

1-1-2019

## Curating gene sets: challenges and opportunities for integrative analysis.

Jason A. Bubier

David P. Hill


Gaurab Mukherjee

Timothy Reynolds

Erich J Baker

*See next page for additional authors*

Follow this and additional works at: <https://mouseion.jax.org/stfb2019>

 Part of the [Life Sciences Commons](#), and the [Medicine and Health Sciences Commons](#)

---

---

**Authors**

Jason A. Bubier, David P. Hill, Gaurab Mukherjee, Timothy Reynolds, Erich J Baker, Alexander Berger, Jake Emerson, Judith A. Blake, and Elissa J Chesler

---



Original article

# Curating gene sets: challenges and opportunities for integrative analysis

Jason Bubier<sup>1,†</sup>, David Hill<sup>1,†</sup>, Gaurab Mukherjee<sup>1,†,\*</sup>,  
Timothy Reynolds<sup>2,†</sup>, Erich J. Baker<sup>2</sup>, Alexander Berger<sup>1</sup>,  
Jake Emerson<sup>1</sup>, Judith A. Blake<sup>1</sup> and Elissa J. Chesler<sup>1</sup>

<sup>1</sup>The Jackson Laboratory, Bar Harbor, ME 04609, USA and <sup>2</sup>Institute of Biomedical Studies, Baylor University, Waco, TX 76706, USA

\*Corresponding author: Email: Gaurab.Mukherjee@jax.org

†These authors contributed equally as primary authors.

Citation details: Bubier, J., Hill, D., Mukherjee, G. *et al.* Curating gene sets: challenges and opportunities for integrative analysis. *Database* (2019) Vol. 2019: article ID baz036; doi:10.1093/database/baz036

Received 29 October 2018; Revised 25 February 2019; Accepted 26 February 2019

## Abstract

Genomic data interpretation often requires analyses that move from a gene-by-gene focus to a focus on sets of genes that are associated with biological phenomena such as molecular processes, phenotypes, diseases, drug interactions or environmental conditions. Unique challenges exist in the curation of gene sets beyond the challenges in curation of individual genes. Here we highlight a literature curation workflow whereby gene sets are curated from peer-reviewed published data into GeneWeaver (GW), a data repository and analysis platform. We describe the system features that allow for a flexible yet precise curation procedure. We illustrate the value of curation by gene sets through analysis of independently curated sets that relate to the integrated stress response, showing that sets curated from independent sources all share significant Jaccard similarity. A suite of reproducible analysis tools is provided in GW as services to carry out interactive functional investigation of user-submitted gene sets within the context of over 150 000 gene sets constructed from publicly available resources and published gene lists. A curation interface supports the ability of users to design and maintain curation workflows of gene sets, including assigning, reviewing and releasing gene sets within a curation project context.

**Database URL:** <https://www.geneweaver.org/>

## Introduction

Biocuration plays a central role in biomedical research and public data resources, such as the Gene Ontology (GO). Until recently, curation efforts in functional genomics have

focused primarily on gene-by-gene associations. The advent of high-throughput technologies and explosion in biological data have led to the need for collaborative and genome-scale curation.

The GeneWeaver (GW) resource functions as both a multi-species, heterogeneous data store and an analysis platform (1). It is designed to harmonize results from disparate experimental systems and functional genomics data including, but not limited to, differential expression profiling, genome-wide association studies (GWAS), gene networks and literature curation.

Curated, empirical gene sets are supplemented by data automatically aggregated from public resources such as ontological annotations by term from GO, Disease Ontology and Mammalian Phenotype Ontology (MP); pathway databases [e.g. Kyoto Encyclopedia of Genes and Genomes (KEGG) and Pathway Commons (PC)]; and curated repositories (e.g. Comparative Toxicogenomics Database and Molecular Signature Database). Table 1 lists the public resource data that are currently used by GW to create these additional gene sets.

