# Construction of a searchable database for gene expression changes in spinal cord injury experiments

Eric C. Rouchka[1,2,3,*], Carlos de Almeida[4,5], Randi B. House[5,6], Jonah C. Daneshmand[3], Julia H. Chariker[2,7], Sujata Saraswat-Ohri[5,8], Cynthia Gomes[5,9], Morgan Sharp[5,8], Alice Shum-Siu[5,8], Greta M. Cesarz[5], Jeffrey C. Petruska[5,8,9], David S.K. Magnuson[4,5,8,9]

[1]Department of Biochemistry and Molecular Genetics, University of Louisville School of Medicine, University of Louisville, Louisville, KY USA
[2]Kentucky IDeA Networks of Biomedical Research Excellence (KY INBRE) Bioinformatics Core, University of Louisville School of Medicine, 522 East Gray Street, Louisville, KY USA 40202
[3]Bioinformatics Program, School of Interdisciplinary and Graduate Studies, University of Louisville, Louisville, KY
[4]Translational Neuroscience Program, School of Interdisciplinary and Graduate Studies, University of Louisville, Louisville, KY
[5]Kentucky Spinal Cord Injury Research Center, School of Medicine, University of Louisville, Louisville, KY
[6]Department of Bioengineering, Speed School of Engineering, University of Louisville, Louisville, KY
[7]Department of Neuroscience Training, School of Medicine, University of Louisville, Louisville, KY
[8]Department of Neurological Surgery, School of Medicine, University of Louisville, Louisville, KY USA
[9]Department of Anatomical Sciences and Neurobiology, School of Medicine, University of Louisville, Louisville, KY

*email: eric.rouchka@louisville.edu

## ABSTRACT

Spinal cord injury (SCI) is a debilitating disease resulting in an estimated 18,000 new cases in the United States on an annual basis. Significant behavioral research on animal models has led to a large amount of data, some of which has been catalogued in the Open Data Commons for Spinal Cord Injury (ODC-SCI). More recently, high throughput sequencing experiments have been utilized to understand molecular mechanisms associated with SCI, with nearly 6,000 samples from over 90 studies available in the Sequence Read Archive. However, to date, no resource is available for efficiently mining high throughput sequencing data from SCI experiments. Therefore, we have developed a protocol for processing RNA-Seq samples from high-throughput sequencing experiments related to SCI resulting in both raw and normalized data that can be efficiently mined for comparisons across studies as well as homologous discovery across species. We have processed 1,196 publicly available RNA-seq samples from 50 bulk RNA-Seq studies across nine different species, resulting in an SQLite database that can be used by the SCI research community for further discovery. We provide both the database as well as a web-based front-end that can be used to query the database for genes of interest, differential gene expression, genes with high variance, and gene set enrichments.

**Keywords**: spinal cord injury, SCI, ODC-SCI, RNA-Seq, bulk RNA-Seq, SQLite, transcriptomics, differential gene expression

## INTRODUCTION

**Spinal Cord Injury (SCI).** Spinal cord injury (SCI) is a debilitating disease resulting in an estimated 18,000 new cases in the United States annually, with over 80% of the cases caused by auto accident, fall, gunshot wound, motorcycle accident, or diving [1-3]. Approximately 1.4 million Americans are living with SCI [4]. Over the last 30 years, significant strides have been made in understanding the systems-wide pathophysiology and behavioral components affected by SCI, including axon growth [5-10], compensatory sprouting [11-17], glial cell roles [18-22], gut dysbiosis [23-26], and the role exercise plays in recovery [27-36]. However, successful translation of potential therapies remains elusive [37]. To capture and understand the heterogeneity of SCI and recovery from injury, researchers created the Open Data Commons for Spinal Cord Injury (odc-sci.org) [38] in 2017 to provide a platform for sharing SCI-related data adhering to the FAIR principles of Findable, Accessible, Interoperable, and Resuable [39]. As of 9/28/2022, a total of 176 complex datasets from 95 labs had been uploaded to ODC-SCI, and the structured data dictionaries that accompany each data set allow the data to be easily accessible and interoperable. While ODC-SCI contains a vast amount of data, one component that it does not catalog is transcriptomic data associated with RNA-Seq studies, largely because this data is generally easily accessible in resources such as the Gene Expression Ominibus (GEO) [40] or the Sequence Read Archive (SRA) [41]. While data from SRA and GEO are accessible, they are not generally interoperable to the average researcher to perform meta-analysis comparisons across experiments.

**High Throughput Datasets.** The expansion in the utilization of high-throughput methodologies such as next generation sequencing along with numerous behavioral studies has led to the public availability of sets of data across the genome-to-phenome spectrum. Efforts to make this data usable to the larger scientific community have pushed for the data to adhere to the principles of being findable, accessible, interoperable, and reusable, or FAIR compliant [39]. The issue of making this data available was first

recognized in a formal manner with microarray experiments, leading to the creation of the Minimum Information About a Microarray Experiment (MIAME) [42]. The MIAME standards led to guidelines for submission of high-throughput datasets to publicly available repositories such as the Gene Expression Omnibus (GEO) [40], The Sequence Read Archive (SRA) [41], ArrayExpress [43], and the Database of Genotypes and Phenotypes (dbGAP) [44]. More specifically, for SCI research, standards were created for describing the minimal information about a spinal cord experiment (MIASCI) [45].

Since the first RNA-Seq dataset related to spinal cord injury was made available in 2013 [46], the number of high throughput sequencing SCI datasets (including RNA-Seq) has grown exponentially (**Figure 1**). More recently, efforts to make sequencing data interoperable has extended into disease-specific domains, with cancer leading the way through The Cancer Genome Atlas (TCGA) [47] and Genomics Data Commons (GDC) [48]. Other efforts have been made to do this at tissue and gene levels through the Gene Tissue Expression (GTEx) database [49]. However, these projects typically classify samples in a binary fashion, into healthy or diseased states. This excludes other potentially pertinent characteristics. In the case of SCI, these additional characteristics may include organism (e.g. rat, mouse, zebrafish), injury type (e.g. contusion, transection, sham), injury force (e.g. 60 kdyn, 200 kdyn, 25 g/cm), injury level (e.g. C5, 5$^{th}$ cervical segment, T10, 10$^{th}$ thoracic segment), tissue type (e.g. spinal cord tissue at epicenter, spinal cord tissue below injury, isolated astrocytes from injured spinal cord, dorsal root ganglion), and time since injury (e.g. 1 hour post-injury, 24 hours post-injury, 60 days post-injury). Given the shortcomings with the interoperability of SCI transcriptomic data, we have developed a pipeline for preparing publicly available datasets for comparison across studies as well as across model organisms, resulting in both an SQLite database for the raw data, as well as a web interface for integrative analysis across studies.

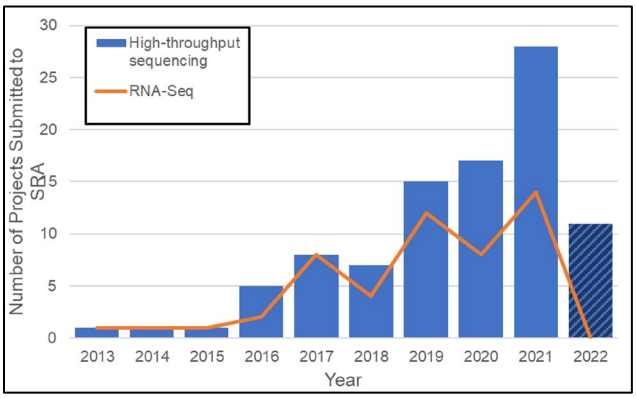


**Figure 1: Growth in number of high-throughput sequencing and RNA-Seq projects related to "spinal cord injury" or SCI in SRA, by release year.**

## METHODS

**Data selection and preparation**. Samples were located within the Sequence Read Archive (SRA) [41] using the search term "Spinal Cord Injury" OR SCI. From these, a total of 90 potential studies of interest were identified (**Supplemental Table 1**). Three additional SRA records related to neuropathic pain (SRP173586), nerve injury (SRP13362), and central nervous system (CNS) injury (SRP094587) were manually added, as were additional samples from the identified SRA studies that were not automatically included, due to differences in the specific search terms. A total of 5,891 samples were identified from these 93 studies, falling into one of twelve categories. The number of samples is inflated, due to one single cell RNA-Seq (scRNA-seq) study (SRP239303) where each of the 3,687 cells sequenced is represented as its own sample. Those samples corresponding to an RNA-seq assay were manually reclassified to the most appropriate assay type as RNA-seq (bulk), scRNA-seq, ribosomal-associated RNA-seq (RAM-seq), and transplanted RNA-seq. A handful of assay types were not consistently labeled and were reclassified from RNA-seq to microRNA (miRNA-seq) or noncoding RNA-Seq (ncRNA-Seq). The final filtered data for each of these classifications is shown in **Table 1**. The 50 bulk RNA-seq studies (**Table 2**) were selected for further processing, representing 1,196 samples from nine different species, the majority being mouse, rat, frog, and human (**Table 3**). RNA-Seq samples were downloaded from SRA using the sratoolkit [50]:

```
fasterq-dump -skip-technical -x -p -split-3
        <SRR RUN NUMBER>
```

**Table 1:** Classification of samples identified.

| Sequencing Type | Number of Studies | Number of Samples |
|---|---|---|
| Single cell RNA-Seq | 15 | 4,062 |
| RNA-Seq (bulk) | 50 | 1,196 |
| Amplicon | 9 | 374 |
| miRNA-Seq | 6 | 118 |
| Other | 2 | 68 |
| RAM-Seq | 1 | 23 |
| ncRNA-Seq | 3 | 22 |
| Transplanted cells | 1 | 19 |
| ChIP-Seq | 1 | 4 |
| RIP-Seq | 1 | 4 |
| ATAC-Seq | 1 | 2 |
| WGS | 1 | 2 |
| CLONE | 1 | 1 |
| Targeted capture | 1 | 1 |
| WXS | 1 | 1 |

Eleven rat samples were omitted (SRR13562711-SRR13562722) due to issues with obtaining the samples with `sratoolkit`.

Samples were then processed for quality assurance and quality control (QA/QC) using `fastQC` [51].

