OpenXGR: a web-server update for genomic summary data interpretation

Chaohui Bao[1,†], Shan Wang[1,†], Lulu Jiang[2,†], Zhongcheng Fang[3], Kexin Zou[4], James Lin[5], Saijuan Chen[1] and Hai Fang[1,*]

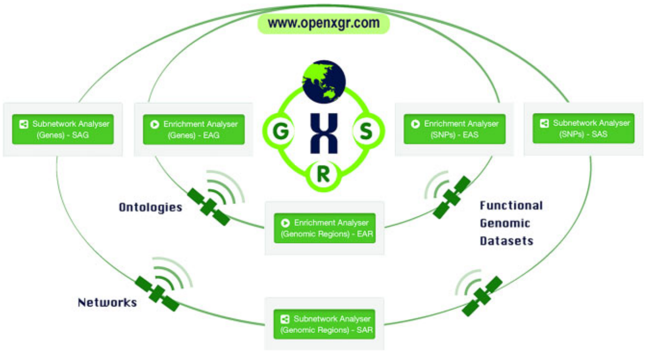
[1]Shanghai Institute of Hematology, State Key Laboratory of Medical Genomics, National Research Center for Translational Medicine at Shanghai, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China, [2]Translational Health Sciences, University of Bristol, Bristol BS1 3NY, UK, [3]Bioinformatics Department, School of Life Sciences and Technology, Tongji University, Shanghai 200092, China, [4]School of Life Sciences, Central South University, Hunan 410083, China and [5]High Performance Computing Center, Shanghai Jiao Tong University, Shanghai 200240, China

Received March 12, 2023; Revised April 21, 2023; Editorial Decision April 22, 2023; Accepted April 25, 2023

ABSTRACT

How to effectively convert genomic summary data into downstream knowledge discovery represents a major challenge in human genomics research. To address this challenge, we have developed efficient and effective approaches and tools. Extending our previously established software tools, we here introduce OpenXGR (http://www.openxgr.com), a newly designed web server that offers almost *real-time* enrichment and subnetwork analyses for a user-input list of genes, SNPs or genomic regions. It achieves so through leveraging ontologies, networks, and functional genomic datasets (such as promoter capture Hi-C, e/pQTL and enhancer-gene maps for linking SNPs or genomic regions to candidate genes). Six analysers are provided, each doing specific interpretations tailored to genomic summary data at various levels. Three enrichment analysers are designed to identify ontology terms enriched for input genes, as well as genes linked from input SNPs or genomic regions. Three subnetwork analysers allow users to identify gene subnetworks from input gene-, SNP- or genomic region-level summary data. With a step-by-step user manual, OpenXGR provides a user-friendly and all-in-one platform for interpreting summary data on the human genome, enabling more integrated and effective knowledge discovery.

GRAPHICAL ABSTRACT


www.openxgr.com
Subnetwork Analyser (Genes) - SAG
Enrichment Analyser (Genes) - EAG
G
X
S
R
Enrichment Analyser (SNPs) - EAS
Subnetwork Analyser (SNPs) - SAS
Ontologies
Enrichment Analyser (Genomic Regions) - EAR
Functional Genomic Datasets
Networks
Subnetwork Analyser (Genomic Regions) - SAR

INTRODUCTION

Human genomics research produces complex raw genomic data that can be simplified into summary-level data that capture essential information ready for sharing and mining. Without loss of generality, we define genomic summary data as a list of genes, SNPs or genomic regions, along with their summary statistics about the significance level (e.g. *P*-values) (1). Gene-level summary data are often generated from differential expression studies (2), SNP-level summary data from genome-wide association studies (3), and genomic region-level summary data from epigenomic studies (4,5). This simplification of data allows for more straightforward analyses, but how to effectively convert genomic summary data into downstream knowledge discovery remains one of the major challenges in human genomics research.

To address the challenges described above, we have developed e**X**ploring **G**enomic **R**elations or XGR (1) by demonstrating how ontologies enhance genomic summary data interpretation and how to enable insights at the gene

[*] To whom correspondence should be addressed. Hai Fang. Tel: +86 177 49786442; Email: fh12355@rjh.com.cn
[†] The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

subnetwork level. A dozen ontologies have been created to annotate genes regarding functions (6), phenotypes (7,8), diseases (9,10) and other attributes. By integrating a reference gene network that consolidates interaction knowledge (11) with genomic summary data, a subset of the gene network can be identified to best explain the data, thereby gaining insights at the gene subnetwork level. Interpreting non-coding SNPs or genomic regions, however, requires additional use of functional genomic datasets, due to the inherent difficulty in linking them to candidate genes. This difficulty can be resolved by leveraging information from promoter capture Hi-C (PCHi-C) datasets that capture physical interactions with gene promoters (12), quantitative trait loci (QTL) datasets that capture genetic regulation with gene expression (eQTL) (13,14) or protein abundance (pQTL) (15), and datasets about enhancer-gene maps that are constructed using the activity-by-contact (ABC) model (16,17).

Extensively extending our XGR software since its first release (1) and incorporating verified approaches and tools (18–25), in this study, we introduce a newly designed web server 'OpenXGR' (Figure 1), which is available at http://www.openxgr.com. Overall, the server is designed to be scalable, efficient and effective, enabling almost *real-time* enrichment and subnetwork analyses for user-input lists of three different entities: genes, SNPs and genomic regions. It is not only limited to the gene- or SNP-centric data types but is also capable of interpreting genomic regions. This generality of capacity for interpreting different entities on the fly is not available elsewhere, thus complementing other popular web servers such as DAVID (26), Enrichr (27) and GREAT (28) that are the most relevant to OpenXGR, and also competitive to standalone tools such as DEPICT (29), MAGMA (30) and jActiveModule (31). OpenXGR achieves this capacity by leveraging increasingly available ontologies, networks, and functional genomic datasets (i.e. PCHi-C, e/pQTL and ABC). Along with a user manual with step-by-step instructions, it offers a user-friendly and all-in-one way to interpret genomic summary data for more integrated and effective knowledge discovery.

In the remaining sections, we will provide an overview of the OpenXGR web server implementation, its six analysers, and the underlying knowledgebase. We will then delve into each analyser that may be of interest to users, with utilities illustrated using practical examples from real-world scenarios, including ageing-related genes (32), gene-level summary data for early human organogenesis (33), SNP-level summary data for chronic inflammatory diseases (34), and genomic region-level summary data for innate immune activation and tolerance (4). Finally, we will conclude with the discussion on the limitations of OpenXGR and the directions for future developments.

MATERIALS AND METHODS

Implementation of the OpenXGR web server

The OpenXGR web server (Figure 1) was newly implemented using the Perl real-time web framework 'Mojolicious' (https://mojolicious.org) and the widely-used 'Bootstrap' (https://getbootstrap.com) to create a mobile-first and responsive design that ensures fast and responsive performance across all major web browsers and mobile devices. All backend computations can be completed within three minutes on the server side to ensure timely delivery of outputs to users. All outputs displayed on the results page are generated using the R package 'bookdown' (https://bookdown.org), providing users with a self-contained dynamic HTML file for download and exploration. Additionally, a user manual with step-by-step instructions is made available where needed to facilitate ease of use and provide guidance for users.

Two types of analysers supported by OpenXGR

The OpenXGR web server offers a range of analysers for conducting enrichment and subnetwork analyses leveraging ontologies and networks. Presently, two types of analysers are supported: one for enrichment analysis designed to identify ontology enrichments, and the other for subnetwork analysis designed to identify gene subnetworks.