GW recognizes the importance of maintaining current data repositories for reference and data integration. GW supports updates of core species data from model organism databases, NCBI and Ensembl on a quarterly basis. Ontology annotation, UniGene data, probe data reference sets and transcript data are updated bi-monthly or as requested. Furthermore, on a nightly basis, GW creates a data source page, which displays the amount of data contained within the resource and relevant timestamps. This page can be found at <https://www.geneweaver.org/data> and is linked through the software and in the help documentation. GW currently supports 10 species, but additional species may be added upon request at [geneweaver.web@gmail.com](mailto:geneweaver.web@gmail.com).

In addition to its function as a data repository, GW also exposes a suite of analysis tools that users can use to investigate public and curated gene sets. These tools, which operate over collections of gene sets, are based on

statistical, combinatoric and graph-theoretic algorithms (1). This allows users to analyze their own experimental data in the larger context of curated, resource-level data sets.

While there are examples of metadatabases or integrative resources that seek to aggregate studies related to disease or genes and provide analysis tools such as MARRVEL (2), ENRICH (3) and the Open Targets Platform (4), here we report the design and building of a distinctive curation interface into GW. This interface allows groups of curators to manage all aspects of the curation workflow for publications and pre-publication research data including task assignment, literature triage, data import and metadata association. Here we describe the use of GW to curate literature-supported gene sets and give a practical example of how curation of a body of literature can be used to query biological information.

### Capture of gene sets into GW environment

The primary goal of curating literature into GW is to transform published, experimentally supported data into high-quality sets of genes that can be used for computational analysis in the context of gene sets from many sources. Once an article is identified as containing relevant data sets for study, determining how to bin the genes studied in the paper into sets, choosing the most accurate identifiers and documenting relevant metadata become paramount. A general curation strategy is to create separate gene sets for each relevant area of biology in the paper. For example, if the authors report on genes that are up- and down-regulated under two different biological conditions, the creation of four gene sets is appropriate (e.g. up-regulated condition 1, up-regulated condition 2, down-regulated condition 1, down-regulated condition 2). Sometimes authors will bin

**Table 1.** Public data resources incorporated into GW data store

Public resource	Description	Available data sets	Reference
Allen Brain Atlas	Differential expression in adult mouse brain structures	802	(5)
Comparative Toxicogenomics Database	Curated chemical–gene interactions	21 630	(6)
Drug Related Gene Database	Addiction-related experiment data	250	(7)
GO	Gene annotations	85 573	(8)
GWAS Catalog	GWAS summary results	3389	(9)
Human Phenotype Ontology	Gene annotations	6276	(10)
KEGG	Gene network and pathway members	1339	(11)
MeSH	Gene annotations	12 069	(12)
MP	Gene annotations	7931	(13)
Molecular Signatures Database	Curated gene sets	3738	(14)
Online Mendelian Inheritance in Man	Curated disease–gene associations	738	(15)
PC	Gene networks and pathways	1149	(16)

genes in a manuscript according to not only experimental results, but also with respect to some other aspect pertinent to the biology of those genes, such as ‘upregulated transporters’. In those cases, additional gene sets may be created. The general curation guideline for the curation of a gene set is that a set should be created if it relates to a significant piece of biology where it would be useful to compare with other sets. There is no limit to the number of genes appropriate in a set, but creation of intersection and union sets (e.g. all differentially expressed genes, both up- and down-regulated) is not necessary because those can be derived within the GW environment.

GW uses a tab-delimited, two-column format for uploading gene sets as bulk lists. The first column contains the gene identifier. We use organism-specific, persistent gene identifiers when possible, generally obtained from the authoritative model organisms database (MOD) resources, due to the ephemeral nature of gene names and symbols. A unique identifier per gene entity is essential because stability of entity recognition is necessary to reference genomic features independent of genome build or exact genomic location or changes in gene nomenclature. To maintain currency with the MOD resources, GW is frequently updated and seamlessly displays the current gene symbols

for data. GW also accepts a variety of standard identifiers and aliases for non-gene data sets such as molecular identifiers or microarray probe identifiers (Table 2). For empirical data, it is preferable to choose identifiers that reflect precisely what was measured in the experiment (e.g. hybridization to a particular probe sequence), rather than what gene was thought to be targeted by that probe at the time of publication or accession. The current best gene mappings for these probes are stored in the GW system and used in analyses.

