-022-00961-8	
Software	MOPLINE	MOPLINE is a computational algorithm (MOPline) that includes missing call recovery combined with high-confidence SV call selection and genotyping using short-read whole genome sequencing (WGS) data	https://github.com/stat-lab/MOPline		
Software	SCAFE	A software suite for analysis of transcribed cis-regulatory elements in single cells	https://github.com/chung-lab/SCAFE	https://doi.org/10.1093/bioinformatics/btac644	SCAFE source code on Github: https://github.com/chung-lab/SCAFE
Software	SkewC	SkewC is a novel quality-assessment method to identify poor quality or affecting single-cells in scRNA-seq experiments. The method relies on the measure of skewness of the gene coverage of each single cell as a quality measure	https://github.com/LSBDT/SkewC	https://doi.org/10.1016/j.isci.2022.103777	
Software	Moirai2	A scientific workflow system for large-scale biomedical datasets. Moirai2 is a fully reconstructed software from Moirai1 with more functionalities	https://github.com/moirai2/moirai2	https://doi.org/10.1186/1471-2105-15-144	
Software	HDRGenome	A tool to identify patient specirfic structure variants and deletions using whole genome sequence data with Hamming Distance Ratio (HDR) method, which is a collaborative development project with Dr. Okazaki in Juntendo University	https://github.com/LSBDT/HDRGenome		

Software	scDeepInsight	A computational tool for analyzing single-cell RNA-seq data to identify cell types	https://pypi.org/project/scdeepinsight/	https://doi.org/10.1093/bib/bbad266	Ready-to-use environment: https://hub.docker.com/r/shangrujia/scdeepinsight The entire code base, including the implementation and the pre-
Software	DeepInsight-3D	A computational tool for predicting response to anticancer drugs using multi-omics and deep learning	https://github.com/alok-ai-lab/DeepInsight3D_pkg	https://doi.org/10.1038/s41598-023-29644-3	https://github.com/alok-ai-lab/DeepInsight3D_pkg
Software	DeepFeature	A computational tool for extracting features from non-image data, such as genes, using deep learning	https://alok-ai-lab.github.io/deepfeature/	https://doi.org/10.1093/bib/bbab297	https://alok-ai-lab.github.io/deepfeature/
Software	DeepInsight	A computational analysis tool to transform non-image data such as genes for deep learning	https://alok-ai-lab.github.io/DeepInsight/	https://doi.org/10.1038/s41598-019-47765-6	http://emu.src.riken.jp/DeepInsight/
Software	IMSindel	A computational tool for analyzing RNA-seq data to detect intermediate size insertions and deletions	https://github.com/NCGG-MGC/IMSindel/	https://doi.org/10.1038/s41598-018-23978-z	https://github.com/NCGG-MGC/IMSindel/
Software	DRAGON	DRAGON: Matlab package of DRAGON clustering approach	http://emu.src.riken.jp/DRAGON/	https://doi.org/10.1186/s12859-017-1965-5	http://emu.src.riken.jp/DRAGON/
Software	2D-EM	We present 2D-EM, a clustering algorithm approach designed for small sample size and high-dimensional datasets. To employ information corresponding to data distribution and facilitate visualization, the sample is folded into its two-dimension (2D) matrix form (or feature matrix). The maximum likelihood estimate is then estimated using a modified expectation-maximization (EM) algorithm.	http://emu.src.riken.jp/2D-EM/	https://doi.org/10.1186/s12859-017-1970-8	http://emu.src.riken.jp/2D-EM/
Software	SIML	Considering the nature of biological data, we proposed a maximum likelihood clustering approach using a stepwise iterative procedure. The advantage of this proposed method is that it not only uses the distance information, but also incorporate variance information for clustering. This method is able to cluster when data appeared in overlapping and complex forms.	http://emu.src.riken.jp/SIML/	https://doi.org/10.1186/s12859-016-1184-5	http://emu.src.riken.jp/SIML/
Software	HML	Taking into consideration the nature of biological data, we propose a maximum likelihood clustering scheme based on a hierarchical framework. Results: This method can perform clustering even when the data belonging to different groups overlap. It can also perform clustering when the number of samples is lower than the data dimensionality.	http://emu.src.riken.jp/HML/	https://doi.org/10.1109/TBME.2016.2542212	http://emu.src.riken.jp/HML/
Software	MCV	Pipeline to identify somatic substitutions and indels with VAF (variant allele frequency) from whole genome sequencing	http://emu.src.riken.jp/MCV/	https://doi.org/10.1038/ng.3547	http://emu.src.riken.jp/MCV/
Software	ARETE	Arete - an analysis toolkit for network-based gene prioritisation	http://emu.src.riken.jp/arete/arete.html	https://doi.org/10.1186/s13040-017-0141-9	http://emu.src.riken.jp/arete/arete.html