Sequencing reads were aligned to the respective reference genomes guided by known transcriptomes (**Table 4**) using `STAR` [52]. Raw read counts were determined using the `featureCounts` R package [53]. Counts were extracted using three separate stranded settings, including `-s 0` (unstranded), `-s 1` (forward stranded), and `-s 2` (reverse stranded). This step was performed to infer the most likely strandedness of the original library construction kit. Strandedness was also calculated using the `infer_experiment.py` script from the `RSeqC` package [54]. The resulting feature counts were reformatted to the same format of `htSeqCount` [55] using a custom script. The count information was then parsed to create a tab-delimited file of raw gene counts and TPM (transcripts per million) normalized counts [56]. Since TPM normalizes to both the library size and the transcript size, we utilized the R `GenomicFeatures` package [57] to determine gene lengths. The exception to this was the axolotl data, where a known transcriptome description is not

**Table 2:** SRA RNA-Seq studies included. Species codes are listed in Table 3.

| SRA Study ID | Org | GEO ID | Study Title | Pubmed |
|---|---|---|---|---|
| SRP255811 | Am | | Preclinical molecular signatures of spinal cord functional restoration: optimizing the metamorphic axolotl (Ambystoma mexicanum) model in regenerative medicine | |
| SRP255836 | Am | | Identification of the molecular signatures and functional restoration of the spinal cord of metamorphic axolotl | |
| SRP334274 | Dr | GSE182869 | Next Generation Sequencing of zebrafish intraspinal serotonergic neurons in the injury segment and distal segments after spinal cord injury (zebrafish) | 34876587 |
| SRP259365 | Hs | GSE149664 | Generation of induced motor neurons (iMNs) from human fibroblasts facilitates locomotor recovery after spinal cord injury | 32571478 |
| SRP265127 | Hs | GSE151371 | Blood biomarkers for spinal cord injury (human) | 33512429 |
| SRP220569 | Md | | Identification of regenerative processes in neonatal spinal cord injury in the opossum (Monodelphis domestica) | |
| DRP003667 | Mm | | Genome-wide expression analysis of reactive astrocytes in the injured spinal cord at 7 days after spinal cord injury, host astrocytes in the naive spinal cord, and transplanted astrocytes in the naive spinal cord at 7 days after being transplanted | |
| DRP003669 | Mm | | Genome-wide expression analysis in the naive spinal cord and the injured spinal cord at 14 day after spinal cord injury | |
| SRP019916 | Mm | GSE45376 | RNA-Seq characterization of spinal cord injury transcriptome in acute/subacute phases: a resource for understanding the pathology at the systems level | 23951329 |
| SRP049253 | Mm | GSE62698 | Spinal cord injury (RNA sequencing data) | 25385836 |
| SRP067494 | Mm | GSE76097 | In vivo analysis of astrocyte ribosome-associated mRNA after traumatic spinal cord injury | 27027288 |
| SRP079387 | Mm | GSE84737 | Macrophage transcriptional profile identifies lipid catabolic pathways that can be therapeutically targeted after spinal cord injury | 28130359 |
| SRP094587 | Mm | GSE90908 | Characterization of meningeal type 2 innate lymphocytes and their response to CNS injury | 27994070 |
| SRP097644 | Mm | GSE93976 | In vivo analysis of injury sites presenting full or attenuated pericyte-derived scarring after spinal cord injury (SCI) | 29502968 |
| SRP101665 | Mm | GSE96054 | Time-course analysis of astrocyte-specific RNA-seq in two severities of spinal cord injury | 27716282 |
| SRP101667 | Mm | GSE96055 | Time-course analysis of microglia-specific RNA-seq in two severities of spinal cord injury | 28420963 |
| SRP133622 | Mm | GSE111216 | Mouse transcriptomics reveals extracellular matrix organization as a major pathway involved in inflammatory and neuropathic pain | 30763288 |
| SRP142367 | Mm | GSE113566 | Microglia and macrophages promote corralling, wound compaction and recovery in spinal cord injury via Plexin-B2 | 32112058 |
| SRP173586 | Mm | GSE123919 | Translational profiling of dorsal root ganglia and spinal cord in a mouse model of neuropathic pain | 30906902 |
| SRP179750 | Mm | GSE125176 | Cellular response of mesenchymal stem cells transplanted into spinal cord injury (house mouse) | 30944028 |
| SRP201114 | Mm | GSE132552 | Transcriptional changes after spinal cord injury: recruitment of afferents distal to the site of injury | |
| SRP226573 | Mm | GSE130227 | Syngeneic, in contrast to allogeneic, mesenchymal stem cells have superior therapeutic potential following spinal cord injury | |
| SRP259320 | Mm | GSE149646 | Ascending dorsal column sensory neurons respond to spinal cord injury and downregulate genes related to lipid metabolism (house mouse) | 33431991 |
| SRP269775 | Mm | GSE153720 | Systematic analysis of purified astrocytes after spinal cord injury unveils lncRNA Zeb2os as a novel molecular target for astrogliosis [RNA-Seq] (house mouse) | 33535036 |
| SRP313384 | Mm | GSE171441 | Gsx1 Promotes Locomotor Functional Recovery After Spinal Cord Injury | 33895323<br>34343529 |
| SRP325651 | Mm | GSE178930 | Next Generation Sequencing Facilitates Quantitative Analysis of Wild and spinal cord injury mice (house mouse) | |
| SRP101364 | Pm | GSE95686 | RNA-Seq analysis after spinal cord injury in lamprey reveals distinct transcriptional responses during functional recovery in spinal cord and brain | 29335507 |
| SRP049326 | Rn | GSE62760 | T Cell Deficiency in Spinal Cord Injury: Altered Locomotor Recovery and Whole-Genome Transcriptional Analysis | 26546062 |
| SRP073355 | Rn | | Transcriptome of Sprague-Dawley Rats -- spinal cord contusion | |
| SRP096190 | Rn | GSE93249 | RNA-Seq analysis of coding and long non-coding RNAs in the sub-chronic and chronic stages of spinal cord injury | 28106101 |
| SRP131816 | Rn | GSE109902 | Transcriptional screen in the target region of sprouting hindlimb corticospinal fibers after thoracic spinal cord injury in rats (Norway rat) | |
| SRP149309 | Rn | GSE115067 | Integrated systems analysis reveals conserved gene networks underlying response to spinal cord injury (Norway rat) | 30277459 |
| SRP166392 | Rn | | Following 6 days differentiation of neural stem cells in vitro | |

**Table 2:** (continued)

| SRA Study ID | Org | GEO ID | Study Title | Pubmed |
|---|---|---|---|---|
| SRP176640 | Rn | GSE124819 | Activity-induced changes in the liver transcriptome after chronic spinal cord injury | 31197156<br>32226812 |
| SRP179652 | Rn | GSE125134 | Analysis of possible mechanisms behind functional recovery following neural progenitor cell transplantation into spinal cord injury. (Norway rat) | 31031190 |
| SRP181953 | Rn | GSE125630 | Transcriptome of dorsal root ganglia caudal to a spinal cord injury with modulated behavioral activity | 31175296<br>32226812 |
| SRP192162 | Rn | GSE129694 | Transcriptional changes in soleus muscle for rats exposed to different activities after contusion injury to the spinal cord and transcriptional changes in soleus muscle with complete spinal cord transection injury | 32226812 |
| SRP202013 | Rn | GSE133093 | Brainstem control of transcription after spinal cord injury (SCI) | 31803022 |
| SRP213314 | Rn | | Transcriptomic analysis of knockdown of ?-synuclein after T3 spinal cord injury in rats | |
| SRP216808 | Rn | GSE135080 | Novel drug-like RAR-Beta agonist induces BRCA1 to prevent neuropathic pain (Norway rat) | 31726373 |
| SRP224959 | Rn | GSE138637 | Genome wide analysis of thoracic spinal cord at 5 days after T9 hemisection injury. | |
| SRP273616 | Rn | | Ketogenic diet-mediated steroid metabolism reprogramming improves the immune microenvironment and myelin growth of spinal cord injury rats through gene analysis and co-expression network analysis | |
| SRP275629 | Rn | GSE155610 | Transcriptome of Subcortical White Matter and Spinal Cord After Spinal Injury and Cortical Stimulation | 34267212 |
| SRP279076 | Rn | GSE156999 | Circular RNAs expression profiles and potential key molecules incompletely transected spinal cord injury | |
| SRP311591 | Rn | | Sequencing for spinal cord injury | |
| SRP082501 | Ts | | Pool of tissues of Trachemys scripta elegans | |
| DRP006873 | Xl | | RNA-seq analysis after spinal cord injury in Xenopus laevis | |
| SRP222957 | Xl | GSE137844 | Comparative Gene Expression Profiling between Xenopus Optic Nerve and Spinal Cord Injury to Identify Genes Involved in Successful Regeneration of Vertebrate CNS Axons | 32758133<br>34979916 |
| SRP300206 | Xl | GSE164204 | Cellular response to spinal cord injury in regenerative and non-regenerative stages in Xenopus laevis | 33526076 |
| SRP302901 | Xl | GSE165343 | High expression profiling analysis of the early response to spinal cord injury identified a key role for mTORC1 signaling | 34686684 |

currently available. In this case, those samples were constructed into *de novo* transcriptomes using Trinity [58].

**Table 3:** Samples by species.

| Species name | Common name | Species Code | Number of RNA-Seq samples |
|---|---|---|---|
| *Mus musculus* | House mouse | Mm | 569 |
| *Rattus norvegicus* | Norway rat | Rn | 297 |
| *Xenopus laevis* | African clawed frog | Xl | 176 |
| *Homo sapiens* | Human | Hs | 65 |
| *Mondelphis domestica* | Gray short-tailed opossum | Md | 42 |
| *Petromyzon marinus* | Sea lamprey | Dm | 22 |
| *Danio rerio* | Zebrafish | Dr | 11 |
| *Ambystoma mexicanum* | Axolotl | Am | 8 |
| *Trachemys scripta elegans* | Red-eared slider turtle | Ts | 6 |

**SQLite database construction**. Tab-delimited files representing data for ten SQLite tables were constructed, and then parsed into the SQLite database. Each table was connected to other tables via foreign keys, including the organism, gene identifier, SRA study identifier, and SRA sample identifier.

**Data exploration**. When a data exploration study is chosen, the samples of interest are filtered by species using the raw sequencing reads determined by `featureCounts`. If samples from at least one of the groups mouse and human; mouse and rat; human and rat; or mouse, human, and rat are selected, then the genes are filtered for homologous genes across species using the NCBI Homologene identifiers [59]. Differential gene expression is calculated for the samples within each species individually, as well as for homologous genes across species using `DESeq2` [60]. The resulting data is sorted by q-value (Storey's FDR correction) from smallest to largest and parsed for up-regulated genes (those with a positive logFC) and down-regulated genes (those with a negative logFC). Heatmaps for the top 250 differentially expressed and top 250 variance genes are constructed, using a gene-wise z-score. Enriched GeneOntology Biological Processes [61] and KEGG pathways [62] are determined using `ClusterProfiler` [63].