The second column in a GW ‘GeneSet upload’ is for a data value or score type. Priority is given to the statistics that are used to associate a particular gene to the GeneSet. This field can also be used to represent values associated with a quantitative measurement of the data such a fold change or z-score. If complete genomic data are available, they are uploaded although the user has the option of filtering significant data in an analysis. The priority for the type of data to include in column two is *Q*-values, *P*-values, effect size, followed by fold change and finally binary (indicating presence or absence in a GeneSet).

A practice central to gene set curation is adherence to metadata standards. There are certain situations that arise that, based on GW curation guidelines, a curator uses their

**Table 2.** Gene identifier types and expression microarray platforms supported in GW

Identifier type	Supported species
CGNC	<i>Gallus gallus</i>
Ensembl gene	All
Ensembl protein	All
Ensembl transcript	All
FlyBase	<i>Drosophila melanogaster</i>
Gene symbol	All
HGNC	<i>Homo sapiens</i>
MGI	<i>Mus musculus</i>
miRBase	All
NCBI gene	All
RGD	<i>Rattus norvegicus</i>
SGD	<i>Saccharomyces cerevisiae</i>
Unigene	All
Wormbase	<i>Caenorhabditis elegans</i>
ZFIN	<i>Danio rerio</i>

Expression microarray platforms		
Manufacturer	Supported species	No. of available platforms
Affymetrix	<i>Caenorhabditis elegans</i> , <i>Drosophila melanogaster</i> , <i>Danio rerio</i> , <i>Homo sapiens</i> , <i>Macaca mulatta</i> , <i>Mus musculus</i> , <i>Rattus norvegicus</i> , <i>Saccharomyces cerevisiae</i>	42
Agilent	<i>Mus musculus</i> , <i>Rattus norvegicus</i>	4
Illumina	<i>Homo sapiens</i> , <i>Mus musculus</i>	4

## GeneSet Information

**Tier III** GS354317 · Genes downregulated in livers of Eif2ak4-mutant mice perfused with all amino acids minus methionine

from Publication Assignment: [157](#)

**DESCRIPTION:** Genes downregulated in livers of Eif2ak4(Gcn2)-mutant mice 30 minutes after perfusion with amino acids lacking methionine. Removing methionine induces amino-acid starvation. Sequences reported in the manuscript were used to search the MGI resource using the batch query form on June 15, 2018. Sequences that did not match were investigated manually. Sequences that were associated with putative assignments to gene identifiers were included. Values represent z-scores. Higher scores were used for genes represented more than once.

**LABEL:** Downregulated in Eif2ak4-/- no Met

**SCORE TYPE:** Effect

**DATE ADDED:** 2018-06-22

**DATE UPDATED:** 2019-01-07

**SPECIES:** Mus musculus

**AUTHORS:** Dang Do AN, Kimball SR, Cavener DR, Jefferson LS

**TITLE:** *elf2alpha kinases GCN2 and PERK modulate transcription and translation of distinct sets of mRNAs in mouse liver.*