Enrichment analysis comprises three analysers that identify enriched ontology terms. These analysers take as input a list of genes, SNPs or genomic regions. One-sided Fisher's exact test is used to calculate Z-scores, odds ratio with its 95% confidence interval (CI), and false discovery rate (FDR) for measuring the significance of enrichments. The following are the three analysers supported by OpenXGR:

(i) *Enrichment analyser for genes (EAG)*, which uses gene-centric ontology annotations to perform enrichment analysis.
(ii) *Enrichment analyser for SNPs (EAS)*, which identifies genes linked from input SNPs (alongside the significance information) and conducts ontology enrichment analysis for the linked genes. Linking SNPs to genes is enabled by genomic proximity or using functional genomic datasets about PCHi-C and e/pQTL.
(iii) *Enrichment analyser for genomic regions (EAR)*, which is similar to *EAS* that first identifies genes linked from input genomic regions using functional genomic datasets about PCHi-C and enhancer-gene maps and then conducts ontology enrichment analysis based on the linked genes.

Subnetwork analysis is performed using three analysers that identify gene subnetworks from input gene-, SNP- or genomic region-level summary data. All subnetwork analysers require the input of the information about the significance level (e.g. P-values). The subnetwork identification is done via a heuristic solver for the prize-collecting Steiner tree problem, demonstrated to be competitive to other state-of-the-art algorithms (1,24). The significance (P-value) of the identified gene subnetwork can be estimated using a degree-preserving node permutation test to count how often it would be expected by chance. The following are the three subnetwork analysers supported by OpenXGR:

(i) *Subnetwork analyser for genes (SAG)*, which takes as input gene-level summary data to identify a subset of the gene network in a manner that the resulting subnetwork contains a desired number of highly scored and interconnected genes.

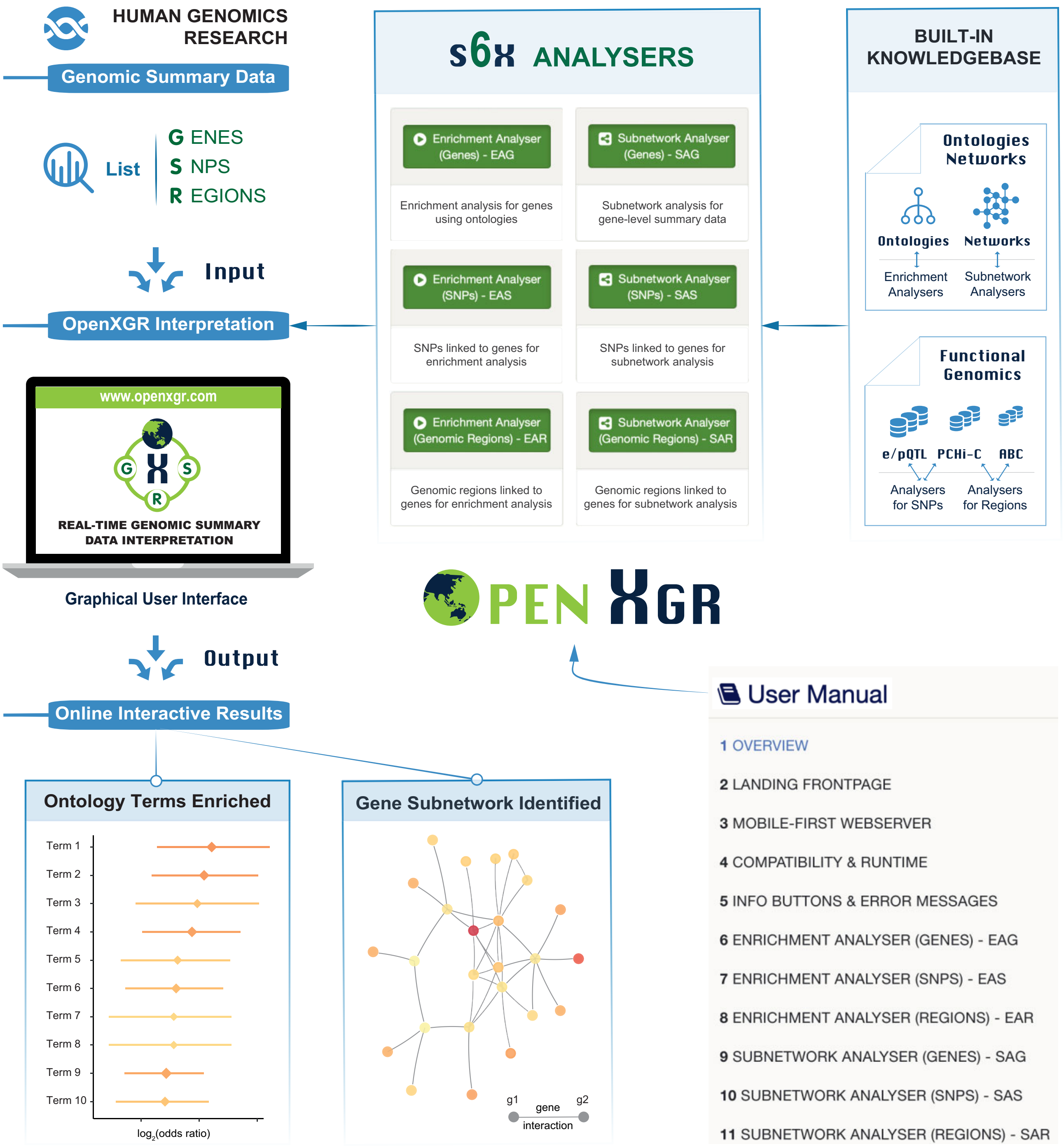

HUMAN GENOMICS RESEARCH
Genomic Summary Data
List
G ENES
S NPS
R EGIONS
Input
OpenXGR Interpretation
www.openxgr.com
G
X
S
R
REAL-TIME GENOMIC SUMMARY DATA INTERPRETATION
Graphical User Interface
Output
Online Interactive Results
s6x ANALYSERS
Enrichment Analyser (Genes) - EAG
Enrichment analysis for genes using ontologies
Subnetwork Analyser (Genes) - SAG
Subnetwork analysis for gene-level summary data
Enrichment Analyser (SNPs) - EAS
SNPs linked to genes for enrichment analysis
Subnetwork Analyser (SNPs) - SAS
SNPs linked to genes for subnetwork analysis
Enrichment Analyser (Genomic Regions) - EAR
Genomic regions linked to genes for enrichment analysis
Subnetwork Analyser (Genomic Regions) - SAR
Genomic regions linked to genes for subnetwork analysis
BUILT-IN KNOWLEDGEBASE
Ontologies Networks
Ontologies
Networks
Enrichment Analysers
Subnetwork Analysers
Functional Genomics
e/pQTL
PCHi-C
ABC
Analysers for SNPs
Analysers for Regions
OPEN XGR
User Manual
1 OVERVIEW
2 LANDING FRONTPAGE
3 MOBILE-FIRST WEBSERVER
4 COMPATIBILITY & RUNTIME
5 INFO BUTTONS & ERROR MESSAGES
6 ENRICHMENT ANALYSER (GENES) - EAG
7 ENRICHMENT ANALYSER (SNPS) - EAS
8 ENRICHMENT ANALYSER (REGIONS) - EAR
9 SUBNETWORK ANALYSER (GENES) - SAG
10 SUBNETWORK ANALYSER (SNPS) - SAS
11 SUBNETWORK ANALYSER (REGIONS) - SAR
Ontology Terms Enriched
Term 1
Term 2
Term 3
Term 4
Term 5
Term 6
Term 7
Term 8
Term 9
Term 10
$\log_2$(odds ratio)
Gene Subnetwork Identified
g1
g2
gene
interaction

Figure 1. Schematic illustration of how the OpenXGR web server works and what to expect from it. The server at http://www.openxgr.com offers six analysers that are designed to interpret various genomic summary data related to genes (G), SNPs (S) and genomic regions (R). By leveraging built-in knowledgebase on ontologies, networks and functional genomics, these analysers allow almost *real-time* enrichment and subnetwork analyses, enabling identification of ontology enrichments and gene subnetworks. A user manual is made available to provide step-by-step instructions on the use.