## RESULTS

**RNA-Seq data.** A total of 1,196 samples from 50 bulk RNA-Seq studies were processed for inclusion in the SQLite database, with the majority originating from either spinal cord or dorsal root ganglion tissue (**Figure 2**). The injury site for these studies was varied, with the most common location thoracic, in particular between thoracic vertebrae 8 (T8) and T10 (**Supplemental Figure S1**). The time since injury was varied, with the majority

**Table 4:** Reference genomes and gene description files used.

| Organism | Reference Genome (Accession) | GTF |
|---|---|---|
| *Mus musculus* | GRCm38.p6 (mm10) | Ensembl v101 (Mus_musculus.GRCm38.101.gtf) |
| *Rattus norvegicus* | Rnor_6.0 (rn6) | Ensembl v101 (Rattus_norvegicus.Rnor_6.0.101.gtf) |
| *Xenopus laevis* | Xenla10.1 | Genbank (GCF_017654675.1_Xenopus_laevis_v10.1_genomic.gtf) |
| *Homo sapiens* | GRC38 (hg38) (GRC38.d1.vd1.fa) | Gencode v22 (gencode.v22.annotation.gtf) |
| *Mondelphis domestica* | ASM229v1 | Ensembl v104 (Monodelphis_domestica.ASM29v1.104.gtf) |
| *Petromyzon marinus* | kPetMar1 (GCF_010993605.1) | Genbank (GCF_0100993605.1_kPetMar1.pri_genomic.gtf) |
| *Danio rerio* | GRCz11 | Ensembl v105 (Danio_rerio_GRCz11.105.chr.gtf) |
| *Trachemys scripta elegans* | CAS_Tse_1.0 (GCF_013100865.1) | Genbank (GCF_013100865.1_CAS_Tse_1.0_genomic.gtf) |
| *Ambystoma mexicanum* | AmbMex60DD (GCA_002915635.3) | (Transcriptome constructed from Trinity) |

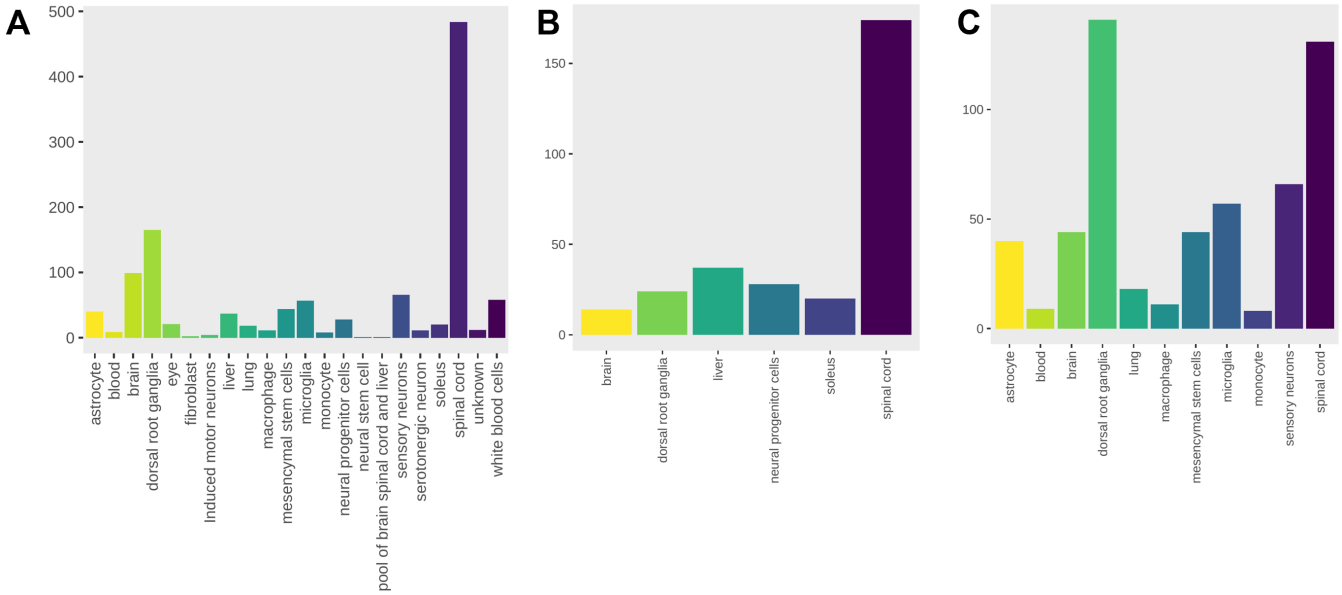


**Figure 2: Tissue types studied in filtered RNA-Seq experiments.** Shown are (A) the tissue types for all samples; (B) tissue types for rat experiments; and (C) tissue types for mouse experiments.

occurring within the first two weeks, although a number of studies extended the time, all the way up to 24 weeks (**Supplemental Figure S2**). Over 30 unique injury types are represented (**Supplemental Figure S3**) with the most frequent being complete transection, contusion, and various controls, including sham and laminectomy. In many cases, the exact control type could not be extracted from the GEO entry and any associated publication and is therefore listed as uninjured or control.

**SQLite database.** After the studies and their corresponding samples were processed, the resulting information was structured into an SQLite relational database consisting of 10 tables, as shown in the entity-relationship (ER) diagram in **Figure 3**. The total database size is just under 2.5 Gb.

**Web front-end**. Utilization of the SQLite database was performed using a web-based application for querying the data. Presented on the web page is information about the constructed database (**Figure 4**), links and summaries for included studies (**Supplemental Figure S4**), downloadable files (**Supplemental Figure S5**), and database exploration (**Figure 5**).

**Data exploration**. The user begins a data exploration by first selecting the organisms, tissues, injury types, injury locations, and specific genes of interest. They are then presented with the resulting sample groups that can be selected for a pairwise analysis (**Supplemental Figure S6**). Once the differential expression analysis is complete, the user is returned a set of results, including the top 25 up- and down-regulated genes (**Supplemental Figure S7**), heatmaps for the top 250 differentially expressed genes (**Supplemental Figure S8**) and genes with the highest variance, PCA plot, volcano plot, heatmap for genes of interest, and GO:BP and KEGG pathway enriched categories.

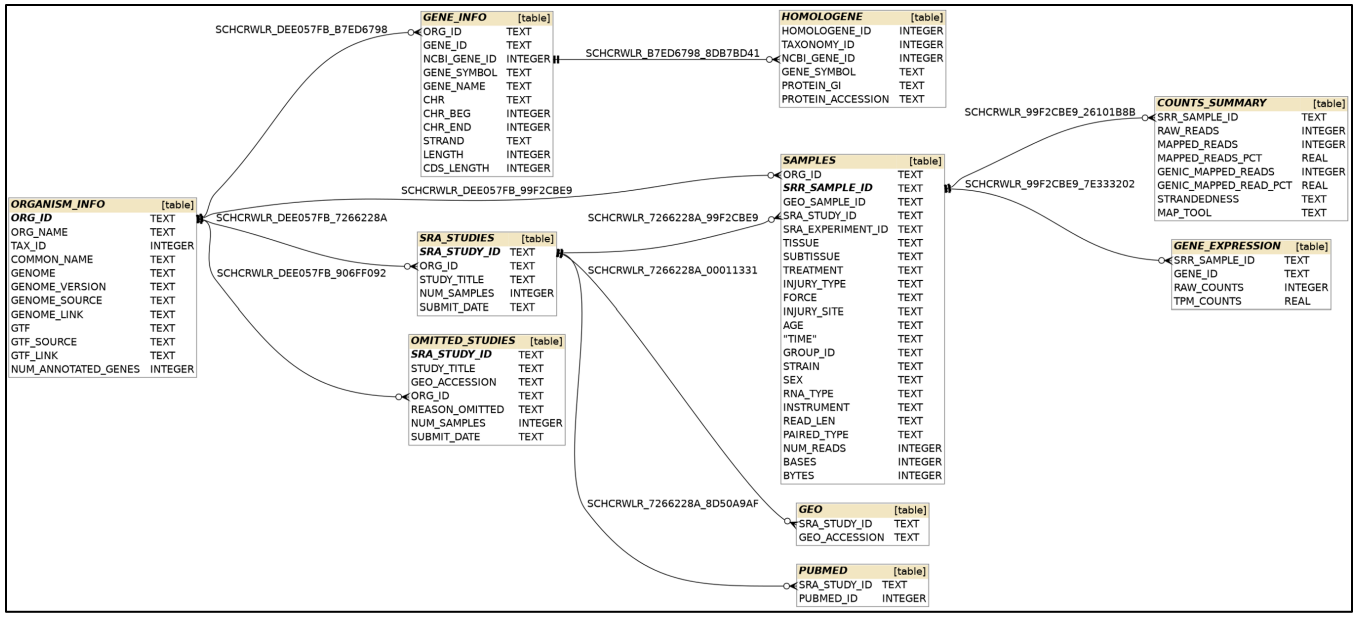


**Figure 3: Entity-relationship (ER) model for the SQLite database**.

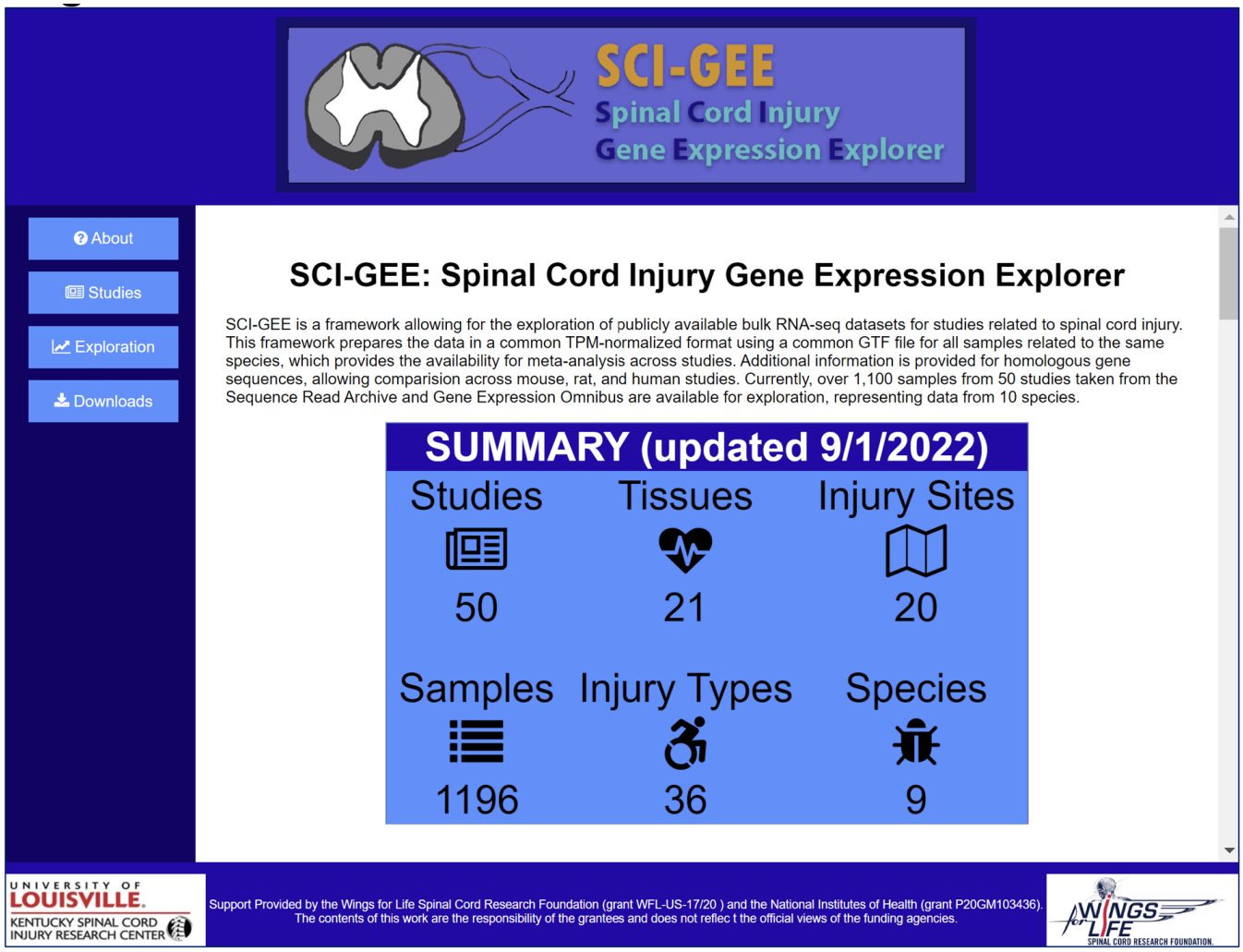


**Figure 4: SCI-GEE summary of SQLite database**. Shown is the web entry point into SCI-GEE with a summary of the number of data points in the database.