**JOURNAL:** *Physiological genomics* Aug 2009, Vol 38, pp. 328-41

**ABSTRACT:** In eukaryotes, selective derepression of mRNA translation through altered utilization of upstream open reading frames (uORF) or internal ribosomal entry sites (IRES) regulatory motifs following exposure to stress is regulated at the initiation stage through the increased phosphorylation of eukaryotic initiation factor 2 on its alpha-subunit (eIF2alpha). While there is only one known eIF2alpha kinase in yeast, general control nonderepressible 2 (GCN2), mammals have evolved to express at least four: GCN2, heme-regulated inhibitor kinase (HRI), double-stranded RNA-activated protein kinase (PKR), and PKR-like endoplasmic reticulum-resident kinase (PERK). So far, the main known distinction among these four kinases is their activation in response to different acute stressors. In the present study, we used the in situ perfused mouse liver model and hybridization array analyses to assess the general translational response to stress regulated by two of these kinases, GCN2 and PERK, and to differentiate between the downstream effects of activating GCN2 versus PERK. The resulting data showed that at least 2.5% of mouse liver mRNAs are subject to derepressed translation following stress. In addition, the data demonstrated that eIF2alpha kinases GCN2 and PERK differentially regulate mRNA transcription and translation, which in the latter case suggests that increased eIF2alpha phosphorylation is not sufficient for derepression of translation. These findings open an avenue for more focused future research toward groups of mRNAs that code for the early cellular stress response proteins. **PUBMED: 19509078**  
Find other GeneSets from this publication [Q](#)

- Export Data
- Similar GeneSets
- Request Curation
- Add Geneset to Project
- Share Geneset w/ Group(s)

## Annotation Information

Adult Mouse Anatomy Gene Ontology

liver (MA:0000358) organ system (MA:0000003) mouse anatomy (MA:0000001) digestive system (MA:0002431)

visceral organ system (MA:0000019) cellular response to amino acid starvation (GO:0034198)

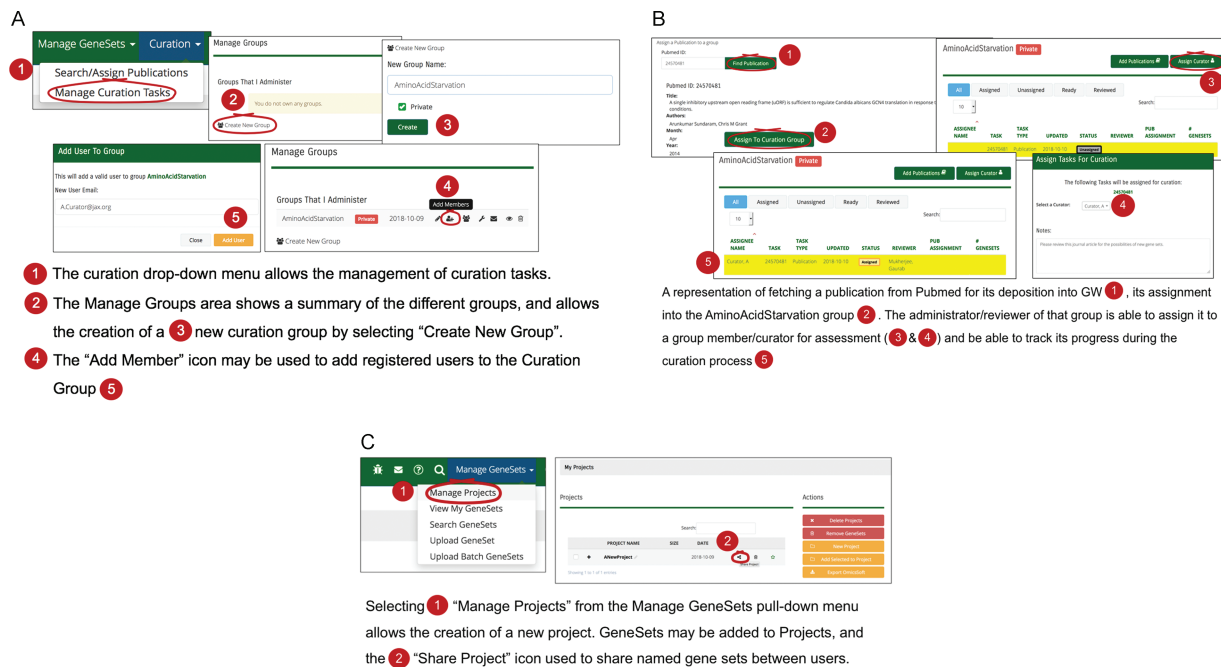
**Figure 1.** Sample GeneSet from GW showing the curator-populated fields of Gene Set Name, Gene Set Figure Label, Gene Set Description and Ontology Annotations. Publication data are populated by providing the PMID during the set upload.

scientific judgment to accurately capture the appropriate data from a publication. When, for example, more than one result is reported for a gene, a common occurrence with legacy microarray data, a consistent approach must be taken in capturing that data. In this situation, the curator would report the gene once, populating column two with either the strongest statistical association score or with the highest fold change value associated with the identical gene(s). In other situations, the fold change for a given gene is not reported as a value but as infinite fold change usually because in one of the reference samples the gene was undetectable. In these cases, the curator chooses a specific value for the gene-value pair to distinguish this situation.