(ii) *Subnetwork analyser for SNPs (SAS)*, which identifies a gene subnetwork from input SNP-level summary data. It first uses genomic proximity, e/pQTL or PCHi-C to link SNPs to genes, and then uses information on the linked genes to identify the gene subnetwork.

(iii) *Subnetwork analyser for genomic regions (SAR)*, which is similar to *SAS* that first identifies genes linked from input genomic regions using PCHi-C datasets or enhancer-gene maps, followed by subnetwork analysis based on the linked genes.

Leveraging knowledgebase on ontologies, networks and functional genomic datasets

Enrichment analysis in OpenXGR is supported by a variety of ontologies that span a wide range of knowledge contexts, ranging from functions and pathways to regulators, from diseases and phenotypes to drugs, and from protein domains and disorders to hallmarks and evolution. Ontologies currently supported are: (a) *functions*: Gene Ontology (GO) (6) (accessed in April 2023), subdivided into GO Biological Process (GOBP), GO Molecular Function (GOMF), and GO Cellular Component (GOCC); (b) *pathways*: KEGG (35) (105.0 release), REACTOME (36) (version 84 release), pathways from MSIGDB (37) (v2023.1.Hs release), and MitoPathways from MitoCarta (38) (3.0 version); (c) *regulators*: ENRICHR Consensus TFs (27) (accessed in April 2023) and TRRUST (39) (2018.04.16 release); (d) *diseases*: Mondo Disease Ontology (MONDO) (9) (v2023-01-04 release), Disease Ontology (10) (March 2023 release), and Experimental Factor Ontology (EFO) for disease traits (40) (3.52.0 release); (e) *phenotypes*: Human Phenotype Ontology (HPO) (8) (June 2022 release) and Mammalian Phenotype Ontology (MPO) (7) (accessed in April 2023); (f) *drugs*: DGIdb druggable categories (41) (2022-Feb release), target tractability buckets (Bucket) (42) (23.02 release), and ChEMBL drug indications (43) (version 32); (g) *domains & disorders*: SCOP (44), Pfam (45), InterPro (46), and Intrinsically Disordered Proteins Ontology (IDPO) (47) and (h) *hallmarks & evolution*: molecular signature hallmarks from MSigDB (48) and Phylostratigraphy (49).

Subnetwork analysis in OpenXGR leverages the knowledge of functional or pathway interaction networks. Functional interaction networks are sourced from the STRING database (11) (version 11.5), with only the 'experiments' and 'databases' source codes considered. Functional interactions are classified as having the highest confidence (≥0.9), high confidence (≥0.7), and medium confidence (≥0.4). Pathway interaction networks are sourced from the KEGG database (35) (105.0 release), with all individual pathways being merged into a gene network.

In OpenXGR, linking SNPs to genes is enabled through genomic proximity, PCHi-C or e/pQTL, while linking genomic regions to genes is achieved based on PCHi-C or enhancer-gene maps. Functional genomic datasets currently supported include PCHi-C in immune-, blood- and brain-related cell types (50–53), plasma pQTL (15), blood eQTL from the eQTLGene Consortium (14), eQTL in immune-related cell types and brain-related tissues (13,54,55), and enhancer-gene maps in ENCODE or Roadmap cell types constructed using the ABC model (16,17).

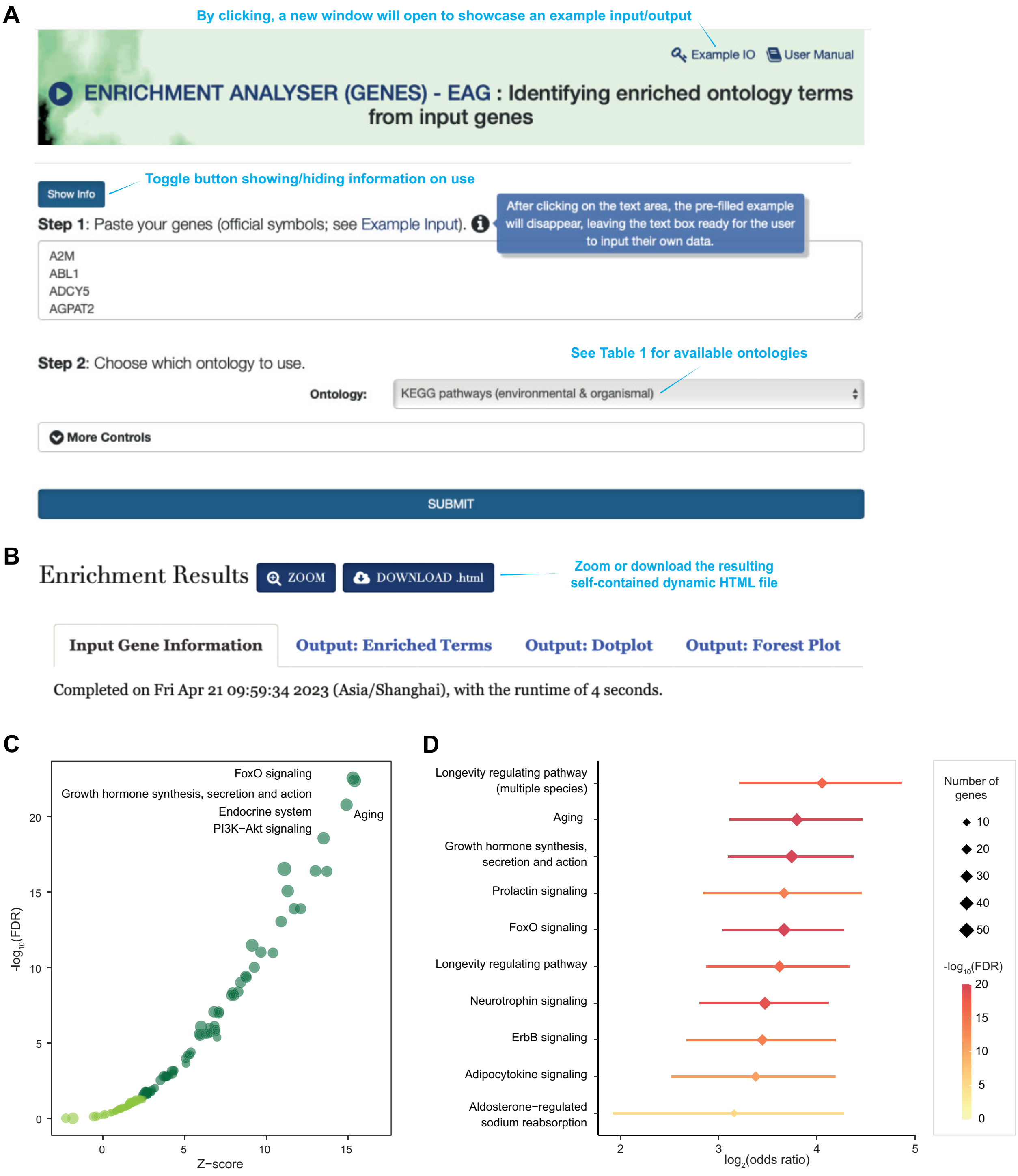
RESULTS

Capabilities of enrichment analysers in identifying enriched ontology terms from input genes, SNPs or genomic regions