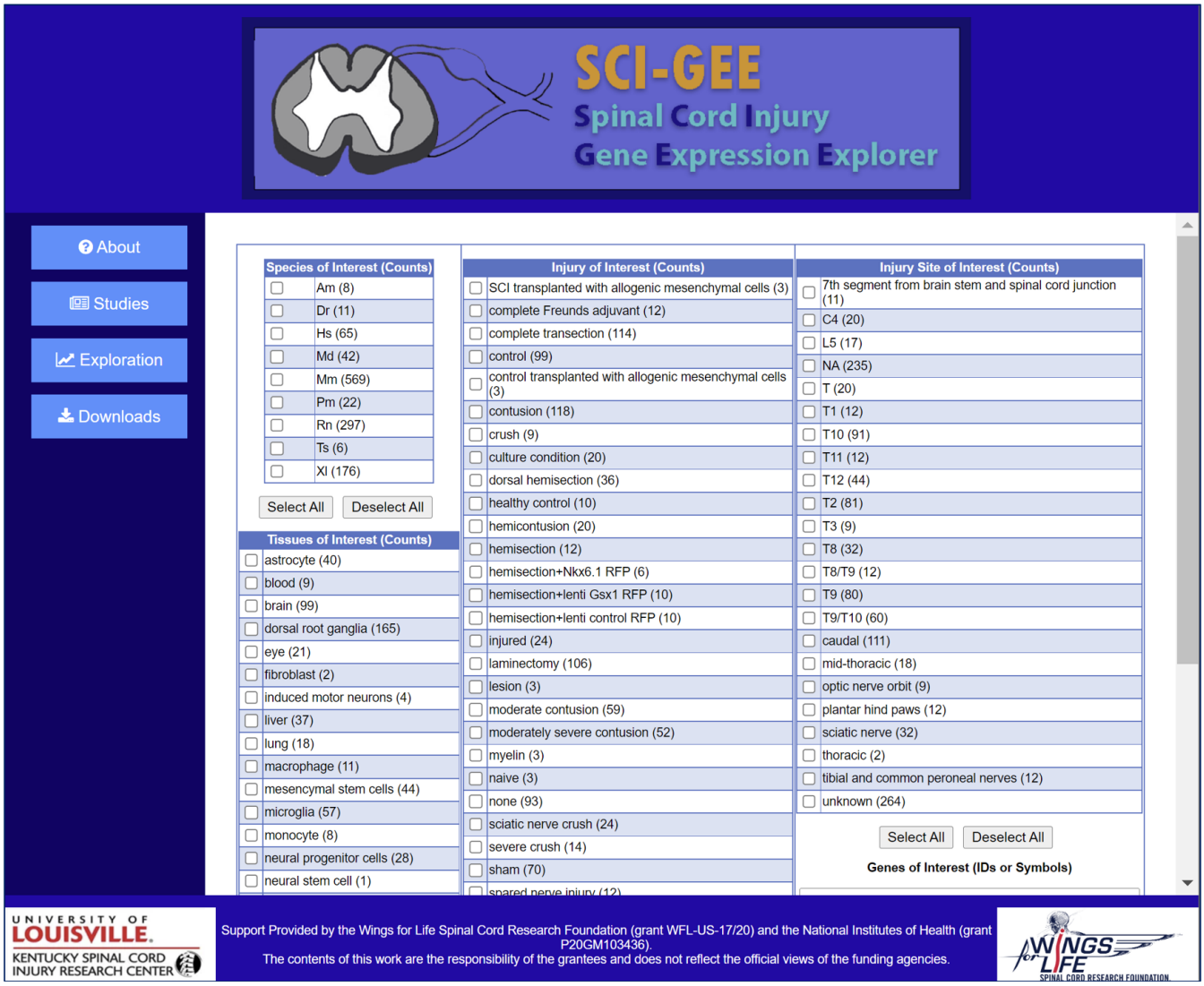


**Figure 5**: **Data exploration page**. SCI-GEE users are presented with choices for species of interest, tissues of interest, injury of interest, and injury site of interest.

## DISCUSSION

**Case Usage Studies.** Demonstration of the utility of SCI-GEE was performed with two usage studies in mouse and rat: 1) analysis of dorsal root ganglion (DRG) in spinal cord injured vs. uninjured samples; and 2) spinal cord tissue in spinal cord injured vs. uninjured samples. Mouse and rat were selected since they have the highest representation in the database. For both studies, results were available for three groups, including mouse and rat homologs; mouse genes only; and rat genes only. In the case of mouse and rat homologs, a total of 16,529 genes were compared based on shared homologene IDs [59], while the mouse samples were compared using 55,487 genes (including both protein coding and noncoding genes), and 32,883 rat genes were utilized for that comparison. While the data analyzed in SCI-GEE is available for download and utilization in other analytical workflows of interest, the results presented here are fully contained within the website functionality. For both tissue sets, the control and injured samples separated as illustrated by the PCA plots for DRG (**Figure 6A**) and spinal cord (**Figure 6B**), although a high degree of variance was found in the injured samples due to variability in other experimental factors, including injury type, severity, and time since injury.

**DRG tissue analysis**. For the DRG tissue dataset, two groups were created across mouse and rat: control samples (including both naïve and sham models) and spinal cord injured samples (**Supplemental Table S2**). The spinal cord injuries were from a variety of models, including hemisection, transection, and contusion. One of the meta-analysis capabilities of SCI-GEE is the ability to perform cross-species analysis across human, mouse, and rat gene expression by looking at the expression of common homologs across species. In these usage cases, shared homologene IDs across mouse and rat were examined. The top 25 up-regulated genes are shown in **Supplemental Table S3**, while the top 25 down-regulated genes are shown in **Supplemental Table S4**. Included in the up-regulated genes in the DRG for the mouse and rat meta-analysis are GADD45A, which is involved in cell cycle and neuronal cell death [64] and may act in a neuroprotective fashion [65-67]; GPR151 which is

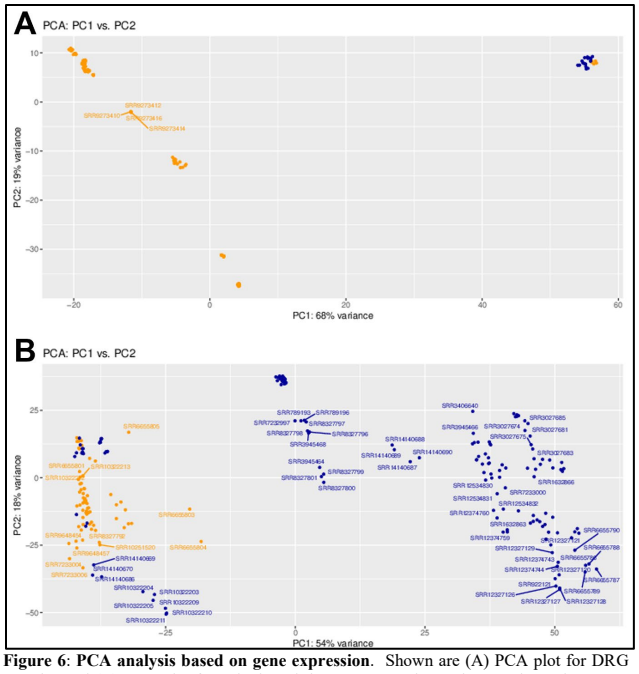


**Figure 6**: **PCA analysis based on gene expression**. Shown are (A) PCA plot for DRG samples and (B) PCA plot for spinal cord tissues. Control samples are shown in orange and injured samples are shown in blue.

involved in neuropathic pain, and may promote axon regeneration [68, 69]; FLRT3 which promotes neurite outgrowth [70-72], CRLF1 which forms complexes with neurotrophic factors to promote survival of neuronal cells [73], and the neuropeptide GAL [74, 75]. At the top of the down-regulated genes are two leucine zipper proteins (FOSB and FOS), two zinc fingers (EGR1 and ZFP36), and CYR61. These all play roles in regulating cell proliferation and apoptosis, indicating that regulatory machinery is turned off in response to spinal cord injury. In the DRG for mouse only, the top 25 up-regulated genes are shown in **Supplemental Table S5**, and the top 25 down-regulated are shown in **Supplemental Table S6**. At the top of the up-regulated genes are GADD45A, FST, GPR151, IGFN1, and TES. The up-regulation of FST (follistatin) and TES (testin LIM domain protein) may be indicative of sex-specific differences in the makeup of the mouse samples, while GPR151 is known to modulate neuropathic pain [76, 77] and IGFN1 is involved in synapse assembly [78]. The top mouse down-regulated genes are similar to those found for mouse and rat homologs. The top 25 up-regulated genes in DRG for the rat only are shown in **Supplemental Table S7**, and the top 25 down-regulated genes are shown in **Supplemental Table S8**. Among the up-regulated genes are several members of the Hox gene family (Hoxc11, Hoxd10, and Hoxd11), which are important for motor neuron patterning [79-82], as well as peripheral myelin protein 2 (Pmp2). The down-regulated genes include Usp5 which modulates neuropathic and inflammatory pain [83], Tuba1a which is necessary for central nervous system development and regeneration [84], Csf1r which promotes microglial proliferation [85, 86], and Slc39a6 which is found in reactive astrocytes [87].

For each of the comparisons (both, mouse and rat, respectively), volcano plots were generated to show the overall pattern of expression, including the fold-change across the x-axis and the q-value significance across the y-axis (**Figures 7A-C**), which

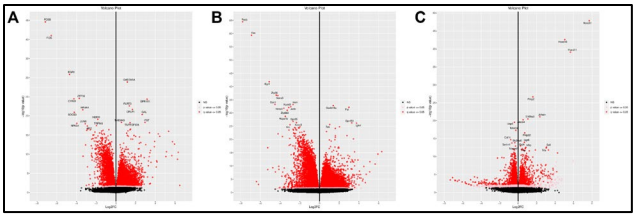


**Figure 7: Volcano plots shown for DRG samples.** Shown are (A) meta-analysis of mouse and rat homologs combined, (B) mouse genes, and (C) rat genes.

illustrates a relative even distribution of the data. Heatmaps were generated showing the top differentially expressed genes (**Figures 8A-C**), with the genes listed across the rows, and the samples across the columns.

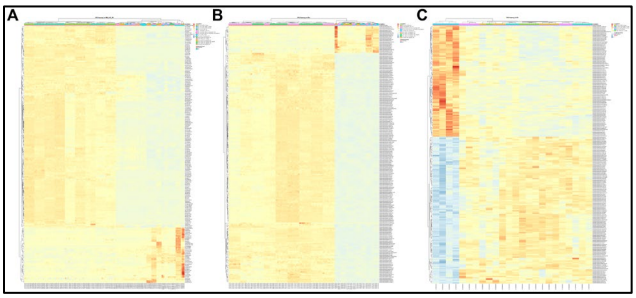


**Figure 8: Heatmaps for top differentially expressed genes in DRG injury vs. control comparison for (A) meta-analysis of mouse and rat homologs; (B) mouse samples; and (C) rat samples.**

**DRG enrichment analysis.** Differentially expressed genes from the mouse and rat comparisons were used as input into ClusterProfiler [63] for analysis of enriched GeneOntology Biological Processes (GO:BP) [61] and KEGG Metabolic Pathways [62]. The top 20 GO:BP enrichments are shown in **Figures 9A** and **9B** for mouse and rat respectively, while the top 20 KEGG enrichments are shown in **Figures 10A** and **10B**. For the mouse differentially expressed genes, several of the top GO:BP annotations are associated with nervous system development, indicating responses for regeneration and collateral sprouting, including synapse organization, regulation of neurogenesis, axonogenesis, epithelial tube morphogenesis, and dendrite development. Others are associated with muscular development, including muscle tissue development, muscle cell differentiation, and striated muscle cell differentiation. A third cluster is related to cell signaling and adhesion, including cell-substrate adhesion, positive regulation of cell adhesion, regulation of metal ion transport, calcium ion transport, and cell junction assembly. Results are similar for the rat, including synaptic vesicle cycle, vesicle-mediated transport in synapse, axonogenesis, regulation of neurotransmitter levels, sensory perception of pain, neurotransmitter transport, neurotransmitter secretion, signal release from synapse, regulation of membrane potential, potassium ion transport, positive regulation of ion transport, protein localization to cell junction, and synaptic vesicle exocytosis. The KEGG pathway enrichment results are a little more difficult to parse, but include pathways of neurodegeneration, axon guidance and synaptic vesicle cycle, along with a number of seemingly unrelated pathways, many of which are tied to proinflammatory responses.