Any of these kinds of rules used in the curation of a gene set will be noted in the gene set description. These rules are described in the curation standard document (<http://geneweaver.org/help/>) and are mentioned in the GeneSet descriptions.

Creation of a gene set also requires capture of additional metadata that is used for searching and the classification of the gene set within the GW environment (Figure 1). The following fields are necessarily included with the curation of a gene set from a curated manuscript:

**Gene Set Name:** a brief title for the gene set, approximately sentence length, which should provide a clear and concise



**Figure 2.** (A) A workflow diagram representing the creation of a new curation group. (B) A diagram showing the management of curation tasks and their assignment to group members. (C) The creation of a new project and the subsequent assignment of gene sets to it, then allows the sharing of that gene set among users.

description of the contents of a gene set interpretable to most users of GW, but with sufficient detail to satisfy a domain expert. This is the major gene set name that is displayed in all search results, project directory and table views of analysis results.

**Gene Set Figure Label:** a brief 23-character abbreviation to facilitate recognition of the gene set in a graph, analysis result or other display.

**Gene Set Description:** a detailed description of the gene set, including rules for its construction, experimental methods and analyses used to generate data, anatomical terms and traceable references to source data including accession information and date. Abbreviations should be avoided.

**Ontology Annotations:** ontology annotations are assigned to gene sets. These appear under the heading of annotation information. Terms from Disease Ontology, Mammalian Phenotype Ontology, Adult Mouse Anatomy, EMBRACE Data and Methods, Chemical Entities of Biological Interest, Medical Subject Headings (MeSH), Experimental Factor Ontology and Human Phenotype Ontology are provided. This is facilitated through the application of the Monarch or NCBO Annotator to provide textual descriptions, including publication abstracts and manually reviewed. Curators may assign terms to gene sets outside of the recommended terms and remove false-positive recommendations.

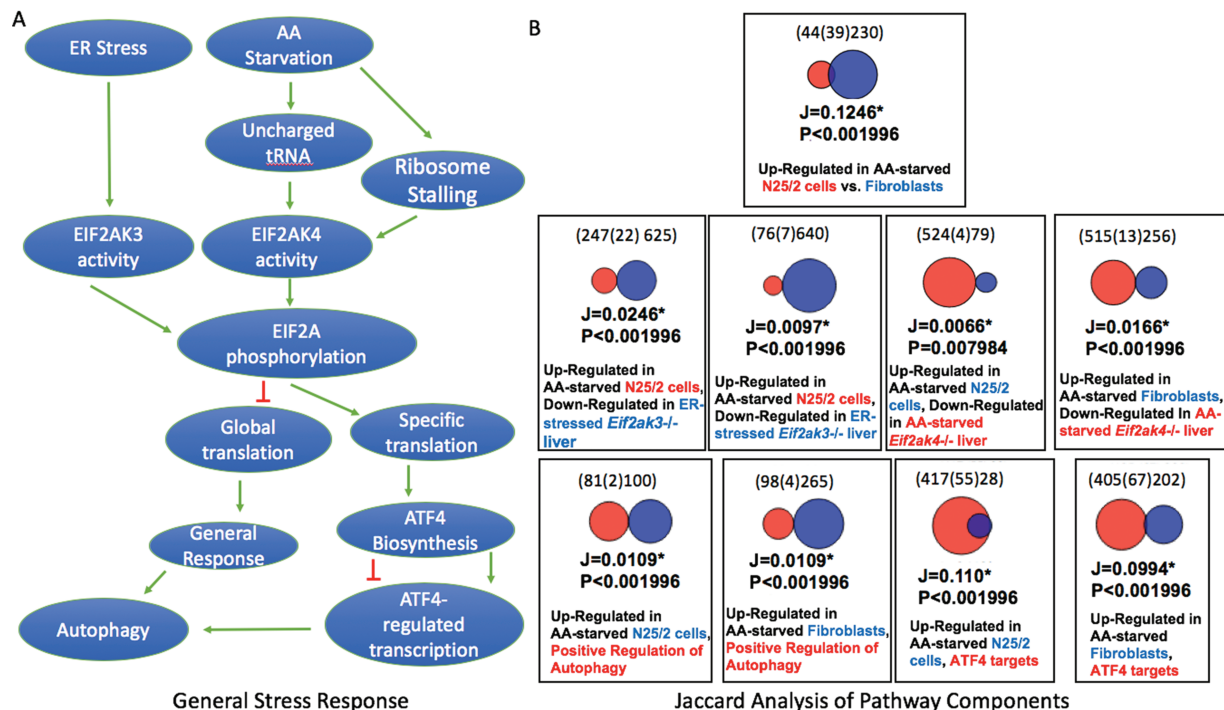
**Publication Information:** PubMed ID, Title, authors, publication information and full text of the abstract.