Enrichment Analyser (Genes)– EAG conducts enrichment analysis for genes leveragingontologies. *EAG* is designed to leverage gene-centric ontology annotations to identify enriched ontology terms from input genes. The tool comprises two major steps, which are outlined in the user-request interface (Figure 2A). The interface takes a list of genes as input, such as ∼300 ageing-related genes as an illustrative example (32). Available ontologies are organised by category (Table 1). Additional parameters can be specified to control the enrichment analysis and results. The interface features a toggle button to show/hide information on the use, including details on input, output and other useful information, as well as a key icon that provides an example input/output showcase. In the results page (Figure 2B), the '*Input Gene Information*' tab lists the input genes and hyperlinks to their corresponding GeneCards pages for additional information and displays the server-side runtime. The '*Output: Enriched Terms*' tab features an interactive table that displays enriched ontology terms, along with their significance information such as Z-scores, FDR, odds ratio and its 95% CI. It also shows member genes that overlap with the input genes. The results are also illustrated in the '*Output: Dotplot*' tab, displaying the top five terms with their respective Z-scores and FDR (Figure 2C), and in the '*Output: Forest Plot*' tab, listing the top enriched terms ordered by odds ratio (Figure 2D). As expected, the most enrichments are ageing-related, such as FoxO signaling, longevity regulating pathways, and ageing. It is worth noting that all enrichment results are embedded into a self-contained dynamic HTML file that can be downloaded and explored interactively in a new browser window. We highly recommend users download this file for subsequent exploration.

EnrichmentAnalyser (SNPs) – EAS links SNPs to candidate genes for enrichment analysis. *EAS* achieves this by linking input SNPs to candidate genes in three steps. The user-request interface requires two pieces of information as input: SNPs and their significance info (p-values). For example, the interface presents an illustrative example of ∼210 SNPs and their reported p-values for chronic inflammatory diseases (34). By default, this analyser considers input SNPs with a *P*-value threshold of $<5 \times 10^{-8}$, and additional SNPs in linkage disequilibrium (R^2 ≥0.8) according to the European population, though other populations are also supported (56). Input and additional SNPs are then linked to genes based on genomic proximity, PCHi-C or e/pQTL (see Table 1). The linked genes are scored based on p-values, threshold and R^2 for SNPs, the distance window for genomic proximity, the strength of gene promoters physically interacting with SNP-harbouring genomic regions for PCHi-C datasets, and the significance level defining e/pQTL, as previously described (1,23). Enriched ontology terms are identified based on enrichment analysis of the linked genes. In addition to a dot plot and a forest plot, the output also includes two tabular displays under the '*Output: Linked Genes*' tab. One lists the linked genes and their scores, which range from 1 to 10. The other is an evidence table showing which SNPs are used to define the linked genes based on which datasets.

EnrichmentAnalyser (Genomic Regions) – EAR links genomic regions to candidate genes for enrichmentanalysis. *EAR* works similarly to *EAS*, but instead of input SNPs, it links input genomic regions to candidate genes and performs enrichment analysis on them. Users specify the genomic coordinates of the input regions, including the chromosome, start, and end positions. The genome build for the input regions is also required, with hg19 used internally as a default and automatically converted if a different build is provided. For example, an input of ∼380 differentially expressed enhancer RNAs (non-coding regions) involved in innate immune activation and tolerance is used as an illustration (4). The linked genes are identified and scored



A

By clicking, a new window will open to showcase an example input/output

Example IO · User Manual

ENRICHMENT ANALYSER (GENES) - EAG : Identifying enriched ontology terms from input genes

Show Info

Toggle button showing/hiding information on use

Step 1: Paste your genes (official symbols; see Example Input).

After clicking on the text area, the pre-filled example will disappear, leaving the text box ready for the user to input their own data.

A2M
ABL1
ADCY5
AGPAT2

See Table 1 for available ontologies

Step 2: Choose which ontology to use.

Ontology: KEGG pathways (environmental & organismal)

More Controls

SUBMIT

B

Enrichment Results · ZOOM · DOWNLOAD .html

Zoom or download the resulting self-contained dynamic HTML file

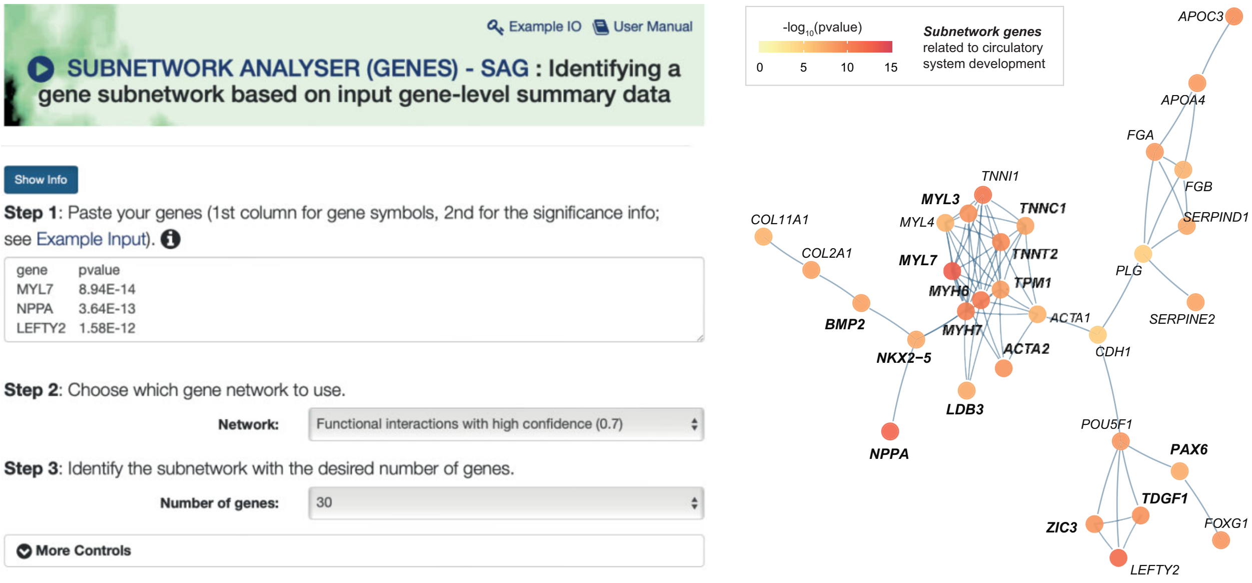
Input Gene Information · Output: Enriched Terms · Output: Dotplot · Output: Forest Plot

Completed on Fri Apr 21 09:59:34 2023 (Asia/Shanghai), with the runtime of 4 seconds.