**Spinal cord tissue analysis**. In the case of the spinal cord tissues, two groups were created across mouse and rat: control samples

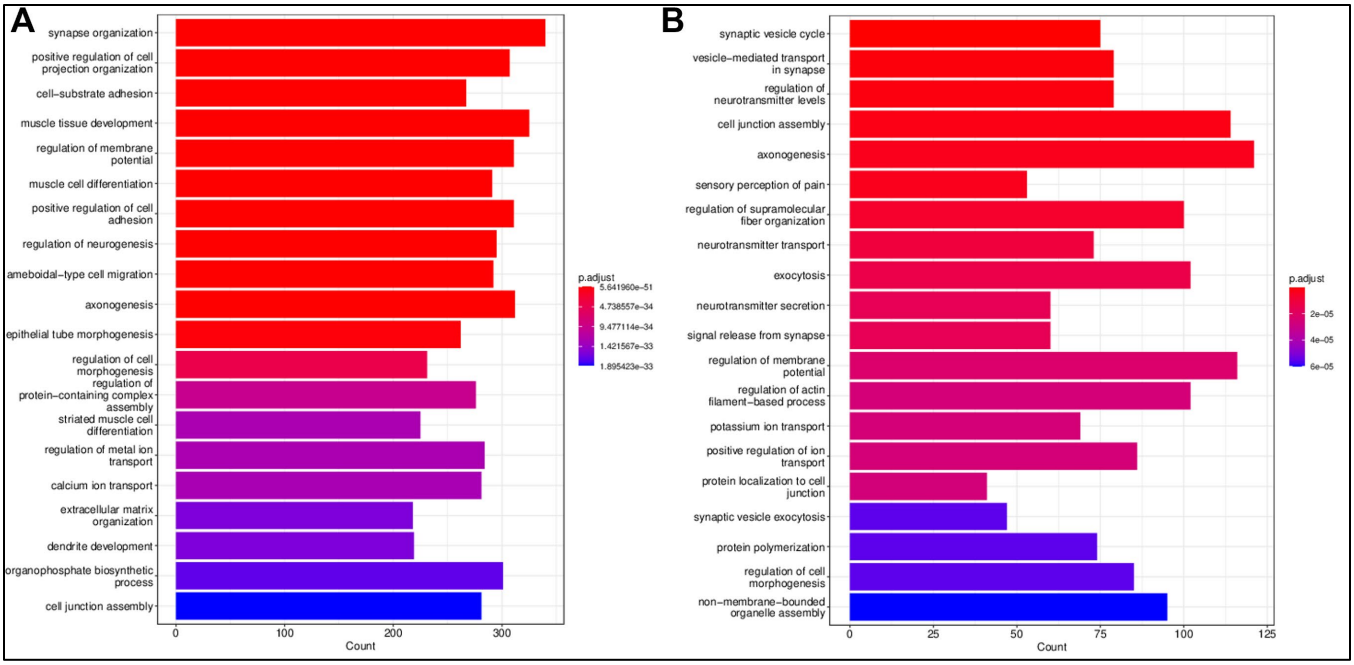


**Figure 9: Top 20 GO:BP enrichments from clusterProfiler for DRG injury vs. control comparison for (A) mouse samples and (B) rat samples.**

(including both naïve and sham models) and spinal cord injured samples (**Supplemental Table S9**). The spinal cord injuries were from a variety of models, including hemisection, transection, hemicontusion and contusion. Among the top up-regulated differentially expressed genes for the meta-analysis across mouse and rat homologs (**Supplemental Table S10**) are SIGLEC1 (CD169) which signals an increase in metallophilic macrophages and an increased immune response [88], GPNMB, a glioma-associated glycoprotein [89], MPEG1, a macrophage/microglia marker gene [90, 91], CCL13 which is produced from M2 macrophages [92], and CD68, a marker for phagocytic microglia [93]. Down-regulated genes (**Supplemental Table S11**) include HMGCS1, HMGCR MSMO1 and IDI1 which are involved in cholesterol metabolism, and whose down-regulation may prevent demyelination [94, 95]. The mouse only study shows similar results to the mouse and rat homologs, with Gpnmb, Ccl2, Bst2, Lyz2, and Trem2 among the top up-regulated genes (**Supplemental Table S12**). Ccl2 is a cytokine functioning in inflammation and pain following spinal cord injury [96] while Bst2 is a neuroinflammation biomarker [97]. Lyz2 is expressed in reactive microglia and macrophages [98], and Trem2 elicits a proinflammatory response in microglia [99]. Down-regulated genes include Cltrn, Aldob, Rapgef5, Mep1a, and Slc22a12

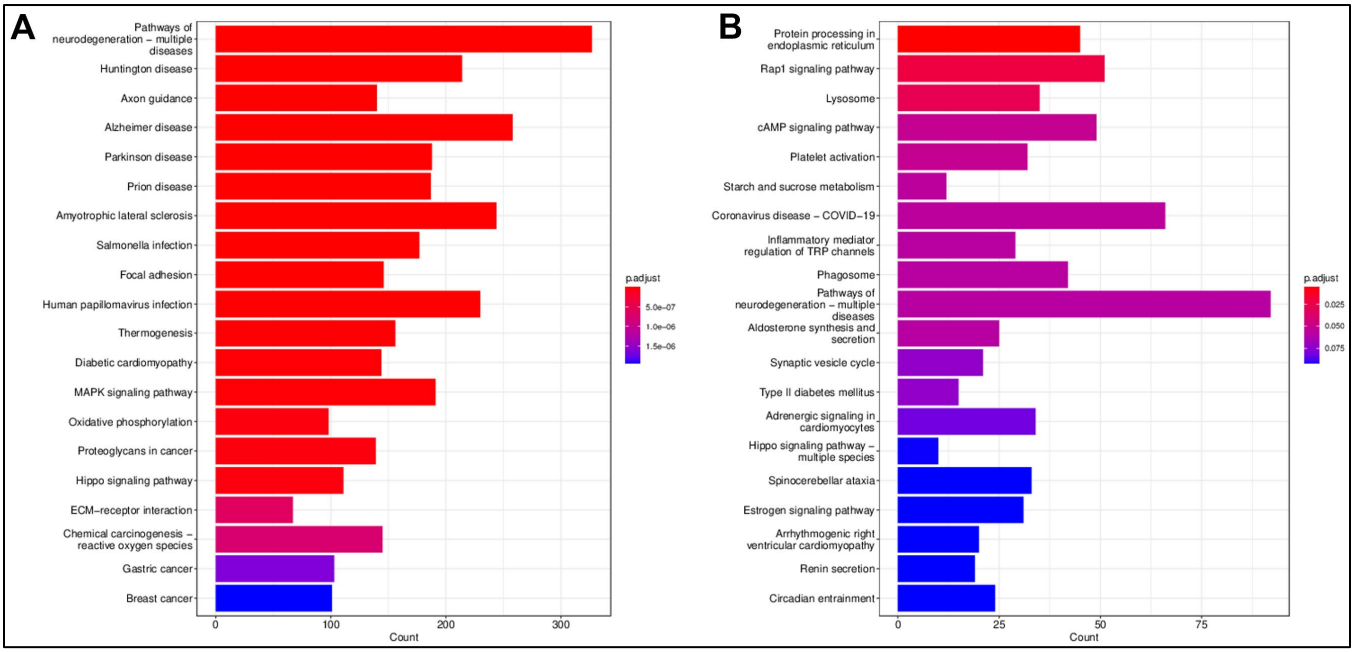


**Figure 10: Top 20 KEGG metabolic pathway enrichments from clusterProfiler for DRG injury vs. control comparison for (A) mouse samples and (B) rat samples.**

(**Supplemental Table S13**). Aldob has a role in glycolysis and glucogenesis [100] while Rapgef5 is involved in signal transduction with the Ras pathway [101]. Up-regulated genes for the rat-only spinal cord tissue includes Siglec1, Lilrb3, Gpnmb, Clec7a, and Mmp12 (**Supplemental Table S14**), similar to the results for the mouse and rat homologs. Ablation of Lilrb3 has been shown to promote neurite outgrowth [102], while Clec7a is a marker of actively proliferating microglia [103] and Mmp12 is involved in myelination. The top down-regulated rat genes include a number of genes with unknown function (**Supplemental Table S15**). Among those that have associated functions are Hmgcr, Mff, and Msmo1. As previously mentioned, Hmgcr and Msmo1 function in cholesterol metabolism, while Mff is associated with lengthening neuron life [104]. Volcano plots and heatmaps are provided in **Figures 11A-B** and **Figures 12A-B**, respectively.

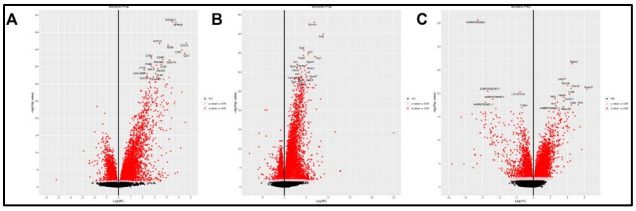


**Figure 11**: **Volcano plots shown for spinal cord samples.** Shown are (A) meta-analysis of mouse and rat homologs, (B) mouse genes, and (C) rat genes.

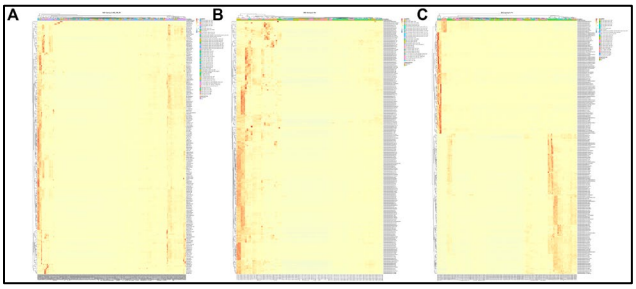


**Figure 12**: **Heatmaps for top differentially expressed genes in spinal cord injury vs. control comparison for (A) meta-analysis of mouse and rat homologs; (B) mouse samples; and (C) rat samples.**

**Spinal cord enrichment analysis**. Among the top 20 GO:BP enrichments (**Figures 13A-B**) are a number of proinflammatory annotations, including negative regulation of immune system process, regulation of inflammatory response, leukocyte migration, leukocyte cell-cell adhesion, cytokine-mediated signalling pathway, leukocyte proliferation, regulation of T cell activation, regulation of leukocyte migration, and myeloid leukocyte activation. Enriched KEGG metabolic pathways (**Figures 14A-B**) include a number of signaling pathways (NF-kappa B signaling pathway, TNF signaling pathway, MAPK signaling pathway, B cell receptor signaling pathway, NOD-like receptor signaling pathway, Toll-like receptor signaling pathway, cytoking-cytokine receptor interaction, and Hippo signaling pathway, indicating that spinal cord injury leads to a number of signaling cascades in the spinal cord itself, consistent with previous findings [105-109].