## Curation workflows and strategies

The curation pull-down menu in GW provides various strategies for managing curation tasks such as searching and assigning publications from peer-reviewed journals. The workflow process allows the creation of a curation group, i.e. a group of curators that have GW editorial accounts (Figure 2A). A lead curator can assign curation tasks to members of the curation group, typically assigning publications for review and gene set entry into the resource to specific curators. Once the gene sets are created for a curation task, the lead curator may choose to review each gene set prior to approval (Figure 2B). Once accepted, the gene set becomes available based on its sharing restrictions. Sharing restrictions can be set that allow data sets still under preliminary analysis by a research group to be private, all the way to having data sets that are freely and publicly available. Once public, a gene set may be added to a 'project' and used for analysis (Figure 2C). A project typically consists of a set of related gene sets that have been defined within the GW environment concerning a biological hypothesis that is under investigation. Related gene sets may be grouped into projects that can be shared publicly or within analysis teams.

## Curated sets can be used to investigate relationships between biological processes

To assess the utility of expert curation into the GW resource, we focused on a well-known pathway i.e. the integrated



**Figure 3.** (A) A schematic representation of two branches of the ISR. Unfolded protein ER stress or amino acid starvation activates two separate kinases, EIF2AK3 or EIF2AK4, which phosphorylates EIF2A. Phosphorylated EIF2A then represses general translation in the cell, but stimulates translation of a subset of response genes including the transcription factor ATF4. A consequence of ATF4 activation is the downstream activation of autophagy. (B) The results of Jaccard similarity analyses using independently curated gene sets that represent different aspects of the ISR. These results show significant similarity between genes up-regulated after amino acid starvation and genes down-regulated in EIF2A kinase mutant cells, ATF4 target genes and genes annotated with the GO term ‘positive regulation of autophagy’ (GO:0010508, downloaded from a Mouse Genome Informatics Query performed on 17 August 2018).

**Table 3.** Gene sets used for analysis of the ISR

Gene set identifier	# of genes	Gene set title	Reference
GS355023	102	Positive regulation of autophagy	(8)
GS354317	528	Genes down-regulated in livers of Eif2ak4-mutant mice perfused with all amino acids minus methionine	(23)
GS354521	647	Genes down-regulated in livers of Eif2ak3-mutant mice treated with tBuHQ	(23)
GS354534	83	Genes up-regulated in amino acid-starved N25/2 cells	(24)
GS355584	472	Atf4 target genes	(25)
GS354663	269	Genes up-regulated in amino acid-starved fibroblasts	(26)

stress response (ISR) pathway, curating genes sets relevant to the execution of the pathway and comparing those gene sets using the Jaccard analysis tool. Jaccard statistics are widely used to estimate the similarity of sets and originate from studies of species habitation of ecological zones. Briefly, these statistics estimate the positive match of categorical data and are considered a preferred alternative to set matching methods that are biased by the large number of negative results that are often found in comparison of genomic data sets. The simplest Jaccard statistic takes the intersection size over the size of the union of two gene sets. In GW, the size of the union is based on the possible matches

between two gene sets, such that gene products that could only be present in one set are excluded, e.g. genes for which there is no known orthologue.