Figure 2. Enrichment analysis for genes using *EAG*. (**A**) User-request interface. The interface takes as input a list of genes and provides available ontologies (listed in Table 1). **(B)** Enrichment results page. It displays a summary of input data under the '*Input Gene Information*' tab, a table of enriched ontology terms under the '*Output: Enriched Terms*' tab, a dot plot of enrichments under the 'Output: Dotplot' tab (see **C**), and a forest plot of enrichments under the '*Output: Forest Plot*' tab (see **D**).

Table 1. A summary of web browser compatibility, analysers, and built-in knowledgebase (ontologies, networks and genomic datasets) in OpenXGR

Objects	**Characteristics to compare**		
Web browser compatibility			
	MacOS (Big Sur)	**Windows(10)**	**Linux (Ubuntu)**
Safari	15.6.1	N/A	N/A
Edge	N/A	108.0.1462.54	N/A
Chrome	108.0.5359.124	108.0.5359.124	108.0.5359.124
Firefox	108.0.1	108.0.1	108.0.1
Analysers			
	Type	**Description**	**Entities**
EAG	Enrichment	Enrichment Analyser (Genes)	Genes
EAS	Enrichment	Enrichment Analyser (SNPs)	SNPs
EAR	Enrichment	Enrichment Analyser (Regions)	Genomic regions
SAG	Subnetwork	Subnetwork Analyser (Genes)	Genes
SAS	Subnetwork	Subnetwork Analyser (SNPs)	SNPs
SAR	Subnetwork	Subnetwork Analyser (Regions)	Genomic regions
Built-in ontologies			
	Category	**Description**	**Used by analysers**
GOBP	Functions	Gene Ontology Biological Process	EAG, EAS, EAR
GOMF	Functions	Gene Ontology Molecular Function	EAG, EAS, EAR
GOCC	Functions	Gene Ontology Cellular Component	EAG, EAS, EAR
KEGG	Pathways	KEGG pathways	EAG, EAS, EAR
REACTOME	Pathways	REACTOME pathways	EAG, EAS, EAR
MSIGDBpath	Pathways	MSIGDB pathways	EAG, EAS, EAR
MITOPATH	Pathways	MitoPathway pathways	EAG, EAS, EAR
CTF	Regulators	ENRICHR Consensus TFs	EAG, EAS, EAR
TRRUST	Regulators	ENRICHR TRRUST TFs	EAG, EAS, EAR
MONDO	Diseases	Mondo Disease Ontology	EAG, EAS, EAR
DO	Diseases	Disease Ontology	EAG, EAS, EAR
EFO	Diseases	Experimental Factor Ontology	EAG, EAS, EAR
HPO	Phenotypes	Human Phenotype Ontology	EAG, EAS, EAR
MPO	Phenotypes	Mammalian Phenotype Ontology	EAG, EAS, EAR
DGIdb	Drugs	DGIdb druggable categories	EAG, EAS, EAR
Bucket	Drugs	Target tractability buckets	EAG, EAS, EAR
ChEMBL	Drugs	ChEMBL drug indications	EAG, EAS, EAR
SCOPsf	Domains & Disorders	SCOP superfamily domains	EAG, EAS, EAR
SCOPfa	Domains & Disorders	SCOP family domains	EAG, EAS, EAR
Pfam	Domains & Disorders	Pfam domains	EAG, EAS, EAR
InterPro	Domains & Disorders	InterPro domains	EAG, EAS, EAR
IDPO	Domains & Disorders	Intrinsically Disordered Proteins Ontology	EAG, EAS, EAR
MSIGDBh	Hallmarks & Evolution	MSIGDB hallmarks	EAG, EAS, EAR
PSG	Hallmarks & Evolution	Phylostratigraphy	EAG, EAS, EAR
Built-in gene networks			
	Category	**Description**	**Used by analysers**
STRING	Functions	Functional interaction networks	SAG, SAS, SAR
KEGG	Pathways	Pathway interaction networks	SAG, SAS, SAR
Built-in functional genomic datasets			
	Category	**Description**	**Used by analysers**
PMID27863249	PCHi-C	Blood-related cell types or states ($n = 16$)	EAS, EAR, SAS, SAR
PMID31501517	PCHi-C	Brain-related tissues ($n = 3$)	EAS, EAR, SAS, SAR
PMID31367015	PCHi-C	Brain-related tissues ($n = 4$)	EAS, EAR, SAS, SAR
PMID25938943	PCHi-C	Blood-related cell types ($n = 2$)	EAS, EAR, SAS, SAR
PMID29875488	pQTL	Plasma eQTL	EAS, SAS
PMID34475573	eQTL	Blood eQTLGen	EAS, SAS
PMID30449622	eQTL	Immune-related cell types ($n = 15$)	EAS, SAS
PMID32913098	eQTL	Brain-related tissues ($n = 13$)	EAS, SAS
PMID33828297	ABC	ENCODE cell types (all combined to be cell type-agnostic)	EAR, SAR
PMID33828297	ABC	Roadmap cell types (all combined to be cell type-agnostic)	EAR, SAR



Example IO User Manual

SUBNETWORK ANALYSER (GENES) - SAG : Identifying a gene subnetwork based on input gene-level summary data

Show Info

Step 1: Paste your genes (1st column for gene symbols, 2nd for the significance info; see Example Input).

```
gene    pvalue
MYL7    8.94E-14
NPPA    3.64E-13
LEFTY2  1.58E-12
```

Step 2: Choose which gene network to use.

Network: Functional interactions with high confidence (0.7)

Step 3: Identify the subnetwork with the desired number of genes.

Number of genes: 30

More Controls

Figure 3. Subnetwork analysis for gene-level summary data using *SAG*. The left panel shows the user-request interface where input gene-level summary data can be provided. The right panel visualises the identified gene subnetwork, with genes/nodes color-coded by their input gene significance information.

based on genomic proximity, PCHi-C, or enhancer-gene maps (see Table 1). The output includes tabular displays of the linked genes and graphical plots of enriched ontology terms. The linked gene table under the '*Output: Linked Genes*' tab displays information on genes linked from input genomic regions, including scores that quantify the degree to which genes are likely modulated by input genomic regions. The evidence table displays information on which regions are linked to genes based on which evidence. Taken together, *EAR* can handle various genomic regions, such as differentially expressed regions, differentially methylated DNA regions, transcription factor binding sites, and epigenetic marks from epigenomic experiments. It assists in the interpretation by identifying ontology enrichments and candidate genes associated with input genomic regions.