**Single cell transcriptomics**. More recent transcriptome work in the SCI field has included studies utilizing single cell RNA-Seq (scRNA-Seq) [98, 110-116], including a recent atlas of spinal cord injury in mice [114]. Due to the complexities of scRNA-Seq data, including cell typing and the disparity between limited sample numbers and large numbers of cells per sample, as well as a different modality for viewing scRNA-Seq data, we chose not to include that data at this time to focus on the more broadly available bulk RNA-Seq datasets. Over time, we anticipate that more scRNA-Seq datasets will be generated for SCI as costs come down. When that happens, we plan to develop a parallel method for integrating scRNA-Seq data. Such a resource would allow for more resolution at a single cell level, providing for an atlas of transcriptional changes for cells affected by SCI.

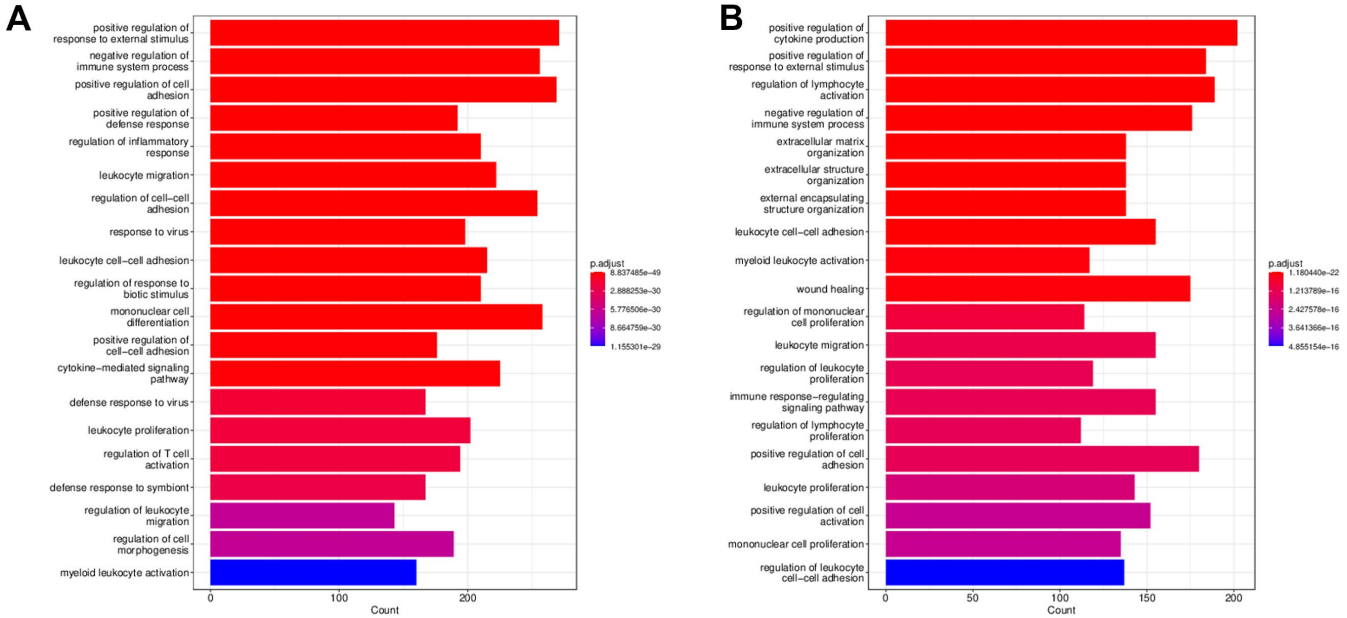


**Figure 13: Top 20 GO:BP enrichments from clusterProfiler for spinal cord injury vs. control comparison for (A) mouse samples and (B) rat samples.**

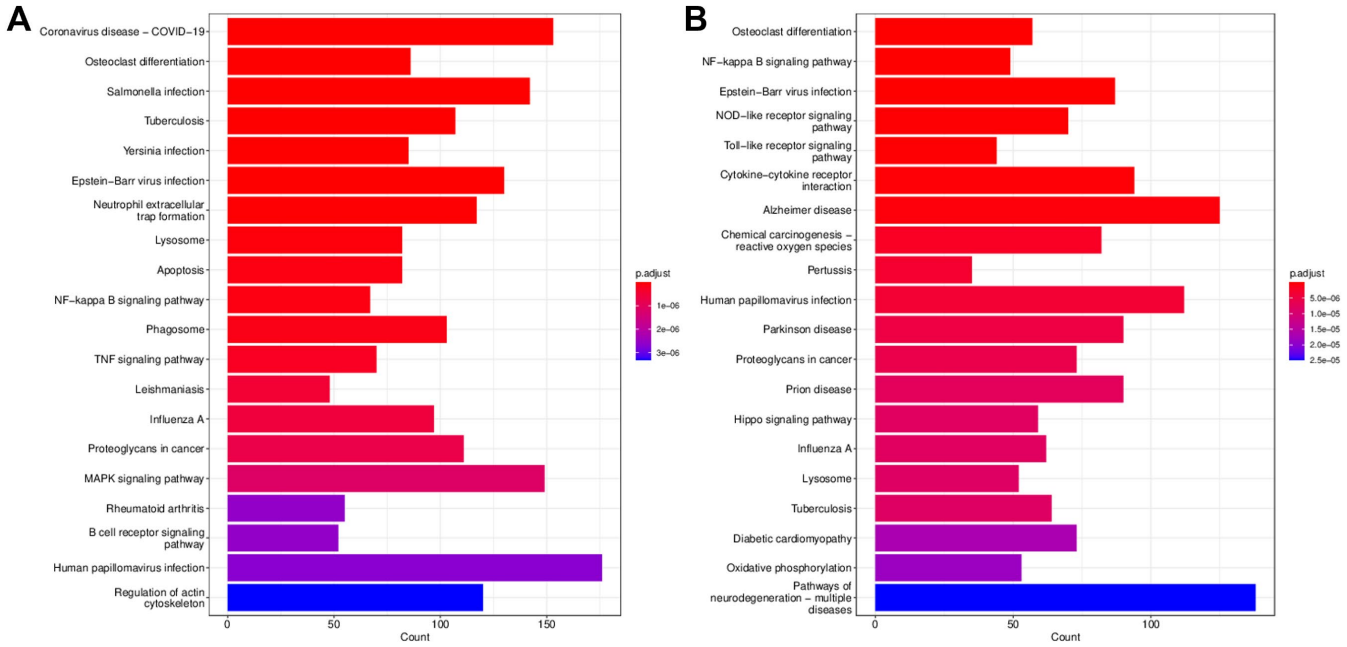


**Figure 14: Top 20 KEGG metabolic pathway enrichments from clusterProfiler for spinal cord injury vs. control comparison for (A) mouse samples and (B) rat samples.**

## CONCLUSION

Our approach for preparing bulk RNA-seq data related to SCI allows for greater interoperability across datasets. This in turn enables for meta-analyses both within and across species. By providing the data in an SQLite format, potential users are able to either utilize our web interface for exploring the available data, or develop their own analysis pipelines the utilizes the prepared data. We hope this work will eventually be incorporated into ODC-SCI data types, allowing for even greater integration along the genome-to-phenome modalities.

## ACKNOWLEDGEMENTS

Support provided by the Wings for Life Spinal Cord Research Foundation (grant WFL-US-17/20) and the National Institutes of Health (grant P20GM103436). The contents of this manuscript do not reflect the official views of the funding agencies.

## AVAILABILITY

The SQLite database, flat files used to construct the database, and the web interface can be accessed at: http://162.215.210.70/~tracks/SCI-GEE/ .

## REFERENCES

[1] National Spinal Cord Injury Statistics Center, "Spinal Cord Injury Model Systems 2021 Annual Report," University of Alabama at Birmingham, https://www.nscisc.uab.edu/PublicDocuments/AR2021_public%20version.pdf , 2021 2021.

[2] National Spinal Cord Injury Statistics Center. "Frequently Asked Questions." https://www.nscisc.uab.edu/Public_Pages/FAQ (accessed 9/28/2022, 2022).

[3] N. B. Jain *et al.*, "Traumatic spinal cord injury in the United States, 1993-2012," *Jama,* vol. 313, no. 22, pp. 2236-2243, 2015.

[4] Christopher and Dana Reeves Foundation. "Stats about paralysis." https://www.christopherreeve.org/living-with-paralysis/stats-about-paralysis (accessed 9/28/2022, 2022).

[5] M. A. Anderson *et al.*, "Required growth facilitators propel axon regeneration across complete spinal cord injury," *Nature,* vol. 561, no. 7723, pp. 396-400, 2018.

[6] B. S. Bregman, M. McAtee, H. N. Dai, and P. L. Kuhn, "Neurotrophic factors increase axonal growth after spinal cord injury and transplantation in the adult rat," *Experimental neurology,* vol. 148, no. 2, pp. 475-494, 1997.

[7] F. Hellal *et al.*, "Microtubule stabilization reduces scarring and causes axon regeneration after spinal cord injury," *Science,* vol. 331, no. 6019, pp. 928-931, 2011.

[8] P. Lu, L. Jones, E. Snyder, and M. Tuszynski, "Neural stem cells constitutively secrete neurotrophic factors and promote extensive host axonal growth after spinal cord injury," *Experimental neurology,* vol. 181, no. 2, pp. 115-129, 2003.

[9] P. Lu, L. Jones, and M. Tuszynski, "BDNF-expressing marrow stromal cells support extensive axonal growth at sites of spinal cord injury," *Experimental neurology,* vol. 191, no. 2, pp. 344-360, 2005.

[10] P. Lu *et al.*, "Long-distance axonal growth from human induced pluripotent stem cells after spinal cord injury," *Neuron,* vol. 83, no. 4, pp. 789-796, 2014.

[11] A. Barritt *et al.*, "Chondroitinase ABC promotes sprouting of intact and injured spinal systems after spinal cord injury," *Journal of Neuroscience,* vol. 26, no. 42, pp. 10856-10867, 2006.

[12] K. Fouad, V. Pedersen, M. E. Schwab, and C. Brösamle, "Cervical sprouting of corticospinal fibers after thoracic spinal cord injury accompanies shifts in evoked motor responses," *Current Biology,* vol. 11, no. 22, pp. 1766-1770, 2001.

[13] J. K. Lee *et al.*, "Assessing spinal axon regeneration and sprouting in Nogo-, MAG-, and OMgp-deficient mice," *Neuron,* vol. 66, no. 5, pp. 663-670, 2010.

[14] I. C. Maier and M. E. Schwab, "Sprouting, regeneration and circuit formation in the injured spinal cord: factors and activity," *Philosophical Transactions of the Royal Society B: Biological Sciences,* vol. 361, no. 1473, pp. 1611-1634, 2006.

[15] O. Raineteau and M. E. Schwab, "Plasticity of motor systems after incomplete spinal cord injury," *Nature Reviews Neuroscience,* vol. 2, no. 4, pp. 263-273, 2001.