Our hypothesis was that curation of a group of gene sets that represent gene expression associated with steps of a well-defined pathway would result in significant Jaccard overlap of the gene sets. The ISR pathway is a well-conserved and well-studied pathway that is amenable to this type of analysis (17). Two branches of the ISR are shown in Figure 3A. The pathway begins with the detection of a cellular stress such as amino acid starvation or unfolded proteins that then activate a protein kinase that



phosphorylates EIF2A (18, 19). The phosphorylation of EIF2A results in the repression of global translation, but the activation of a small set of genes that coordinate the response (20). One of those genes is ATF4, which regulates transcription of downstream response genes (21). Some of those genes in turn up-regulate autophagy (22). To illustrate the value of curated gene sets in the GW resource, we independently curated sets that were relevant to the ISR. The sets included genes that were up-regulated after amino acid starvation in two cell lines, genes that were down-regulated in amino acid-starved *Eif2ak4* (GCN2) mutant livers, genes that were down-regulated in endoplasmic reticulum (ER) stress *Eif2ak3* (PERK) livers, genes that bound the ATF4 transcription factor within 3 kb of their promoters and mouse genes that are annotated to the GO term ‘positive regulation of autophagy’ (Table 3). Our results show that the gene sets up-regulated in the amino acid-starved cells overlapped significantly with each other,  $P < 0.002$  (the upper limit value reported by GW) and overlapped significantly with the gene sets identifying the various downstream steps of the ISR,  $P < 0.002$ .

## Conclusions

GeneWeaver in contrast to many bioinformatics resources, provides a curation component that focuses on gene sets as the reportable entity. Working from both the data-download and the literature-curation aspects, GW now comprises over 150 000 high-quality gene sets that are accessible for data analysis. Selected gene sets can be defined and analyzed within the GW analysis platform, and/or provisional gene sets can be analyzed in a restricted access environment. The ability to compute on gene sets from multiple resources is essential in today’s genome systems and comparative functional analysis environments. These analyses allow experimental data to be immediately placed in the context of other experiments and of rigorously curated pathways and gene functional annotations.

## Acknowledgements

The authors would like to thank Robert Burgess, Emily Spaulding and Beena Kadhakhuza for their suggestions and comments toward this work.

## Funding

National Cancer Institute at the National Institutes of Health (P30CA034196); The Jackson Laboratory Center for Precision Genetics of the National Institutes of Health (U54 OD020351); and the National Institute on Drug Abuse and the National Institute on Alcohol Abuse and Alcoholism at the National Institutes of Health (R01 AA18776).

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

*Conflict of interest.* None declared.