Software	TFBIND	TFBIND : Software for searching transcription factor binding sites (including TATA boxes, GC boxes, CCAAT boxes, transcription start sites (TSS)). This tool uses weight matrix in transcription factor database TRANSFAC R.3.4 developed by Dr. Wingender et al, and the cut-offs originally estimated by our research.	https://tfbind.hgc.jp	https://doi.org/10.1093/bioinformatics/15.7.622	https://tfbind.hgc.jp
Software	VCMM	VCMM: Variant Caller with Multinomial probabilistic Model	http://emu.src.riken.jp/VCMM/	https://doi.org/10.1038/srep02161	http://emu.src.riken.jp/VCMM/
Software	NGS analyzer	Tools for identifying SNV (single nucleotide variations) and indels (insertions and deletions) from short read sequence data. Open source code for k computer. This program uses BWA for mapping, samtools for removing PCR duplication and VCMM (Shigemizu et al submitted) for variant calling. Developed by Tesuo Abe, Eiji Nakamura, Akihiro Fujimoto and Tatsuhiko Tsunoda.	http://emu.src.riken.jp/k_softwares/		http://emu.src.riken.jp/k_softwares/
Software	ExRAT	Tools for association study with interaction. Open source code for k and other computers. Developed by Eiji Nakamura, Reiichiro Nakamichi, Takashi Morizono, and Tatsuhiko Tsunoda.	http://emu.src.riken.jp/k_softwares/		http://emu.src.riken.jp/k_softwares/

Software	ParaHaplo	Tools for haplotype-based whole-genome association study using parallel computing. Open source code for k and other computers. Developed by Kazuharu Misawa.	http://emu.src.riken.jp/k_softwares/		http://emu.src.riken.jp/k_softwares/
Software	hzAnalyzer	hzAnalyzer: R and Java for Detection and Analysis of Contiguous Homozygosity in High-density Genotyping Datasets.	http://emu.src.riken.jp/hzAnalyzer/	https://doi.org/10.1186/gb-2011-12-3-r21	http://emu.src.riken.jp/hzAnalyzer/
Software	CisNetView	CisNetView: Tool to analyze gene networks by integrating genome-wide protein-protein, protein-DNA, and expression data.	http://emu.yokohama.riken.jp/CisNetView/cgi-bin/Draw_GRNet.cgi		http://emu.yokohama.riken.jp/CisNetView/cgi-bin/Draw_GRNet.cgi
Software	MotifCombinator	MotifCombinator: Tool to search for significant combinations of transcription factor-binding motifs, using regression methods.	http://emu.src.riken.jp/combinator	https://doi.org/10.1186/1471-2105-8-100	http://emu.src.riken.jp/combinator
Software	MOCSphaser	MOCSphaser: Tool to infer haplotypes composed of copy numbers alleles and SNP alleles.	http://emu.src.riken.jp/cgi-bin/Download_MOCSphaser.cgi	https://doi.org/10.1093/bioinformatics/btn242	http://emu.src.riken.jp/cgi-bin/Download_MOCSphaser.cgi
Software	CNVphaser	CNVphaser: Tool to infer CNV haplotypes with variations in both copy numbers and nucleotide sequences.	http://emu.src.riken.jp/cgi-bin/Download_CNVphaser.cgi	https://doi.org/10.1016/j.ajhg.2008.06.021	http://emu.src.riken.jp/cgi-bin/Download_CNVphaser.cgi
Software	CNVphaserPro	CNVphaserPro: CNV phasing tool from high-throughput data (with uncertainty).	https://figshare.com/articles/software/CNVphaserPro/22283992	https://doi.org/10.1534/g3.111.000174	https://figshare.com/articles/software/CNVphaserPro/22283992
Software	MS-DIAL	MS-DIAL was launched as a universal program for untargeted metabolomics that supports multiple instruments (GC/MS, GC/MS/MS, LC/MS, and LC/MS/MS) and MS vendors (Agilent, Bruker, LECO, Sciex, Shimadzu, Thermo, and Waters)	http://prime.psc.riken.jp/compms/msdial/main.html	https://doi.org/10.1038/s41587-020-0531-2	
Protocol	CAGE（Cap Analysis of Gene Expression）	A method for genome-wide identification of transcription start sites and simultaneously measure gene expression	https://fantom.gsc.riken.jp/protocols/basic.html	https://doi.org/10.1073/pnas.2136655100 https://doi.org/10.1038/nmeth0306-211 https://doi.org/10.1007/978-1-0716-1597-3_4	
Protocol	RADICL-seq	A method for comprehensively analyzing the interaction between RNA and chromatin localized in the nucleus to understand the function of RNA bound to chromatin		https://doi.org/10.1038/s41467-020-14337-6 https://doi.org/10.1093/nar/gkaa1054	FANTOM web resource: https://fantom.gsc.riken.jp/6/datafiles/RADICL-Seq/
Protocol	Retrotransposon capture	A method to enrich NGS libraries for repeat element sequences prior to sequencing in order to annotate non-reference repeat-mediated structural variants with downstream bioinformatic pipelines.		https://doi.org/10.1016/j.cell.2022.06.032 https://doi.org/10.1016/j.xpro.2022.102027 https://doi.org/10.1002/ctm2.1226	
Protocol	CapTrap long read RNA sequencing (CTR-seq)	A method for genome-wide long read full length cap trapped RNA sequencing with Oxford Nanopore Technology and PacBio.		https://doi.org/10.1038/s41467-021-23143-7 https://doi.org/10.1101/2023.06.16.543444	
Protocol	SINEUPs	A method for antisense long noncoding RNAs that enhance protein production.		https://dx.doi.org/10.3791/58627	
Protocol	SkewC-based scRNA-seq QC	Computational approach to evaluate scRNA-seq data quality and gene body coverage with SkewC		https://doi.org/10.1016/j.xpro.2022.102038	