Capabilities of subnetwork analysers in identifying gene subnetworks from input gene-, SNP- or genomic region-level summary data

SubnetworkAnalyser (Genes) – SAG performs subnetwork analysis for gene-level summary data leveraging networks. *SAG* is an analyser designed to exploit knowledge of protein interactions or pathway-derived gene interactions to identify gene subnetworks from input gene-level summary data (Figure 3). A typical example would be a list of differentially expressed genes with their corresponding significance information. An illustrative example provided in the user-input interface is the list of stage-transitive differential genes between Carnegie stages 9 and 10 during early human organogenesis (33). Functional interaction networks are sourced from the STRING database (11), and by default, the high-confidence interactions are used, corresponding to ∼14 800 genes and ∼203 900 interactions. Pathway interaction networks are derived by merging pathways from the KEGG database (35), collectively forming a gene network with ∼6000 genes and ∼59 000 interactions. *SAG* aims to identify a subset of the gene network in such a way that the resulting gene subnetwork (or 'pathway crosstalk' if pathway interactions are used) contains most, if not all, of the most significantly and differentially expressed genes in this illustrative example (Figure 3). Interestingly, most of the subnetwork genes are involved in circulatory system development (*ACTA2*, *BMP2*, *LDB3*, *MYH6*, *MYH7*, *MYL3*, *MYL7*, *NKX2-5*, *NPPA*, *PAX6*, *TDGF1*, *TNNC1*, *TNNT2*, *TPM1* and *ZIC3*), amongst which eight genes (*BMP2*, *MYH6*, *MYH7*, *MYL3*, *NKX2-5*, *TNNC1*, *TNNT2* and *TPM1*) are related to cardiac muscle tissue morphogenesis. In other words, the identified subnetwork likely explains primordial development of heart taking place at Carnegie stage 10 (33). Users can specify the desired number of nodes/genes in the resulting subnetwork, and the output is returned via a well-established iterative search procedure (1,24). In summary, *SAG* takes a list of genes along with their significance information, such as differential genes showcased here, and returns a tabular display of the subnetwork genes and a network-like visualisation of the subnetwork (with nodes/genes colour-coded by input gene significance information).

SubnetworkAnalyser (SNPs) – SAS links SNPs to candidate genes for subnetwork analysis. *SAS* is designed to perform subnetwork analysis using input SNP-level summary data, with the goal of linking SNPs to genes for subsequent analysis. The first three steps in user-request interface are identical to those in *EAS*. Instead of specifying which ontology to use, users must indicate which gene network to use and provide specifications to control the desired number of the resulting subnetwork genes. Using the same example as in the previous section for *EAS*, under the '*Output: Gene Subnetwork*' tab in the subnetwork results page, the output subnetwork is visualised, with genes/nodes colour-coded by linked gene scores. Interestingly, most of subnetwork genes are involved in C-type lectin receptor signaling (*CARD9*, *CYLD*, *IL10*, *IL12B*, *IL2*, *IRF1*, *NFKB1* and

RHOA), JAK-STAT signaling (*IL10*, *IL12B*, *IL19*, *IL2*, *IL23R*, *JAK2*, *PDGFB*, *PTPN2* and *SOCS1*), and TNF signaling (*CCL2*, *FOS*, *IRF1*, *NFKB1*, *NOD2* and *TNFRSF1A*). These findings are consistent with the importance of these pathways in inflammation and inflammatory diseases (57–59). In summary, *SAS* is a valuable online tool that links SNPs to genes, enabling the identification of subnetworks that are critical to the understanding of the genetic basis of complex diseases. The resulting gene subnetwork is returned in a tabular display and a network-like visualisation, which facilitates further analysis of candidate genes, particularly enrichment analysis of subnetwork genes to identify enriched pathways.

SubnetworkAnalyser (Genomic Regions) – SAR links genomic regions to candidate genes for subnetwork analysis. Similar to *SAS*, this analyser is specially designed for subnetwork analysis using input summary data but at the genomic region level. The first three steps in user-request interface are identical to those in *EAR*. Users need to indicate gene networks to use and specify the desired number of the resulting subnetwork genes. Using real-world summary data on non-coding enhancer RNAs differentially expressed upon innate immune activation and tolerance (4), the output subnetwork is retuned under the tab '*Output: Gene Subnetwork*' in the results page. Enrichment analysis of the resulting subnetwork genes via *EAG* identifies JAK-STAT signaling (*CDKN1A*, *CISH*, *IL10*, *IL10RA*, *IL19*, *IL20*, *IL7*, *IL7R* and *JAK2*) as the most significant pathway (FDR = 2.0×10^{-6}; odds ratio = 16.0; 95% CI = [6.07, 39.7]), highlighting its crucial role in mediating innate immune activation and tolerance (57).

DISCUSSION

We have developed OpenXGR to meet the growing demand for efficient and effective interpretation of the ever-increasing volume of summary-level data in genomics. Designed as a versatile and user-friendly web server, it can interpret a wide range of genomic summary data related to three different entities (namely, genes, SNPs and genomic regions). This represents a significant development in human genomics research, as it has the potential to facilitate a more comprehensive understanding of genomic summary data and more effective downstream knowledge discovery.

One of the unique features of OpenXGR is its ability to identify gene subnetworks from input summary data at the gene, SNP and genomic region levels, providing valuable insights into the functional relationships between genes (or linked genes) and aiding in researchers to identify potential pathways or networks that best explain specific biological processes or diseases. Another significant advancement offered by OpenXGR is its use of functional genomic datasets, such as PCHi-C, e/pQTL and enhancer-gene maps, to link non-coding SNPs or genomic regions to candidate genes. This enables interpretation of input SNPs and genomic regions, regardless of their location in the genome, which is often difficult or lacking in existing tools (for interpreting non-coding entities).

However, we recognise that there are limitations to OpenXGR regarding the availability of functional genomic datasets, which currently primarily support blood- and brain-related contexts. Thus, our first aim for future developments is to expand the supporting functional genomic datasets to include a diverse range of cell types, states and tissues. Additionally, enrichment and subnetwork analyses are currently limited to the human genome, so our second aim is to support model organisms, for example, the extension to the mouse genome already on the agenda. This will expand the capacity of OpenXGR in interpreting genomic summary data beyond human. Looking further ahead, we are excited about the opportunity of employing large language models (60) to support genomic summary data interpretation, either in generating ontology annotations and gene networks or in providing outputs in a conversational way similar to ChatGPT. Other future efforts will focus on improving the selection panel of available options (e.g. cell type-specific information on PCHi-C, eQTLs and enhancer-gene maps), supporting enrichment and subnetwork analyses for protein structural domains taken from the dcGO resource (25,61), increasing the user base, and committing to the OpenXGR web server update twice a year. In the long run, OpenXGR will function as an interactive, user-friendly and all-in-one platform that accelerates genomic summary data interpretation by leveraging ontologies, networks, and functional genomics as well.

DATA AVAILABILITY

The OpenXGR web server is easily accessible at http://www.openxgr.com, where the user manual is also available that provides step-by-step instructions on how to get started (http://www.openxgr.com/OpenXGRbooklet/index.html). The source code is made available on GitHub at https://github.com/hfang-bristol/OpenXGR-site and Figshare at https://doi.org/10.6084/m9.figshare.22679284.v1. For added convenience, OpenXGR can also be accessed through the mirror site at http://www.genomicsummary.com/OpenXGR, along with the user manual at http://www.genomicsummary.com/OpenXGRbooklet/index.html.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to two referees and numerous anonymous users who generously tested OpenXGR during its development and deployment. Their insightful feedbacks greatly contributed to the improvement of the web server.

FUNDING

National Natural Science Foundation of China [32170663 to H.F.]; Shanghai Pujiang Program [21PJ1409600 to H.F.]; Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning [H.F.]; Innovative Research Team of High-Level Local Universities in Shanghai. Funding for open access charge: National Natural Science Foundation of China.