## References

1. Baker,E.J., Jay,J.J., Bubier,J.A. *et al.* (2012) GeneWeaver: a web-based system for integrative functional genomics. *Nucleic Acids Res.*, **40**, D1067–D1076.
2. Wang,J., Al-Ouran,R., Hu,Y. *et al.* (2017) MARRVEL: integration of human and model organism genetic resources to facilitate functional annotation of the human genome. *Am. J. Hum. Genet.*, **100**, 843–853.
3. Kuleshov,M.V., Jones,M.R., Rouillard,A.D. *et al.* (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.*, **44**, W90–W97.
4. Koscielny,G., An,P., Carvalho-Silva,D. *et al.* (2017) Open Targets: a platform for therapeutic target identification and validation. *Nucleic Acids Res.*, **45**, D985–D994.
5. Lein,E.S., Hawrylycz,M.J., Ao,N. *et al.* (2007) Genome-wide atlas of gene expression in the adult mouse brain. *Nature*, **445**, 168–176.
6. Davis,A.P., Grondin,C.J., Johnson,R.J. *et al.* (2017) The Comparative Toxicogenomics Database: update 2017. *Nucleic Acids Res.*, **45**, D972–D978.
7. Gardner,D., Akil,H., Ascoli,G.A. *et al.* (2008) The neuroscience information framework: a data and knowledge environment for neuroscience. *Neuroinformatics*, **6**, 149–160.
8. Ashburner,M., Ball,C.A., Blake,J.A. *et al.* (2000) Gene ontology: tool for the unification of biology. *Nat. Genet.*, **25**, 25–29.
9. MacArthur,J., Bowler,E., Cerezo,M. *et al.* (2017) The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.*, **45**, D896–D901.
10. Köhler,S., Doelken,S.C., Mungall,C.J. *et al.* (2014) The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Res.*, **42**, D966–D974.
11. Kanehisa,M., Furumichi,M., Tanabe,M. *et al.* (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.*, **45**, D353–D361.
12. Rogers,F.B. (1963) Medical subject headings. *Bull. Med. Libr. Assoc.*, **51**, 114–116.
13. Smith,C.L. and Eppig,J.T. (2012) The mammalian phenotype ontology as a unifying standard for experimental and high-throughput phenotyping data. *Mamm. Genome*, **23**, 653–668.
14. Liberzon,A., Subramanian,A., Pinchback,R. *et al.* (2011) Molecular signatures database (MSigDB) 3.0. *Bioinformatics*, **27**, 1739–1740.
15. Hamosh,A., Scott,A.F., Amberger,J.S. *et al.* (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.*, **33**, D514–D517.
16. Cerami,E.G., Gross,B.E., Demir,E. *et al.* (2011) Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.*, **39**, D685–D690.

17. Pakos-Zebrucka,K., Koryga,I., Mnich,K. *et al.* (2016) The integrated stress response. *EMBO Rep.*, **17**, 1374–1395.
18. Ishimura,R., Nagy,G., Dotu,I. *et al.* (2016) Activation of GCN2 kinase by ribosome stalling links translation elongation with translation initiation. *Elife*, **5**, 1–22.
19. Zhang,P., McGrath,B.C., Reinert,J. *et al.* (2002) The GCN2 eIF2alpha kinase is required for adaptation to amino acid deprivation in mice. *Mol. Cell. Biol.*, **22**, 6681–6688.
20. Harding,H.P., Novoa,I., Zhang,Y. *et al.* (2000) Regulated translation initiation controls stress-induced gene expression in mammalian cells. *Mol. Cell*, **6**, 1099–1108.
21. Vattam,K.M. and Wek,R.C. (2004) Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.*, **101**, 11269–11274.
22. B'chir,W., Maurin,A.-C., Carraro,V. *et al.* (2013) The eIF2 $\alpha$ /ATF4 pathway is essential for stress-induced autophagy gene expression. *Nucleic Acids Res.*, **41**, 7683–7699.
23. Dang Do,A.N., Kimball,S.R., Cavener,D.R. *et al.* (2009) eIF2 $\alpha$  kinases GCN2 and PERK modulate transcription and translation of distinct sets of mRNAs in mouse liver. *Physiol. Genomics*, **38**, 328–341.
24. Hellsten,S.V., Lekholm,E., Ahmad,T. *et al.* (2017) The gene expression of numerous SLC transporters is altered in the immortalized hypothalamic cell line N25/2 following amino acid starvation. *FEBS Open Bio*, **7**, 249–264.
25. Han,J., Back,S.H., Hur,J. *et al.* (2013) ER-stress-induced transcriptional regulation increases protein synthesis leading to cell death. *Nat. Cell Biol.*, **15**, 481–490.
26. Deval,C., Chaveroux,C., Maurin,A.-C. *et al.* (2009) Amino acid limitation regulates the expression of genes involved in several specific biological processes through GCN2-dependent and GCN2-independent pathways. *FEBS J.*, **276**, 707–718.