[16] M. E. Schwab, "Repairing the injured spinal cord," *Science,* vol. 295, no. 5557, pp. 1029-1031, 2002.

[17] N. Weidner, A. Ner, N. Salimi, and M. H. Tuszynski, "Spontaneous corticospinal axonal plasticity and functional recovery after adult central nervous system injury," *Proceedings of the National Academy of Sciences,* vol. 98, no. 6, pp. 3513-3518, 2001.

[18] F. Barnabé-Heider *et al.*, "Origin of new glial cells in intact and injured adult spinal cord," *Cell stem cell,* vol. 7, no. 4, pp. 470-482, 2010.

[19] A. D. Gaudet and L. K. Fonken, "Glial cells shape pathology and repair after spinal cord injury," *Neurotherapeutics,* vol. 15, no. 3, pp. 554-577, 2018.

[20] Y. S. Gwak, J. Kang, G. C. Unabia, and C. E. Hulsebosch, "Spatial and temporal activation of spinal glial cells: role of gliopathy in central neuropathic pain following spinal cord injury in rats," *Experimental neurology,* vol. 234, no. 2, pp. 362-372, 2012.
[21] D. Kim *et al.*, "A critical role of toll-like receptor 2 in nerve injury-induced spinal cord glial cell activation and pain hypersensitivity," *Journal of Biological Chemistry,* vol. 282, no. 20, pp. 14975-14983, 2007.
[22] X. Z. Liu *et al.*, "Neuronal and glial apoptosis after traumatic spinal cord injury," *Journal of Neuroscience,* vol. 17, no. 14, pp. 5395-5406, 1997.
[23] B. Gungor, E. Adiguzel, I. Gursel, B. Yilmaz, and M. Gursel, "Intestinal microbiota in patients with spinal cord injury," *PloS one,* vol. 11, no. 1, p. e0145878, 2016.
[24] K. A. Kigerl, J. C. Hall, L. Wang, X. Mo, Z. Yu, and P. G. Popovich, "Gut dysbiosis impairs recovery after spinal cord injury," *Journal of Experimental Medicine,* vol. 213, no. 12, pp. 2603-2620, 2016.
[25] K. A. Kigerl, K. Mostacada, and P. G. Popovich, "Gut microbiota are disease-modifying factors after traumatic spinal cord injury," *Neurotherapeutics,* vol. 15, no. 1, pp. 60-67, 2018.
[26] C. Zhang *et al.*, "Gut microbiota dysbiosis in male patients with chronic traumatic complete spinal cord injury," *Journal of translational medicine,* vol. 16, no. 1, pp. 1-16, 2018.
[27] C. R. Battistuzzo, R. J. Callister, R. Callister, and M. P. Galea, "A systematic review of exercise training to promote locomotor recovery in animal models of spinal cord injury," *Journal of neurotrauma,* vol. 29, no. 8, pp. 1600-1613, 2012.
[28] J. H. Chariker *et al.*, "Activity/exercise-induced changes in the liver transcriptome after chronic spinal cord injury," *Scientific data,* vol. 6, no. 1, pp. 1-9, 2019.
[29] E. E. Dupont-Versteegden *et al.*, "Exercise-induced gene expression in soleus muscle is dependent on time after spinal cord injury in rats," *Muscle & Nerve: Official Journal of the American Association of Electrodiagnostic Medicine,* vol. 29, no. 1, pp. 73-81, 2004.
[30] C. Engesser-Cesar, A. J. Anderson, D. M. Basso, V. Edgerton, and C. W. Cotman, "Voluntary wheel running improves recovery from a moderate spinal cord injury," *Journal of neurotrauma,* vol. 22, no. 1, pp. 157-171, 2005.
[31] J. Fu, H. Wang, L. Deng, and J. Li, "Exercise training promotes functional recovery after spinal cord injury," *Neural plasticity,* vol. 2016, 2016.
[32] K. J. Hutchinson, F. Gómez-Pinilla, M. J. Crowe, Z. Ying, and D. M. Basso, "Three exercise paradigms differentially improve sensory recovery after spinal cord contusion in rats," *Brain,* vol. 127, no. 6, pp. 1403-1414, 2004.
[33] H. R. Sandrow-Feinberg and J. D. Houlé, "Exercise after spinal cord injury as an agent for neuroprotection, regeneration and rehabilitation," *Brain research,* vol. 1619, pp. 12-21, 2015.
[34] R. R. Smith *et al.*, "Swim training initiated acutely after spinal cord injury is ineffective and induces extravasation in and around the epicenter," *Journal of neurotrauma,* vol. 26, no. 7, pp. 1017-1027, 2009.
[35] R. R. Smith *et al.*, "Effects of swimming on functional recovery after incomplete spinal cord injury in rats," *Journal of neurotrauma,* vol. 23, no. 6, pp. 908-919, 2006.
[36] N. L. Van Meeteren, R. Eggers, A. J. Lankhorst, W. H. Gispen, and F. P. Hamers, "Locomotor recovery after spinal cord contusion injury in rats is improved by spontaneous exercise," *Journal of neurotrauma,* vol. 20, no. 10, pp. 1029-1037, 2003.
[37] A. Anjum *et al.*, "Spinal cord injury: pathophysiology, multimolecular interactions, and underlying recovery mechanisms," *International journal of molecular sciences,* vol. 21, no. 20, p. 7533, 2020.
[38] A. Callahan *et al.*, "Developing a data sharing community for spinal cord injury research," *Experimental neurology,* vol. 295, pp. 135-143, 2017.
[39] M. D. Wilkinson *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific data,* vol. 3, no. 1, pp. 1-9, 2016.
[40] R. Edgar, M. Domrachev, and A. E. Lash, "Gene Expression Omnibus: NCBI gene expression and hybridization array data repository," *Nucleic acids research,* vol. 30, no. 1, pp. 207-210, 2002.
[41] R. Leinonen, H. Sugawara, M. Shumway, and I. N. S. D. Collaboration, "The sequence read archive," *Nucleic acids research,* vol. 39, no. suppl_1, pp. D19-D21, 2010.
[42] A. Brazma *et al.*, "Minimum information about a microarray experiment (MIAME)—toward standards for microarray data," *Nature genetics,* vol. 29, no. 4, pp. 365-371, 2001.
[43] A. Brazma *et al.*, "ArrayExpress—a public repository for microarray gene expression data at the EBI," *Nucleic acids research,* vol. 31, no. 1, pp. 68-71, 2003.
[44] M. D. Mailman *et al.*, "The NCBI dbGaP database of genotypes and phenotypes," *Nature genetics,* vol. 39, no. 10, pp. 1181-1186, 2007.
[45] V. P. Lemmon *et al.*, "Minimum information about a spinal cord injury experiment: a proposed reporting standard for spinal cord injury experiments," *Journal of neurotrauma,* vol. 31, no. 15, pp. 1354-1361, 2014.
[46] K. Chen *et al.*, "RNA-seq characterization of spinal cord injury transcriptome in acute/subacute phases: a resource for understanding the pathology at the systems level," *PloS one,* vol. 8, no. 8, p. e72567, 2013.
[47] J. N. Weinstein *et al.*, "The cancer genome atlas pan-cancer analysis project," *Nature genetics,* vol. 45, no. 10, pp. 1113-1120, 2013.
[48] R. L. Grossman *et al.*, "Toward a shared vision for cancer genomic data," *New England Journal of Medicine,* vol. 375, no. 12, pp. 1109-1112, 2016.
[49] J. Lonsdale *et al.*, "The genotype-tissue expression (GTEx) project," *Nature genetics,* vol. 45, no. 6, pp. 580-585, 2013.
[50] E. W. Sayers, C. O'Sullivan, and I. Karsch-Mizrachi, "Using GenBank and SRA," in *Plant Bioinformatics*: Springer, 2022, pp. 1-25.
[51] S. Andrews, "FastQC: a quality control tool for high throughput sequence data," ed: Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom, 2010.
[52] A. Dobin *et al.*, "STAR: ultrafast universal RNA-seq aligner," *Bioinformatics,* vol. 29, no. 1, pp. 15-21, 2013.
[53] Y. Liao, G. K. Smyth, and W. Shi, "featureCounts: an efficient general purpose program for assigning sequence reads to genomic features," *Bioinformatics,* vol. 30, no. 7, pp. 923-930, 2014.
[54] L. Wang, S. Wang, and W. Li, "RSeQC: quality control of RNA-seq experiments," *Bioinformatics,* vol. 28, no. 16, pp. 2184-2185, 2012.
[55] S. Anders, P. T. Pyl, and W. Huber, "HTSeq—a Python framework to work with high-throughput sequencing data," *bioinformatics,* vol. 31, no. 2, pp. 166-169, 2015.
[56] G. P. Wagner, K. Kin, and V. J. Lynch, "Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples," *Theory in biosciences,* vol. 131, no. 4, pp. 281-285, 2012.
[57] M. Lawrence *et al.*, "Software for computing and annotating genomic ranges," *PLoS computational biology,* vol. 9, no. 8, p. e1003118, 2013.
[58] B. J. Haas *et al.*, "De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis," *Nature protocols,* vol. 8, no. 8, pp. 1494-1512, 2013.
[59] N. R. Coordinators, "Database resources of the national center for biotechnology information," *Nucleic acids research,* vol. 44, no. D1, pp. D7-D19, 2016.
[60] M. I. Love, W. Huber, and S. Anders, "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2," *Genome biology,* vol. 15, no. 12, pp. 1-21, 2014.
[61] M. Ashburner *et al.*, "Gene ontology: tool for the unification of biology," *Nature genetics,* vol. 25, no. 1, pp. 25-29, 2000.
[62] M. Kanehisa and S. Goto, "KEGG: kyoto encyclopedia of genes and genomes," *Nucleic acids research,* vol. 28, no. 1, pp. 27-30, 2000.
[63] G. Yu, L.-G. Wang, Y. Han, and Q.-Y. He, "clusterProfiler: an R package for comparing biological themes among gene clusters," *Omics: a journal of integrative biology,* vol. 16, no. 5, pp. 284-287, 2012.
[64] S. Di Giovanni, S. M. Knoblach, C. Brandoli, S. A. Aden, E. P. Hoffman, and A. I. Faden, "Gene profiling in spinal cord injury shows role of cell cycle in neuronal death," *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society,* vol. 53, no. 4, pp. 454-468, 2003.
[65] K. Befort, L. Karchewski, C. Lanoue, and C. J. Woolf, "Selective up-regulation of the growth arrest DNA damage-inducible gene Gadd45 alpha in sensory and motor neurons after peripheral nerve injury," *European Journal of Neuroscience,* vol. 18, no. 4, pp. 911-922, 2003.
[66] A. Blesch *et al.*, "Conditioning lesions before or after spinal cord injury recruit broad genetic mechanisms that sustain axonal regeneration: superiority to camp-mediated effects," *Experimental neurology,* vol. 235, no. 1, pp. 162-173, 2012.
[67] C. R. Lin *et al.*, "GADD45A protects against cell death in dorsal root ganglion neurons following peripheral nerve injury," *Journal of neuroscience research,* vol. 89, no. 5, pp. 689-699, 2011.
[68] F. E. Holmes, N. Kerr, Y.-J. Chen, P. Vanderplank, C. A. McArdle, and D. Wynick, "Targeted disruption of the orphan receptor Gpr151 does not alter pain-related behaviour despite a strong induction in dorsal root ganglion expression in a model of neuropathic pain," *Molecular and Cellular Neuroscience,* vol. 78, pp. 35-40, 2017.
[69] B. Lee *et al.*, "Promoting axon regeneration by enhancing the non-coding function of the injury-responsive coding gene Gpr151," *bioRxiv,* 2021.
[70] M. Robinson *et al.*, "FLRT3 is expressed in sensory neurons after peripheral nerve injury and regulates neurite outgrowth," *Molecular and Cellular Neuroscience,* vol. 27, no. 2, pp. 202-214, 2004.
[71] K. Tanabe, I. Bonilla, J. A. Winkles, and S. M. Strittmatter, "Fibroblast growth factor-inducible-14 is induced in axotomized neurons and promotes neurite outgrowth," *Journal of Neuroscience,* vol. 23, no. 29, pp. 9675-9686, 2003.
[72] L. Tsuji *et al.*, "FLRT3, a cell surface molecule containing LRR repeats and a FNIII domain, promotes neurite outgrowth," *Biochemical and biophysical research communications,* vol. 313, no. 4, pp. 1086-1091, 2004.
[73] L. Crisponi, I. Buers, and F. Rutsch, "CRLF1 and CLCF1 in Development, Health and Disease," *International journal of molecular sciences,* vol. 23, no. 2, p. 992, 2022.
[74] S. Hobson *et al.*, "Galanin acts as a trophic factor to the central and peripheral nervous systems," *Cellular and molecular life sciences: CMLS,* vol. 65, no. 12, pp. 1806-1812, 2008.