Conflict of interest statement. None declared

REFERENCES

1. Fang,H., Knezevic,B., Burnham,K.L. and Knight,J.C. (2016) XGR software for enhanced interpretation of genomic summary data, illustrated by application to immunological traits. *Genome Med.*, **8**, 129.
2. Stark,R., Grzelak,M. and Hadfield,J. (2019) RNA sequencing: the teenage years. *Nat. Rev. Genet.*, **20**, 631–656.
3. Tam,V., Patel,N., Turcotte,M., Bossé,Y., Paré,G. and Meyre,D. (2019) Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.*, **20**, 467–484.
4. Zhang,P., Amarasinghe,H., Whalley,J.P., Tay,C., Fang,H., Migliorini,G., Brown,A., Allcock,A., Scozzafava,G., Rath,P. *et al.* (2022) Epigenomic analysis reveals a dynamic and context-specific macrophage enhancer landscape associated with innate immune activation and tolerance. *Genome Biol.*, **23**, 136.
5. Kundaje,A., Meuleman,W., Ernst,J., Bilenky,M., Yen,A., Heravi-Moussavi,A., Kheradpour,P., Zhang,Z., Wang,J., Ziller,M.J. *et al.* (2015) Integrative analysis of 111 reference human epigenomes. *Nature*, **518**, 317–330.
6. Carbon,S., Douglass,E., Good,B.M., Unni,D.R., Harris,N.L., Mungall,C.J., Basu,S., Chisholm,R.L., Dodson,R.J., Hartline,E. *et al.* (2021) The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res.*, **49**, D325–D334.
7. Bogue,M.A., Philip,V.M., Walton,D.O., Grubb,S.C., Dunn,M.H., Kolishovski,G., Emerson,J., Mukherjee,G., Stearns,T., He,H. *et al.* (2020) Mouse Phenome Database: a data repository and analysis suite for curated primary mouse phenotype data. *Nucleic Acids Res.*, **48**, D716–D723.
8. Köhler,S., Gargano,M., Matentzoglu,N., Carmody,L.C., Lewis-Smith,D., Vasilevsky,N.A., Danis,D., Balagura,G., Baynam,G., Brower,A.M. *et al.* (2021) The human phenotype ontology in 2021. *Nucleic Acids Res.*, **49**, D1207–D1217.
9. Shefchek,K.A., Harris,N.L., Gargano,M., Matentzoglu,N., Unni,D., Brush,M., Keith,D., Conlin,T., Vasilevsky,N., Zhang,X.A. *et al.* (2020) The Monarch Initiative in 2019: an integrative data and analytic platform connecting phenotypes to genotypes across species. *Nucleic Acids Res.*, **48**, D704–D715.
10. Schriml,L.M., Munro,J.B., Schor,M., Olley,D., McCracken,C., Felix,V., Baron,J.A., Jackson,R., Bello,S.M., Bearer,C. *et al.* (2022) The Human Disease Ontology 2022 update. *Nucleic Acids Res.*, **50**, D1255–D1261.
11. Szklarczyk,D., Gable,A.L., Nastou,K.C., Lyon,D., Kirsch,R., Pyysalo,S., Doncheva,N.T., Legeay,M., Fang,T., Bork,P. *et al.* (2021) The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.*, **49**, D605–D612.
12. Schoenfelder,S. and Fraser,P. (2019) Long-range enhancer–promoter contacts in gene expression control. *Nat. Rev. Genet.*, **20**, 437–455.
13. Kerimov,N., Hayhurst,J.D., Peikova,K., Manning,J.R., Walter,P., Kolberg,L., Samoviča,M., Sakthivel,M.P., Kuzmin,I., Trevanion,S.J. *et al.* (2021) A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nat. Genet.*, **53**, 1290–1299.
14. Võsa,U., Claringbould,A., Westra,H.-J., Bonder,M.J., Deelen,P., Zeng,B., Kirsten,H., Saha,A., Kreuzhuber,R., Yazar,S. *et al.* (2021) Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.*, **53**, 1300–1310.
15. Sun,B.B., Maranville,J.C., Peters,J.E., Stacey,D., Staley,J.R., Blackshaw,J., Burgess,S., Jiang,T., Paige,E., Surendran,P. *et al.* (2018) Genomic atlas of the human plasma proteome. *Nature*, **558**, 73–79.
16. Fulco,C.P., Nasser,J., Jones,T.R., Munson,G., Bergman,D.T., Subramanian,V., Grossman,S.R., Anyoha,R., Patwardhan,T.A., Nguyen,T.H. *et al.* (2019) Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.*, **51**, 1664–1669.
17. Nasser,J., Bergman,D.T., Fulco,C.P., Guckelberger,P., Doughty,B.R., Patwardhan,T.A., Jones,T.R., Nguyen,T.H., Ulirsch,J.C., Lekschas,F. *et al.* (2021) Genome-wide enhancer maps link risk variants to disease genes. *Nature*, **593**, 238–243.
18. Fang,H. and Knight,J.C. (2022) Priority index: database of genetic targets in immune-mediated disease. *Nucleic Acids Res.*, **50**, D1358–D1367.
19. Bao,C., Wang,H. and Fang,H. (2022) Genomic evidence supports the recognition of endometriosis as an inflammatory systemic disease and reveals disease-specific therapeutic potentials of targeting neutrophil degranulation. *Front. Immunol.*, **13**, 758440.
20. Fang,H. (2022) PiER: web-based facilities tailored for genetic target prioritisation harnessing human disease genetics, functional genomics and protein interactions. *Nucleic Acids Res.*, **50**, W583–W592.
21. Fang,H. and Jiang,L. (2021) Genetic prioritization, therapeutic repositioning and cross-disease comparisons reveal inflammatory targets tractable for kidney stone disease. *Front. Immunol.*, **12**, 687291.
22. Fang,H., Chen,L. and Knight,J.C. (2020) From genome-wide association studies to rational drug target prioritisation in inflammatory arthritis. *Lancet Rheumatol.*, **2**, e50–e62.
23. Fang,H. and The ULTRA-DD ConsortiumThe ULTRA-DD Consortium, De Wolf,H., Knezevic,B., Burnham,K.L., Osgood,J., Sanniti,A., Lledó Lara,A., Kasela,S., De Cesco,S. *et al.* (2019) A genetics-led approach defines the drug target landscape of 30 immune-related traits. *Nat. Genet.*, **51**, 1082–1091.
24. Fang,H. and Gough,J. (2014) The 'dnet' approach promotes emerging research on cancer patient survival. *Genome Med.*, **6**, 64.
25. Fang,H. and Gough,J. (2013) dcGO: database of domain-centric ontologies on functions, phenotypes, diseases and more. *Nucleic Acids Res.*, **41**, D536–D544.
26. Sherman,B.T., Hao,M., Qiu,J., Jiao,X., Baseler,M.W., Lane,H.C., Imamichi,T. and Chang,W. (2022) DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.*, **50**, W216–W221.
27. Kuleshov,M.V., Jones,M.R., Rouillard,A.D., Fernandez,N.F., Duan,Q., Wang,Z., Koplev,S., Jenkins,S.L., Jagodnik,K.M., Lachmann,A. *et al.* (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.*, **44**, W90–W97.
28. McLean,C.Y., Bristor,D., Hiller,M., Clarke,S.L., Schaar,B.T., Lowe,C.B., Wenger,A.M. and Bejerano,G. (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.*, **28**, 495–501.
29. Pers,T.H., Karjalainen,J.M., Chan,Y., Westra,H.J., Wood,A.R., Yang,J., Lui,J.C., Vedantam,S., Gustafsson,S., Esko,T. *et al.* (2015) Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.*, **6**, 5890.
30. de Leeuw,C.A., Mooij,J.M., Heskes,T. and Posthuma,D. (2015) MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.*, **11**, e1004219.
31. Ideker,T., Ozier,O., Schwikowski,B. and Andrew,F. (2002) Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics*, **18**, S233–S240.
32. Tacutu,R., Thornton,D., Johnson,E., Budovsky,A., Barardo,D., Craig,T., DIana,E., Lehmann,G., Toren,D., Wang,J. *et al.* (2018) Human Ageing Genomic Resources: new and updated databases. *Nucleic Acids Res.*, **46**, D1083–D1090.
33. Fang,H., Yang,Y., Li,C., Fu,S., Yang,Z., Jin,G., Wang,K., Zhang,J. and Jin,Y. (2010) Transcriptome analysis of early organogenesis in human embryos. *Dev. Cell*, **19**, 174–184.
34. Ellinghaus,D., Jostins,L., Spain,S.L., Cortes,A., Bethune,J., Han,B., Park,Y.R., Raychaudhuri,S., Pouget,J.G., Hubenthal,M. *et al.* (2016) Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat. Genet.*, **48**, 510–518.
35. Kanehisa,M., Furumichi,M., Sato,Y., Kawashima,M. and Ishiguro-Watanabe,M. (2023) KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.*, **51**, D587–D592.
36. Gillespie,M., Jassal,B., Stephan,R., Milacic,M., Rothfels,K., Senff-Ribeiro,A., Griss,J., Sevilla,C., Matthews,L., Gong,C. *et al.* (2022) The reactome pathway knowledgebase 2022. *Nucleic Acids Res.*, **50**, D687–D692.
37. 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.
38. Rath,S., Sharma,R., Gupta,R., Ast,T., Chan,C., Durham,T.J., Goodman,R.P., Grabarek,Z., Haas,M.E., Hung,W.H.W. *et al.* (2021) MitoCarta3.0: an updated mitochondrial proteome now with

sub-organelle localization and pathway annotations. *Nucleic Acids Res.*, **49**, D1541–D1547.
39. Han,H., Cho,J.-W., Lee,S., Yun,A., Kim,H., Bae,D., Yang,S., Kim,C.Y., Lee,M., Kim,E. *et al.* (2018) TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res.*, **46**, D380–D386.
40. Sollis,E., Mosaku,A., Abid,A., Buniello,A., Cerezo,M., Gil,L., Groza,T., Güneş,O., Hall,P., Hayhurst,J. *et al.* (2023) The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res.*, **51**, D977–D985.
41. Freshour,S.L., Kiwala,S., Cotto,K.C., Coffman,A.C., McMichael,J.F., Song,J.J., Griffith,M., Griffith,O.L. and Wagner,A.H. (2021) Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts. *Nucleic Acids Res.*, **49**, D1144–D1151.
42. Ochoa,D., Hercules,A., Carmona,M., Suveges,D., Baker,J., Malangone,C., Lopez,I., Miranda,A., Cruz-Castillo,C., Fumis,L. *et al.* (2023) The next-generation Open Targets Platform: reimagined, redesigned, rebuilt. *Nucleic Acids Res.*, **51**, D1353–D1359.
43. Mendez,D., Gaulton,A., Bento,A.P., Chambers,J., De Veij,M., Félix,E., Magariños,M.P., Mosquera,J.F., Mutowo,P., Nowotka,M. *et al.* (2019) ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.*, **47**, D930–D940.
44. Murzin,A.G., Brenner,S.E., Hubbard,T. and Chothia,C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, **247**, 536–540.
45. Mistry,J., Chuguransky,S., Williams,L., Qureshi,M., Salazar,G.A., Sonnhammer,E.L.L., Tosatto,S.C.E., Paladin,L., Raj,S., Richardson,L.J. *et al.* (2021) Pfam: the protein families database in 2021. *Nucleic Acids Res.*, **49**, D412–D419.
46. Blum,M., Chang,H.Y., Chuguransky,S., Grego,T., Kandasaamy,S., Mitchell,A., Nuka,G., Paysan-Lafosse,T., Qureshi,M., Raj,S. *et al.* (2021) The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.*, **49**, D344–D354.
47. Salladini,E., Lazar,T., Pancsa,R., Chemes,B., Pe,S., Santos,J., Acs,V., Farahi,N., Dobson,L., Chasapi,A. *et al.* (2022) DisProt in 2022: improved quality and accessibility of protein intrinsic disorder annotation. *Nucleic Acids Res.*, **50**, D480–D487.
48. Liberzon,A., Birger,C., Thorvaldsdóttir,H., Ghandi,M., Mesirov,J.P. and Tamayo,P. (2015) The molecular signatures database hallmark gene set collection. *Cell Syst.*, **1**, 417–425.
49. Trigos,A.S., Pearson,R.B., Papenfuss,A.T. and Goode,D.L. (2017) Altered interactions between unicellular and multicellular genes drive hallmarks of transformation in a diverse range of solid tumors. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 6406–6411.
50. Mifsud,B., Tavares-Cadete,F., Young,A.N., Sugar,R., Schoenfelder,S., Ferreira,L., Wingett,S.W., Andrews,S., Grey,W., Ewels,P.A. *et al.* (2015) Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.*, **47**, 598–606.
51. Javierre,B.M., Burren,O.S., Wilder,S.P., Kreuzhuber,R., Hill,S.M., Sewitz,S., Cairns,J., Wingett,S.W., Várnai,C., Thiecke,M.J. *et al.* (2016) Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell*, **167**, 1369–1384.
52. Jung,I., Schmitt,A., Diao,Y., Lee,A.J., Liu,T., Yang,D., Tan,C., Eom,J., Chan,M., Chee,S. *et al.* (2019) A compendium of promoter-centered long-range chromatin interactions in the human genome. *Nat. Genet.*, **51**, 1442–1449.
53. Song,M., Yang,X., Ren,X., Maliskova,L., Li,B., Jones,I.R., Wang,C., Jacob,F., Wu,K., Traglia,M. *et al.* (2019) Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. *Nat. Genet.*, **51**, 1252–1262.
54. Schmiedel,B.J., Singh,D., Madrigal,A., Valdovino-Gonzalez,A.G., White,B.M., Zapardiel-Gonzalo,J., Ha,B., Altay,G., Greenbaum,J.A., McVicker,G. *et al.* (2018) Impact of genetic polymorphisms on human immune cell gene expression. *Cell*, **175**, 1701–1715.
55. The GTEx Consortium (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*, **369**, 1318–1330.
56. 1000 Genomes Project Consortium (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491**, 56–65.
57. Banerjee,S., Biehl,A., Gadina,M., Hasni,S. and Schwartz,D.M. (2017) JAK–STAT signaling as a target for inflammatory and autoimmune diseases: current and future prospects. *Drugs*, **77**, 521–546.
58. del Fresno,C., Iborra,S., Saz-Leal,P., Martínez-López,M. and Sancho,D. (2018) Flexible signaling of Myeloid C-type lectin receptors in immunity and inflammation. *Front. Immunol.*, **9**, 804.
59. van Loo,G. and Bertrand,M.J.M. (2022) Death by TNF: a road to inflammation. *Nat. Rev. Immunol.*, **23**, 289–303.
60. Brown,T.B., Mann,B., Ryder,N., Subbiah,M., Kaplan,J., Dhariwal,P., Neelakantan,A., Shyam,P., Sastry,G., Askell,A. *et al.* (2020) Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.*, **33**, 1877–1901.
61. Bao,C., Lu,C., Lin,J., Gough,J. and Fang,H. (2023) The dcGO domain-centric ontology database in 2023: new website and extended annotations for protein structural domains. *J. Mol. Biol.*, **435**, 168093.