[75] D. Wynick, S. W. Thompson, and S. B. McMahon, "The role of galanin as a multi-functional neuropeptide in the nervous system," *Current opinion in pharmacology,* vol. 1, no. 1, pp. 73-77, 2001.
[76] B.-C. Jiang *et al.*, "Demethylation of G-protein-coupled receptor 151 promoter facilitates the binding of Krüppel-like factor 5 and enhances neuropathic pain after nerve injury in mice," *Journal of Neuroscience,* vol. 38, no. 49, pp. 10535-10551, 2018.
[77] L.-P. Xia *et al.*, "GPR151 in nociceptors modulates neuropathic pain via regulating P2X3 function and microglial activation," *Brain,* vol. 144, no. 11, pp. 3405-3420, 2021.
[78] P. Gaudet, M. S. Livstone, S. E. Lewis, and P. D. Thomas, "Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium," *Briefings in bioinformatics,* vol. 12, no. 5, pp. 449-462, 2011.
[79] H. Kumamaru *et al.*, "Generation and post-injury integration of human spinal cord neural stem cells," *Nature methods,* vol. 15, no. 9, pp. 723-731, 2018.
[80] I. M. Pereira, A. Marote, A. J. Salgado, and N. A. Silva, "Filling the gap: neural stem cells as a promising therapy for spinal cord injury," *Pharmaceuticals,* vol. 12, no. 2, p. 65, 2019.
[81] N. White and S. E. Sakiyama-Elbert, "Derivation of specific neural populations from pluripotent cells for understanding and treatment of spinal cord injury," *Developmental Dynamics,* vol. 248, no. 1, pp. 78-87, 2019.
[82] C. Yu *et al.*, "The application of neural stem/progenitor cells for regenerative therapy of spinal cord injury," *Current stem cell research & therapy,* vol. 14, no. 6, pp. 495-503, 2019.
[83] A. García-Caballero *et al.*, "The deubiquitinating enzyme USP5 modulates neuropathic and inflammatory pain by enhancing Cav3. 2 channel activity," *Neuron,* vol. 83, no. 5, pp. 1144-1158, 2014.
[84] M. B. Veldman, M. A. Bemben, and D. Goldman, "Tuba1a gene expression is regulated by KLF6/7 and is necessary for CNS development and regeneration in zebrafish," *Molecular and Cellular Neuroscience,* vol. 43, no. 4, pp. 370-383, 2010.
[85] H. Fu, Y. Zhao, D. Hu, S. Wang, T. Yu, and L. Zhang, "Depletion of microglia exacerbates injury and impairs function recovery after spinal cord injury in mice," *Cell death & disease,* vol. 11, no. 7, pp. 1-12, 2020.
[86] Y. N. Gerber *et al.*, "CSF1R inhibition reduces microglia proliferation, promotes tissue preservation and improves motor recovery after spinal cord injury," *Frontiers in cellular neuroscience,* vol. 12, p. 368, 2018.
[87] S. Okada *et al.*, "Conditional ablation of Stat3 or Socs3 discloses a dual role for reactive astrocytes after spinal cord injury," *Nature medicine,* vol. 12, no. 7, pp. 829-834, 2006.
[88] B. T. Noble, F. H. Brennan, and P. G. Popovich, "The spleen as a neuroimmune interface after spinal cord injury," *Journal of Neuroimmunology,* vol. 321, pp. 1-11, 2018.
[89] J. J. Huang, W. J. Ma, and S. Yokoyama, "Expression and immunolocalization of Gpnmb, a glioma-associated glycoprotein, in normal and inflamed central nervous systems of adult rats," *Brain and behavior,* vol. 2, no. 2, pp. 85-96, 2012.
[90] L. Cavone *et al.*, "A unique macrophage subpopulation signals directly to progenitor cells to promote regenerative neurogenesis in the zebrafish spinal cord," *Developmental cell,* vol. 56, no. 11, pp. 1617-1630. e6, 2021.
[91] Y. Goldshmit *et al.*, "Blockage of lysophosphatidic acid signaling improves spinal cord injury outcomes," *The American journal of pathology,* vol. 181, no. 3, pp. 978-992, 2012.
[92] A. Jaerve and H. W. Müller, "Chemokines in CNS injury and repair," *Cell and tissue research,* vol. 349, no. 1, pp. 229-248, 2012.
[93] A. I. Faden, J. Wu, B. A. Stoica, and D. J. Loane, "Progressive inflammation-mediated neurodegeneration after traumatic brain or spinal cord injury," *British journal of pharmacology,* vol. 173, no. 4, pp. 681-691, 2016.
[94] G. Chen, X. Fang, and M. Yu, "Regulation of gene expression in rats with spinal cord injury based on microarray data," *Molecular medicine reports,* vol. 12, no. 2, pp. 2465-2472, 2015.
[95] J. Zhang *et al.*, "Astrocytic YAP prevents the demyelination through promoting expression of cholesterol synthesis genes in experimental autoimmune encephalomyelitis," *Cell death & disease,* vol. 12, no. 10, pp. 1-11, 2021.
[96] J. Van Steenwinckel *et al.*, "CCL2 released from neuronal synaptic vesicles in the spinal cord is a major mediator of local inflammation and pain after peripheral nerve injury," *Journal of Neuroscience,* vol. 31, no. 15, pp. 5865-5875, 2011.
[97] X. Xu *et al.*, "Bone Marrow Stromal Cell Antigen 2: Is a Potential Neuroinflammation Biomarker of SOD1G93A Mouse Model of Amyotrophic Lateral Sclerosis in Pre-symptomatic Stage," *Frontiers in Neuroscience,* vol. 15, 2021.
[98] S. Wahane *et al.*, "Diversified transcriptional responses of myeloid and glial cells in spinal cord injury shaped by HDAC3 activity," *Science Advances,* vol. 7, no. 9, p. eabd8811, 2021.
[99] M. Kobayashi, H. Konishi, A. Sayo, T. Takai, and H. Kiyama, "TREM2/DAP12 signal elicits proinflammatory response in microglia and exacerbates neuropathic pain," *Journal of Neuroscience,* vol. 36, no. 43, pp. 11138-11150, 2016.
[100] A. D. Gaudet *et al.*, "Spinal cord injury in rats dysregulates diurnal rhythms of fecal output and liver metabolic indicators," *Journal of Neurotrauma,* vol. 36, no. 12, pp. 1923-1934, 2019.
[101] S. S. Dhar *et al.*, "MLL4 is required to maintain broad H3K4me3 peaks and super-enhancers at tumor suppressor genes," *Molecular cell,* vol. 70, no. 5, pp. 825-841. e6, 2018.
[102] T. L. Dickendesher *et al.*, "NgR1 and NgR3 are receptors for chondroitin sulfate proteoglycans," *Nature neuroscience,* vol. 15, no. 5, pp. 703-712, 2012.
[103] S. Brockie, J. Hong, and M. G. Fehlings, "The role of microglia in modulating neuroinflammation after spinal cord injury," *International Journal of Molecular Sciences,* vol. 22, no. 18, p. 9706, 2021.
[104] M. Cordaro, G. Casili, I. Paterniti, S. Cuzzocrea, and E. Esposito, "Fumaric acid esters attenuate secondary degeneration after spinal cord injury," *Journal of neurotrauma,* vol. 34, no. 21, pp. 3027-3040, 2017.
[105] N. Abe and V. Cavalli, "Nerve injury signaling," *Current opinion in neurobiology,* vol. 18, no. 3, pp. 276-283, 2008.
[106] D. Bastien and S. Lacroix, "Cytokine pathways regulating glial and leukocyte function after spinal cord and peripheral nerve injury," *Experimental neurology,* vol. 258, pp. 62-77, 2014.
[107] T. Genovese *et al.*, "Effects of palmitoylethanolamide on signaling pathways implicated in the development of spinal cord injury," *Journal of Pharmacology and Experimental Therapeutics,* vol. 326, no. 1, pp. 12-23, 2008.
[108] R. W. Keane, A. R. Davis, and W. D. Dietrich, "Inflammatory and apoptotic signaling after spinal cord injury," *Journal of neurotrauma,* vol. 23, no. 3-4, pp. 335-344, 2006.
[109] C.-G. Yu and R. P. Yezierski, "Activation of the ERK1/2 signaling cascade by excitotoxic spinal cord injury," *Molecular brain research,* vol. 138, no. 2, pp. 244-255, 2005.
[110] O. Avraham, R. Feng, E. E. Ewan, J. Rustenhoven, G. Zhao, and V. Cavalli, "Profiling sensory neuron microenvironment after peripheral and central axon injury reveals key pathways for neural repair," *Elife,* vol. 10, p. e68457, 2021.
[111] E. M. Floriddia *et al.*, "Distinct oligodendrocyte populations have spatial preference and different responses to spinal cord injury," *Nature communications,* vol. 11, no. 1, pp. 1-15, 2020.
[112] J. Hou *et al.*, "Heterogeneity analysis of astrocytes following spinal cord injury at single-cell resolution," *The FASEB Journal,* vol. 36, no. 8, p. e22442, 2022.
[113] Y. Li *et al.*, "Microglia-organized scar-free spinal cord repair in neonatal mice," *Nature,* vol. 587, no. 7835, pp. 613-618, 2020.
[114] K. J. Matson *et al.*, "Single cell atlas of spinal cord injury in mice reveals a pro-regenerative signature in spinocerebellar neurons," *Nature Communications,* vol. 13, no. 1, pp. 1-16, 2022.
[115] J. W. Squair *et al.*, "Confronting false discoveries in single-cell differential expression," *Nature communications,* vol. 12, no. 1, pp. 1-15, 2021.
[116] J. Wang *et al.*, "Single-cell transcriptome analysis reveals the immune heterogeneity and the repopulation of microglia by Hif1α in mice after spinal cord injury," *Cell death & disease,* vol. 13, no. 5, pp. 1-12, 2022.